Evaluating and Communicating the Reliability of Data from Wearable Devices for Health Tracking

Abstract:

Wearable health devices have become universal tools for health monitoring, collecting vast amounts of data including heart rate, sleep patterns, physical activity, and various biometric data. However, the reliability of the data is still a concern that affects clinical decision-making, user trust, and the overall utility of these devices [1]. The project aims to develop a robust system for evaluating wearable device data quality and communicating reliability data to both healthcare professionals and end-users through visualisation interfaces that are easily understood depending on which side it is being viewed to.

Introduction:

Studies [2] have shown substantial data quality issues across wearable devices. Research shows that data completeness can vary differently depending on the way data is recorded, with streaming approaches to devices experiencing up to 49% data loss compared to just 9% for onboard storage. Signal quality varies significantly by each metric, with blood volume pulse showing the lowest reliability at a mean of 60.2%, followed by electrodermal activity at 70.4%, while temperature measurements prove most reliable at 96.1% [2]. Consumer-grade wearables exhibit variable accuracy across different metrics, with heart rate measurements underestimating actual values by 7-9 beats per minute during exercise, step counts showing 9-24% error rates, and energy expenditure calculations displaying 6-43% error rates [2].

This proposed system will investigate the reliability of wearable device data through quality algorithmic assessment, to develop an effective method of presenting the calculated uncertainties to users. This will include a data quality assessment framework based on proven methods, which will be combined with an interactive dashboard featuring various uncertainty visualisation techniques.

The framework will evaluate data in terms of completeness, on-body detection, and signal quality. The dashboard will cater to both clinical and consumer audiences, adjusting its complexity according to the user's expertise. Insights gained from the technical implementation and user evaluation will help set up best practices for communicating reliability, assess the effectiveness of different uncertainty visualisation techniques, and contribute to the research on human-computer interaction in healthcare and the wearable device market.

Background:

Previous projects related to the proposed project include a study on data quality evaluation in wearable monitoring devices [3]. The paper discusses how to perform multi-dimensional quality assessments on wearable sensor data using the Empatica E4 device for epilepsy monitoring. It uses metrics like data completeness, on-body detection, and signal quality evaluation. The research showed that quality varies across different sensor types and recording methods. Specifically, streaming approaches experienced up to 49% of data loss, while onboard storage methods only lost about 9%. This difference shows a major weakness in real-time data transmission, which could affect the reliability of health monitoring systems [3].

The framework created by Nasseri et al. used advanced signal processing techniques specific to each sensor type. For the blood volume pulse data's quality assessment, researchers calculated spectral entropy to find periods of poor signal quality. The study revealed that blood volume pulse had the lowest reliability among all the modalities, with a quality score of only 60.2%. Electrodermal activity evaluation used rate of amplitude change analysis, achieving a reliability of 70.4% mean quality. Temperature measurements were the most reliable, reaching 96.1% mean quality, while accelerometry data was evaluated through threshold-based algorithms for on-body detection. The varied approaches of the validation findings, conducted across different sites with different patient populations, enhance the reliability of these findings [3].

Furthermore, research on uncertainty in health wearables also relates to this project. This work involved investigating the sources and implications of uncertainty in wearable device data, highlighting that uncertainty arises from sensor limitations, algorithmic processing, contextual factors, and individual physiological differences which lead to interpreting the data presented incorrectly. The research identified that healthcare professionals often lack training in interpreting wearable data, while consumers may over-rely on device metrics without understanding limitations [4].

As a result, this project would specifically investigate the application of uncertainty visualisation techniques to wearable health device data, rather than general uncertainty visualisation or wearable data presentation alone. It will enable information such as signal

quality scores, data completeness percentages, and reliability indicators to be communicated effectively, allowing for an analysis of how data transparency affects user trust and professional decision-making.

The Proposed Project:

The main aim of this project is to investigate the reliability of wearable device data through quality assessments and by using this data to develop visualisable communication strategies to allow for non-tech focused end users and also for clinical situations. In order to achieve this, the following objectives will need to be involved:

Creating a Data Quality Assessment Framework - A comprehensive framework will need to be developed to evaluate wearable device data across multiple different metrics, so that reliability can be quantitatively measured. The framework will implement algorithms for data completeness assessment, on-body detection, and signal quality assessments adapted from established methodologies developed by other researchers.

