Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France – August, 2025
Paper No. ICBES 185
DOI: 10.11159/icbes25.185

REBECCA: A System for Real World Behavioural Data in Breast Cancer Clinical Research and Patient Care

Paraskevas Bourgos[1], Filopoimin Lykokanellos[1], Ioannis Sarafis[2], Alexandros Papadopoulos[2], Leonidas Alagialoglou[2], Vasileios Papapanagiotou[2,3], Aristotelis Ballas[4], Niki Kiriakidou[4], Christos Diou[4], Lazaros Apostolidis[5], Anna Barachanou[5], Symeon Papadopoulos[5], Nikolaos Androulidakis[6], Emmanouil Kafetzakis[6], Ioannis Giannoulakis[6], Anastasios Delopoulos[2], on behalf of the REBECCA consortium
[1]Netcompany S.A., 2b, rue Nicolas Bové, Luxembourg, Luxembourg
[2]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece
[3]Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden
[4]Department of Informatics and Telematics, Harokopio University of Athens, Greece
[5]Information Technologies Institute, Centre for Research and Technology Hellas, Greece
[6]Eight Bells Ltd Nicosia, Cyprus

Abstract - The REBECCA platform is a digital, cloud-based system designed to support clinical research and patient treatment in breast cancer patients (BCPs) through continuous, real-time monitoring of their behavioural, physical, and emotional well-being. It integrates heterogeneous data sources—wearable sensors, mobile applications, electronic questionnaires, web browsing behaviour, social media interactions, and electronic health records (EHRs)—into a unified and scalable infrastructure. At the core of the platform lies a robust system architecture that combines asynchronous communication via a message broker with distributed data streaming technologies, enabling high-frequency data ingestion and processing. Data from wearable devices and patient-reported outcomes are collected through a mobile application, while EHRs are securely transferred from REDCap servers located at clinical centres. All collected data are stored in secure cloud databases, forming datasets that reflects the patient's lifestyle, treatment journey, and psychosocial status. The platform features an interactive web-based clinical dashboard that allows healthcare professionals to assess individual patient trajectories, evaluate treatment and post treatment impacts, and derive actionable insights for care optimization and research. As a tool for research, it can complement traditional Randomised Control Trials (RCTs). The REBECCA system enables continuous monitoring, predictive modelling, and personalized decision support, integrating real-world data streams and supporting asynchronous data communication among analytic services. This paper presents the system architecture, data integration pipeline, and core functionalities of the REBECCA platform, highlighting its role as an enabler for next-generation clinical monitoring in oncology research and patient-centred care.

***Keywords*:** Real-World Data, Real-World Evidence, Clinical Monitoring, Breast Cancer Support, Real-time Monitoring, Electronic Health Records (EHRs), Predictive Modelling, Patient-Centred Care

1 INTRODUCTION

Modern healthcare is experiencing a transformative shift from traditional siloed systems to comprehensive integrated health monitoring platforms that synthesize diverse data streams into unified architectures. This evolution represents a fundamental advancement over previous single-modality approaches that have typically focused on isolated health aspects such as physiological data from wearables, patient-reported outcomes, or retrospective electronic health record analysis. Contemporary integrated platforms combine raw physiological data, mobile health applications, structured questionnaires, historical health records, and AI-generated indicators within single architectures [1][2]. These systems enable real-time aggregation and analysis of multimodal data types, facilitating synthesis of emotional wellbeing estimates, behavioural compliance metrics, and online activity indicators into comprehensive patient profiles [3]. The NeuroPredict platform exemplifies this approach by managing neurodegenerative diseases through monitoring both physical and mental health, leveraging artificial intelligence and Internet of Medical Things (IoMT) technologies to provide predictive models and early-warning services for clinicians [9][12]. Advanced AI and machine learning techniques significantly amplify platform innovation capabilities. Recent studies demonstrate that AI-driven analytics extract meaningful features from multimodal data including mood, sleep, activity, and social engagement patterns, enabling health risk prediction and proactive personalized recommendations [6]. Transformer-based approaches and deep neural networks forecast emotional states and detect behavioural instabilities in real-time, improving clinical intervention accuracy and timeliness [3]. AI models demonstrate high accuracy in predicting mental health deterioration in adolescents, identifying early warning signs from both self-reported and passively collected data [4][6]. Mobile health applications serve as central components facilitating collection of passive sensor data and active patient-reported outcomes at scale [10]. Recent studies highlight mHealth app effectiveness in reducing patient engagement barriers, improving compliance, and enabling large-scale geographically diverse recruitment for clinical trials and routine care [5][10]. These applications support rapid structured collection of adverse event data and compliance metrics, enriching datasets available for analysis and decision-making. Integrated platforms differentiate themselves through intuitive web interfaces and clinician-facing visualizations that synthesize compliance indices, early-warning alerts, and behavioural trends [7][11]. These dashboards support rapid evidence-based clinical decisions, contrasting with earlier systems providing only retrospective or fragmented insights [7]. Modern platforms offer dynamic real-time monitoring and risk prediction, significantly enhancing patient outcomes and healthcare efficiency [11]. The principal innovation lies in holistic integration of raw data, mobile applications, questionnaires, historical data, AI-generated indicators, early-warning services, emotional and behavioural analytics, compliance metrics, and clinician-oriented visualizations [1][8]. This comprehensive approach advances personalized and preventive healthcare while addressing longstanding challenges of data fragmentation, limited patient engagement, and reactive clinical workflows.

2 SYSTEM OVERVIEW AND ARCHITECTURE

REBECCA creates a digital cloud platform for monitoring breast cancer patient (BCPs) continuously through multiple data sources. The system collects data from wearable devices via a smartphone app, electronic questionnaires, web browsing history, social media interactions, and electronic health records from clinical centres. All collected data resides in cloud databases for analysis, providing insights into patients' lifestyle, behaviour, medical history, and changes in functional/emotional status. The platform employs advanced analytics including indicator extraction algorithms, machine learning, and causal analysis models to transform raw data into actionable information covering: (a) Behavioural and Physiological indicators, (b) Online behavioural indicators, (c) EHR indicators, (d) Environmental indicators, (e) Data quality assessment, (f) Causal inference, and (g) Decision support. Healthcare professionals can access this comprehensive information through a clinical dashboard interface. The REBECCA system ensures secure authentication across all user-facing components, including the Clinical Dashboard, the REBECCA app, and Browser Plugin. Clinicians manage patients through a clinical dashboard, while the patient app collects data via GPS, Garmin™ wearables, photos, and questionnaires, on both Android and iOS. Companion users, designed by the patient (e.g., spouse, friends) can provide additional information for the patient. The browser plugin tracks online behaviour and sentiment and offers visualisations. CI/CD pipelines support testing and deployment across clinical centres. REBECCA's Early Warning System is integrated with the Clinical Dashboard to provide timely warnings based on patient activity trends.

2.1 Architecture Overview

The REBECCA system architecture is designed as a secure, modular, and cloud-based infrastructure that supports scalable digital health services for BCPs. It integrates multimodal real-world data (RWD) into a cohesive ecosystem while ensuring privacy, secure communication, and high system availability. At the edge, clinical site systems—including EHR platforms and data anonymisation modules—send data securely via HTTPS to the REBECCA Load Balancer. Patient data is collected through the smartphone app, and browser plugin, which interact with the platform using encrypted HTTPS channels. The Load Balancer ensures secure traffic routing, forwarding requests to relevant services while interfacing with key components such as the Authentication Endpoint, Data Encryption modules, and Platform Monitoring tools. Central to the REBECCA Cloud Services are the microservices deployments, which are managed through a dedicated cloud infrastructure and organized under an API Gateway. These microservices include controllers for different data sources and interfaces: the REDCap Synchronisation Service, Clinical Dashboard Controller, Browser Plugin Controller, and the REBECCA App Cloud Controllers. These services communicate with two main data storage entities: the NoSQL REBECCA