Designing Uncertainty Visualisation Techniques - Once the data quality assessment framework can evaluate data reliability, multiple visualisation approaches will be designed and implemented to communicate uncertainty properly and effectively. This will involve using metrics such as confidence intervals, color-coded quality indicators, density plots, and interactive time-series displays. The visualisations will be designed to serve both clinical professionals and consumer users, with adaptive complexity based on expertise level.

Building an Interactive Dashboard - An interactive web-based dashboard will be developed to integrate these components. The dashboard will display wearable device data alongside reliability indicators, allowing users to explore quality variations over time and across different sensor modalities. The interface will incorporate HCI principles on health data contexts [5].

Evaluating Effectiveness Through User Studies - When the dashboard is functional, user evaluation will be carried out in the form of task-based usability testing with 5-20 participants. This will assess whether users can correctly interpret uncertainty data, whether reliability communication affects trust, and whether the visualisations support

better decision-making. Both quantitative (task success, completion time, errors) and qualitative feedback will be collected and analysed at the end of this project.

Methodology:

The main software engineering approach used will be an Agile/Iterative method with rapid prototyping due to the need for user feedback during the design process. The project will be developed based on user feedback to test early prototypes.

For the evaluation, a sample of 5-20 participants representing target user bases (potentially including both individuals with wearable device experience and those with clinical backgrounds if accessible) will be recruited. Using 5-20 participants will allow for identification of usability issues and collection of feedback within the project timeframe, following standard usability testing practices. Participants must fill out the ethics forms, including participant information sheets, consent forms, and data collection forms, so that they understand the nature of the study and how their feedback and data will be used. To recruit study participants, advertisements such as posters and online notices will be distributed within Lancaster University and potentially through university health services connections and friends if available [5].

Due to the project goals, there will be two main technical phases of development: the framework implementation (the quality assessment algorithms), and the interface development (the visualisation dashboard). Additionally, there will be an evaluation phase involving user studies and feedback.

The framework implementation will involve developing Python-based algorithms to assess wearable device data quality. The metrics to be implemented will include:

Data completeness - Calculating the percentage of expected data samples that are recorded, showing gaps and missing data periods

On-body detection - Determining when the device is worn versus removed, using threshold-based algorithms on accelerometry, Electrodermal activity, and temperature signals

Signal quality assessment - Evaluating the reliability of physiological measurements using modality-specific approaches:

Blood volume pulse quality via spectral entropy calculation

Electrodermal activity quality via rate of amplitude change analysis

Temperature recording quality using rate of amplitude change and different threshold methods

Accelerometry quality via movement artifact detection

Aggregate reliability scoring - Combining individual metrics into overall reliability indicators for time periods, such as a score or different colour

Contextual metadata generation - Factors that affect data quality such as activity level, time of day, and device placement on the body

To ensure that the framework produces accurate quality assessments, validation will be performed using publicly available wearable device datasets where reference measurements exist, or through manual labeling of sample data periods. This will ensure that the algorithms produced correctly identify artifacts, data loss, and periods of unreliable measurements.

Programme of Work:

The project will begin late October 2025 (27th October), running until mid-March 2026. The project will be separated into the following stages:

Data Source Research and Requirements Analysis - This will involve conducting an overall review of research on wearable device reliability, data quality assessment methodologies, uncertainty visualisation methods, different datasets, and what data they provide, and HCI for health data interfaces. Requirements for the system will be defined based on gaps found in existing approaches and user needs which will be documented. A requirements specification document will be produced detailing metrics to be implemented, visualisation techniques to be made, and evaluation criteria to be employed. This will take around 2 weeks.

Framework Design and Algorithm Development - This will involve designing the quality assessment framework architecture and implementing core algorithms for data completeness assessment, on-body detection, signal quality evaluation, and aggregate reliability scoring based on established methods. The framework will be developed in Python and tested on sample wearable device data to confirm its functionality. This stage will take 5-6 weeks.

Visualisation Design and Prototyping - This will involve designing multiple uncertainty visualisation approaches including time-series plots with quality overlays, confidence intervals, quality indicator panels, and interactive filtering interfaces. Low-fidelity prototypes (wireframes and mockups) will be created for different visualisation techniques and reviewed for effectiveness before implementation. This will take 2-3 weeks.