Operational Database and the Sensor Data Cloud Repository, both of which use replicated architectures to ensure fault tolerance and data redundancy. The Data Broker facilitates asynchronous communication across services, enabling real-time and decoupled data exchange between components. Built on Apache Kafka, this messaging layer supports a highly responsive and scalable system environment. Downstream, the Data Analysis Microservices Deployment layer houses specialized analytical engines. These are grouped into two main categories: indicator extraction modules (e.g., behavioural, emotional, EHR, environmental) and decision-support services (e.g., data quality assessment, causal inference, and decision support modules). These services operate on the data collected and stored in the cloud, and each has a dedicated REST API, processing logic, and a database to store intermediate or final results. Finally, the REBECCA Clinical Dashboard interfaces with these analysis modules to deliver visualizations and insights to healthcare professionals, closing the loop from data ingestion to decision-making support.

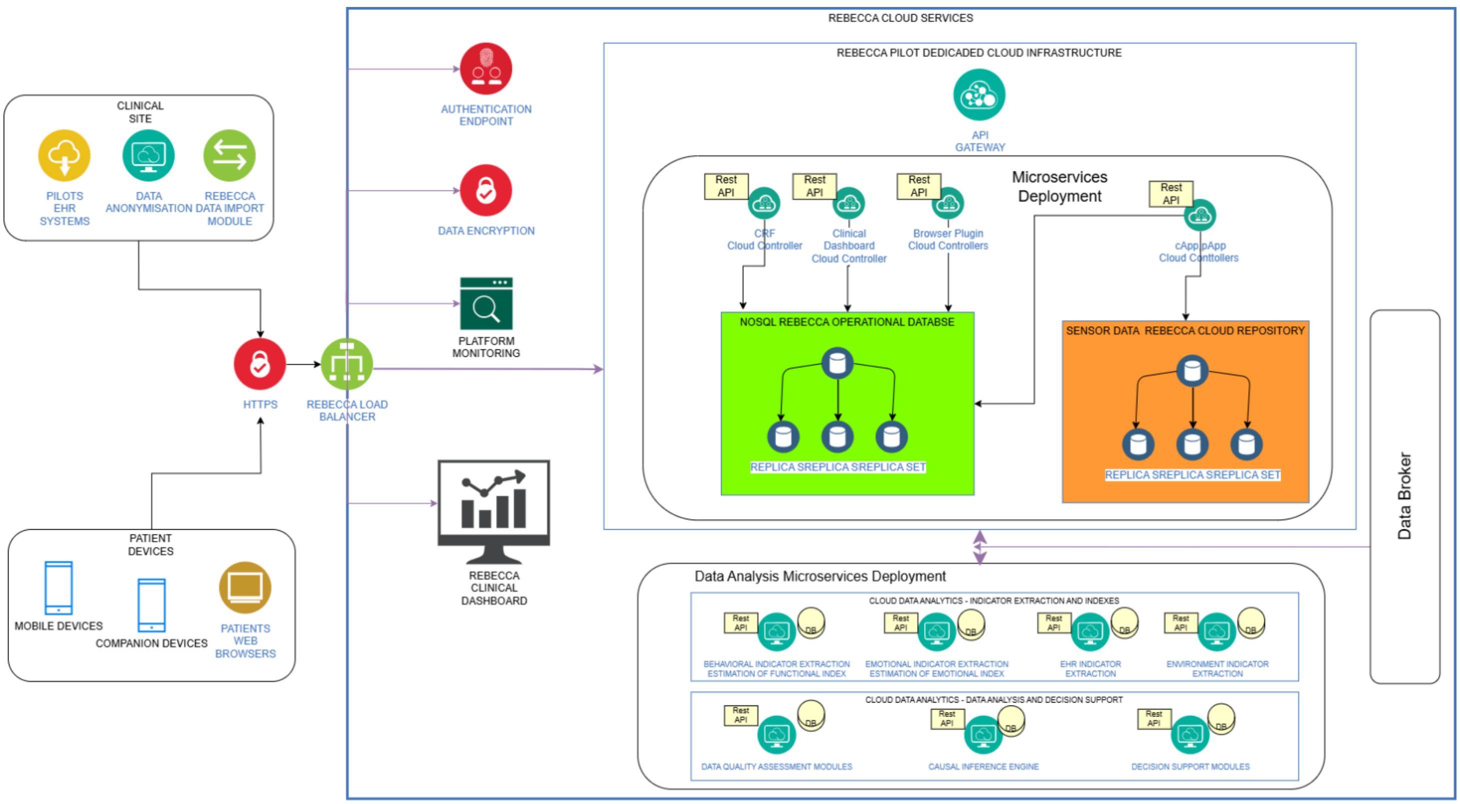


Figure 1 REBECCA System Architecture

3 SYSTEM COMPONENTS

3.1 The REBECCA mobile app

The REBECCA mobile app serves as the central tool for patient monitoring in REBECCA. Developed in Flutter for Android and iOS, the app features two modes of operation depending on the user type: a) **pApp mode** – for data collection from the BCPs, and b) **cApp mode** – for data collection from a BCP's companion.

In **pApp** mode, the app primary collects continuous, passive, and unobtrusive data from BCPs. This includes a) behaviour and lifestyle data from Garmin™ wearable devices; b) continuous background location data. The wearable data includes per minute measurements of steps, stress level, heart rate, resting heart rate, distance travelled, and sleep sessions. The location data are captured with a frequency of at least one sample per minute and are used to analyse patients' mobility patterns with known links to functional status, such as mobility diversity, radius of gyration, and time spent and home and work. Collection of self-reported data is also supported in pApp mode, in the form of user-annotated photos (meals, drinks, and life events – both positive and negative), as well as short in-app questionnaires inspired by standardised Patient Reported Outcomes Measures (PROMs). The figure below shows example screens of the REBECCA app in pApp mode.

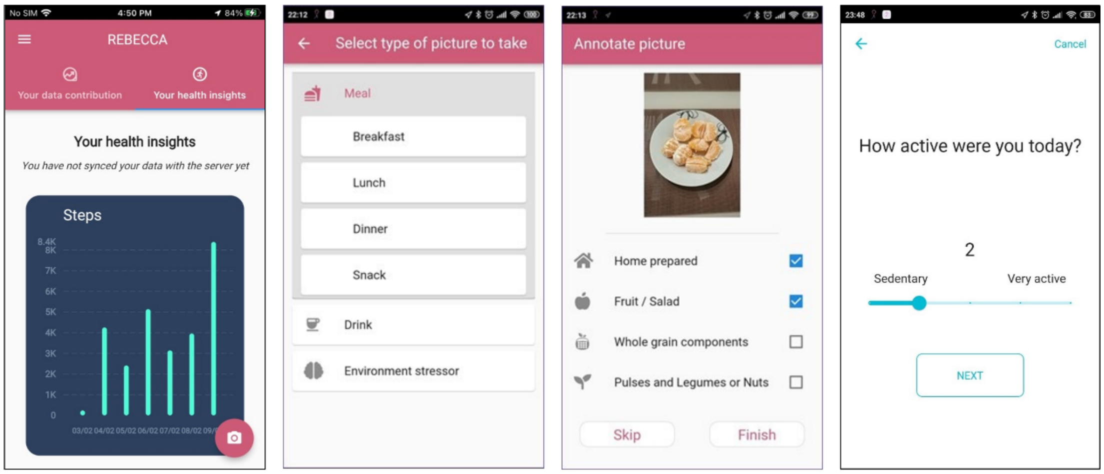


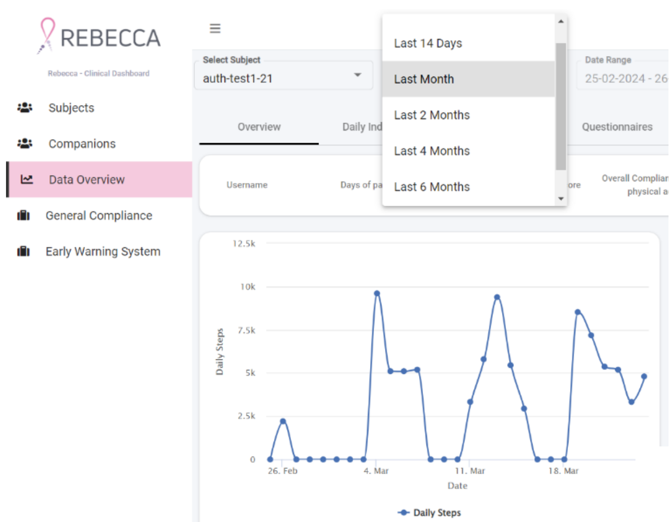
Figure 2 Example screens of REBECCA app; from left to right: user feedback, photo capture, picture annotation, and in-app questionnaires