Ethics Application Preparation - In parallel with visualisation design, the ethics application will be prepared in consultation with the supervisor (Ms. Emma Wilson). This will involve developing a detailed study plan including study design (aims, research design, procedure, and analysis plan), participant details (type of participants, target number of 5-20, age range 18+, and recruitment process through university channels), data handling plan (means of collection, maintenance of anonymity and confidentiality, secure storage using university-approved systems, and security measures), and compliant participant-facing documents (Participant Information Sheet and Consent Form). The application will be sent through the SCC UG Ethics Unified Application workflow as needed. This preparation will take 1-2 weeks and must be completed before any participant's recruitment can begin, as ethical approval is mandatory before recruiting participants or collecting data.

Mid-Year Progress Report and Refinements - During the Christmas break of approximately 2 weeks, a progress report will be written documenting work completed to date, preliminary findings for the framework validation, and plans for the evaluation phase. Any refinements to the quality assessment framework based on testing results will also be implemented.

Dashboard Development - This will involve developing the interactive web-based dashboard that integrates the quality assessment framework with designed visualisations. Backend development will use REST API creation, while frontend development will implement the user interface using modern JavaScript frameworks (React or Vue.js) and visualisation libraries (D3.js, Chart.js). The system will be tested internally to ensure functionality before user evaluation. This stage will take 4-5 weeks.

System Integration and Testing - This will involve integration testing to verify all components work correctly together, internal usability testing to identify key issues, and refinements based on testing results. During this phase, the ethics application will be finalized and submitted for supervisor approval. This stage will take 2 weeks, with ethics approval expected to be obtained by the end of this period.

Participant Recruitment - Once formal ethics approval has been granted through the SCC UG Ethics Unified Application, participant recruitment will begin. Recruitment will occur

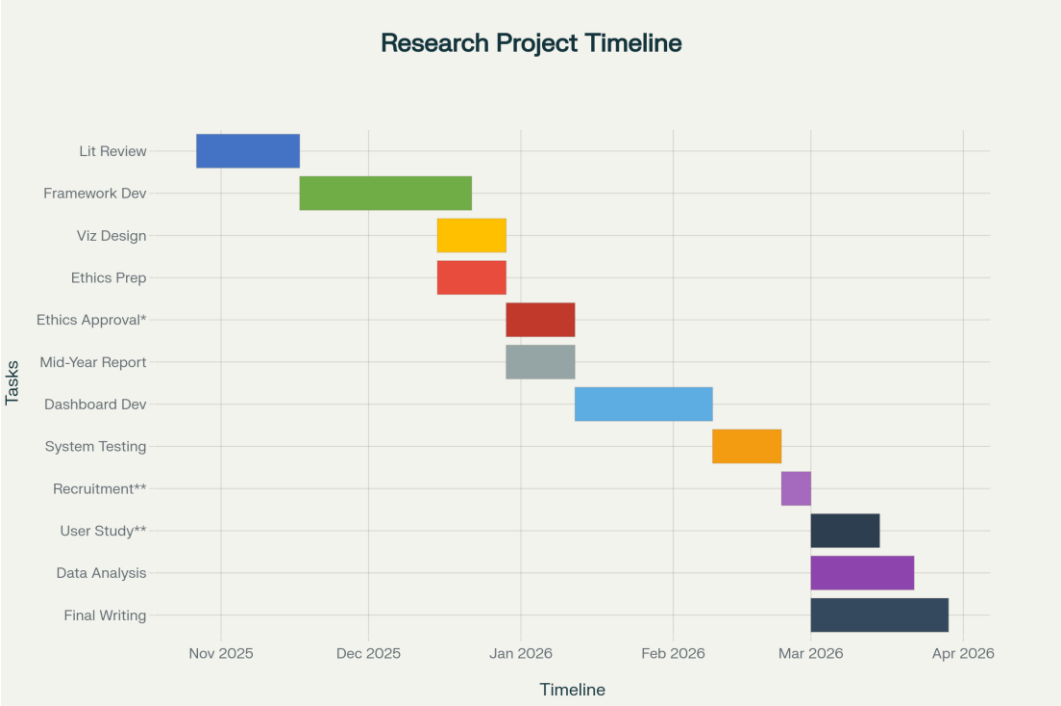
through department mailing lists, and potentially university participants and friends. Participants must fill out the proper ethics forms, including participant information sheets and consent forms, so that they understand the nature of the study, how their feedback will be used, their right to withdraw, and how their data will be stored and protected. Only after participants have provided consent can they be scheduled for evaluation sessions. It is essential that the study is conducted according to the approved procedure; any changes would require consultation and potentially a new ethics application. Recruitment will take 1-2 weeks to schedule 5-20 participants.