In **cApp** mode, data collected related to the BCP's functional status are collected by an optional companion designated by the BCP. The companion is typically a person close to the BCP, such as caregiver, a family member, or friend, and he/she is responsible for regularly completing a short in-app questionnaire. Thus, the companion acts as an independent information source for monitoring BCP's condition. To support compliance, companion users receive reminder notifications to complete the questionnaire every two weeks.

3.2 Clinical Dashboard

The Clinical Dashboard (CD) is used by the REBECCA clinicians and scientific personnel to monitor data and information over the available clinical trials. The CD functionalities are framed under a secure environment which ensures authorized access to the provided information concerning the breast cancer patients. Dedicated CDs have been established for each clinical centre with a corresponding set of user accounts created accordingly.

Through the dashboard's secure environment, the personnel (researchers, doctors, study nurses) can perform the



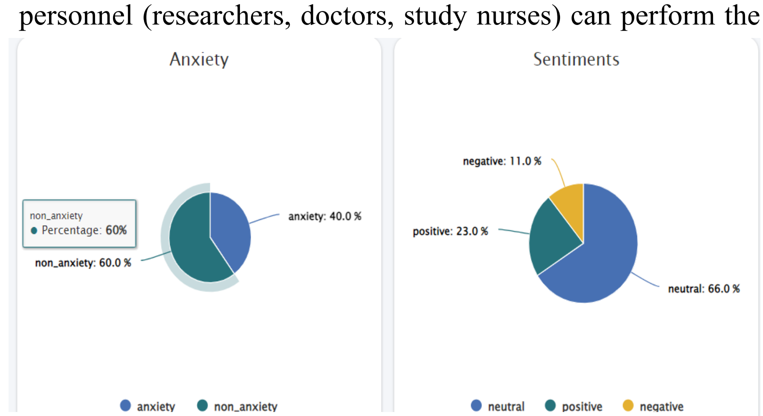


Figure 3 (left) Frame in a Specific Time Range and (right) Online Activity account administration for patient and companion accounts (inspect active accounts, add new, or remove) and inspect data, analytics, and early warnings. A subset of the CD's functionalities are illustrated in Figure 3.

3.3 Browser Plugin

As part of the REBECCA platform, a custom browser plugin was developed to monitor and collect online activity data from participating BCPs. The plugin captures a range of online activity data from the web, including website visits, search queries, and interactions on popular platforms such as Facebook and YouTube (e.g. posts and comments). It operates silently in the background with minimal user intervention, ensuring a seamless and non-intrusive user experience that aligns with ethical standards and privacy expectations. The plugin is easy to install through the Chrome Web Store[1] and is designed with a modular architecture, comprising two core components: the User Interface (UI) and the Backend Service.

The **User Interface** allows patients to: (a) Customize monitoring preferences, such as pausing tracking or excluding specific websites; (b) View collected data records (e.g., visited websites, search queries, social media interactions); (c) Delete individual records they are uncomfortable sharing. The UI also includes a **self-reflective dashboard**, accessible via the plugin's options page. This interactive interface visualizes the user's online behaviour indicators and emotional insights using metrics such as: (a) Time spent online and session frequency; (b) Activity distribution across days and hours; (c) Topic classification of accessed content; (d) Sentiment, stress, anxiety and emotional trend analysis. All visualizations are filterable by custom timeframes, allowing users to observe behavioural changes over specific periods.

The Backend Service is responsible for: (a) Secure transmission of online activity data from the plugin to the REBECCA server; (b) Storage of pseudonymized data in a MongoDB instance; (c) RESTful communication between the plugin and the backend API. All data transmissions are encrypted, and no personally identifiable information is collected.