User Evaluation Study - This will involve conducting usability testing sessions with the recruited participants who have provided informed consent. Each session (30ish minutes) will include task-based scenarios, think-aloud protocols, and semi-structured interviews following the procedure approved in the ethics application. Participants will be given time to read the participant's information sheet and ask questions before beginning. At the end of each session, participants will have opportunities to ask follow-up questions, and they will be provided with supervisor contact information for any queries. Quantitative metrics (task success rates, completion times, errors) and qualitative feedback will be collected. All data will be anonymized and stored securely according to the approved data handling plan using university-approved storage systems. This will take 2 weeks for all evaluation sessions to be completed.

Data Analysis and Results Interpretation - Once the user studies have taken place, this stage of the project will be carried out to analyse collected metrics and qualitative feedback. Statistical analysis of task performance, calculation of usability scores, and analysis of interview data will be done. All analysis will maintain participant anonymity and confidentiality. The analyzed results will be interpreted to assess whether uncertainty communication improved comprehension and decision-making, and initial writing of results and discussion sections for the dissertation will begin. This will take around 3 weeks.

Final Dissertation Writing - The final weeks will be dedicated to completing the dissertation document, including all project phases, presenting results with appropriate visualisations, discussing implications, and drawing evidence-based conclusions. The dissertation will document that formal ethics approval was obtained prior to participant recruitment and data collection. Preparation for the final presentation will also occur during this period. This will take around 3-4 weeks .

Gantt Chart:



Resources Required:

Access to different wearable device datasets, e.g., Empatica E4, Physionet, MIMIC-III etc, Python with essential libraries such as:

NumPy - for numerical computations

SciPy - for signal processing (spectral entropy calculations)

Pandas - for data manipulation and analysis

SQLite or PostgreSQL - for data storage

Frontend Development framework e.g., Javascript with React or Vue.js, to produce interactive user interfaces, Chart.js or plotly.js for chart generation, and HTML5 and CSS for styling. Also needed would be a FetchAPI or Axios for communication with the backend, and considering mobile friendly frameworks e.g., Tailwind CSS/Bootstrap 5

Lastly for Visualisation, Figma for creating wireframes and prototypes during the design phase.

References:

[1] Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M.A., Simango, B., Buote, R., Heerden, D.V., Luan, H., Cullen, K., Slade, L. and Taylor, N.G.A. (2020). Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR mHealth and uHealth*, 8(9), p.e18694. doi:https://doi.org/10.2196/18694

[2] Shei, R.-J., Holder, I.G., Oumsang, A.S., Paris, B.A. and Paris, H.L. (2022). Wearable Activity Trackers–advanced Technology or Advanced marketing? *European Journal of Applied Physiology*, 122(9)

[3] Böttcher, S., Vieluf, S., Bruno, E., Joseph, B., Epitashvili, N., Biondi, A., Zabler, N., Glasstetter, M., Dümpelmann, M., Van Laerhoven, K., Nasseri, M., Brinkman, B.H., Richardson, M.P., Schulze-Bonhage, A. and Loddenkemper, T. (2022). Data quality evaluation in wearable monitoring. *Scientific Reports*, [online] 12(1), pp.1–16. doi: https://doi.org/10.1038/s41598-022-25949-x

[4] Knowles, B. (2018). Uncertainty in Current and Future Health Wearables – Communications of the ACM. *Acm.org*. [online] doi:https://doi.org/10.1145//3199201.

[5] Vasile Crudu (2025). *Top Usability Metrics & Methods for Evaluating Wearable Health Apps*. [online] Moldstud.com. Available at: https://moldstud.com/articles/p-top-usability-metrics-methods-for-evaluating-wearable-health-apps.

Todo list:

- Make test comparison between chestband data and wristband data
- Change weightings and how the accelerometer affects the end percentage data
- Cross reference EDA and temp to create new off body true/false variable
-