3.4 REDCap Synchronisation Service

In the REBECCA system, REDCap Synchronisation Service is responsible for periodically synchronising data recorded in the REDCap installations at the clinical centres, such as medication, blood exams, Patient Reported Outcome Measures (PROMs). The service transfers ~~stores~~ the data in a MongoDB, which offers a flexible schema for handling diverse REDCap instruments. In addition, it provides an API for making the data accessible to other REBECCA components.

3.5 Data Analysis Modules

The data analysis modules of REBECCA serve two main purposes. First, to consolidate all data collected from REBECCA app, wearable and browser plugin and use them to extract behavioural, physiological and online behaviour indicators. This function is performed out by the **Behavioural Indicator Engine (BIE)** and the **Online Behaviour Indicator Engine (OBIE)**. Second, to detect any deterioration of BCP's status over time, using the extracted indicators. This function is carried out by the **Early Warning System (EWS)**.

3.5.1 Behavioural Indicator Engine

The Behavioural Indicator Engine (BIE) is cloud service implementing the indicator extraction algorithms. The indicators are quantities that accurately and objectively describe parts of the BCP's behaviour and physical functioning status in their everyday life and are an essential component of real-world data that can help to assess the QoL of BCPs. The BIE extracts daily indicator for each patient in the following categories:

- Physical activity and psychological indicators recorded by the smartwatch which include daily steps, daily calories burned (active and total), mean heart rate, resting heart rate, stress levels.
- Sleep indicators for detected sleep sessions which include total sleep duration, number of wakeups, and time spent in awake, light, REM and deep sleeps.

[1] https://chromewebstore.google.com/detail/rebecca/mekkepnpeljplmcncdnhmcmlcaljpekd

• Mobility indicators which are calculated using the continuous GPS signal. The Points of Interest of the patient are calculated (i.e., places that have been visited) and the home and work locations are also determined. Based on this, time spent at home or work is calculated. In addition, the Radius of Gyration [13] and Mobility Diversity [13] are extracted from the GPS signal.

• Self-reported PROMs which are the answers to short in-app questionnaire.

3.5.2 Online Behaviour Indicator Engine

Similar to BIE the Online Behaviour Indicator Engine (OBIE) is a cloud service implementing the online behaviour extraction algorithms. It is designed to extract indicators from the patient's online activity. It complements the BIE by focusing on cognitive and emotional aspects inferred from the user's interactions on social media platforms such as Google, YouTube, and Facebook. These indicators provide valuable insights into the mental and emotional state of BCPs, contributing to a comprehensive understanding of their real-world behaviour and QoL.

OBIE processes textual and metadata elements from online activity logs (e.g., search queries, video titles, posts, comments, URL visits) to extract the following indicators:

- Topics of Interest: Recurring themes and subjects (e.g. health, sports, etc.), enabling assessment of shifting interests.
- Emotions: Text-based emotion recognition into emotional classes such as joy, sadness, anger, worry, neutral, etc.
- Sentiment: Positive or negative tone of online interactions, offering a general measure of user outlook and mood.
- Stress: Linguistic patterns linked to tension, pressure, or feeling overwhelmed are ~~analyzed~~analysed to detect levels of stress.
- Anxiety: Content suggestive of worry, fear, or anticipatory concern, capturing changes in mental well-being over time.

3.5.3 Early Warning System

An Early Warning System (EWS) can be defined as a set of functionalities needed to generate and disseminate timely and meaningful information of possible extreme events [14]. In REBECCA, the EWS is implemented for the timely notification of medical practitioners regarding the deterioration of a BCP's wellbeing or the level of compliance in a clinical study. Specifically, the REBECCA EWS acts as a: (a) data logging and a data compliance calculation service and (b) behaviour monitoring service. Regarding (a), it calculates the level of compliance of a participant for each respective component in REBECCA (mobile applications, browser plugin and, smart watch) and returns a compliance index. The compliance index is a float spanning from 0 to 1 and represents the level of a patient's contributed data to the system.

The behaviour monitoring service on the other hand, is tasked with alerting the clinicians when a participant's behaviour deteriorates. The service calculates the Exponentially Weighted Moving Average (EWMA) of selected behavioural indicators and then checks if the value of said indicator is below 15% of the average. As EWMA takes into consideration the historical

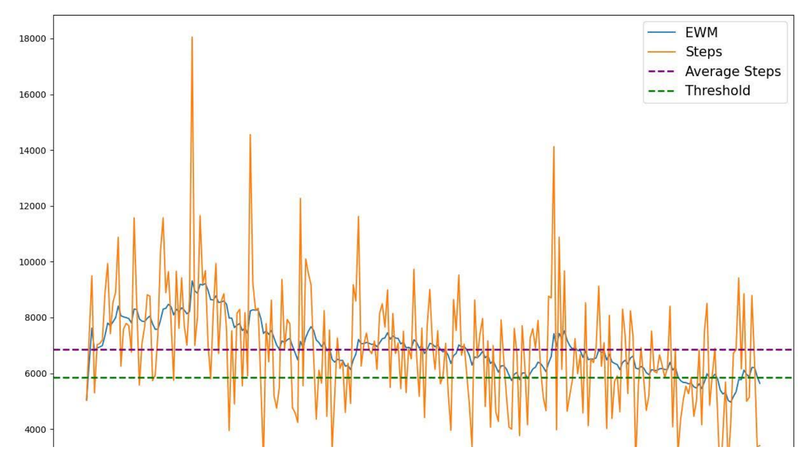
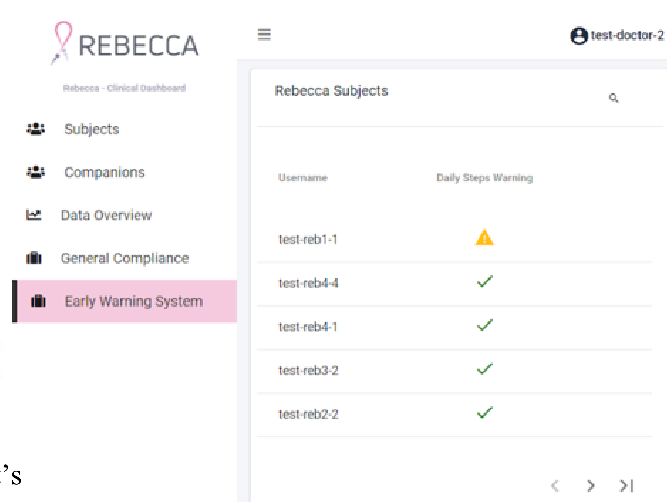


Figure 5 An example of the calculated EWM on a participant's daily steps.



behaviour of each participant, it is a more robust way to measure the level of change, compared with a flat threshold. A concrete example of the above functionality can be found in monitoring a participant's daily steps. We plot the "steps" time-series of a patient (Orange line) and Calculate the Exponential Weighted Moving Average of the series (Blue line).

A warning is produced when the total steps of a participant begin to deteriorate as depicted in Figure 5. In addition to these indicators, the EWS also monitors the answers participants have filled in the daily or weekly questionnaires available in the REBECCA mobile app and notify the clinicians if there is a deterioration reflected in their responses Figure 4.

4 SECURITY AND PRIVACY FRAMEWORK

The REBECCA platform implements a comprehensive Security and Privacy Framework encompassing data protection, authentication, and secure communication. Data exchanged between user devices and servers is encrypted using SSL/TLS/HTTPS, with storage secured via LUKS encryption, strict firewall rules, SSH key-based access, and protected GitHub branches. For authentication and access management, Keycloak provides centralized single sign-on, identity integration, and fine-grained access control, with dedicated endpoints per clinic to ensure data segregation. Additionally, Docker overlay networks with encrypted IPSEC tunnels support secure, distributed deployment of components, with separate virtual machines and storage systems per clinical site to maintain data privacy and isolation. The REBECCA platform adopts modern software engineering practices with a strong focus on automation and quality assurance. An integration matrix developed using a Design Structure Matrix (DSM) approach, maps component interactions based on system architecture to ensure precise interface definition and minimize incompatibilities. A robust CI/CD stack supports standardized, automated development and release workflows, aligning with DevSecOps principles.

5 SYSTEM DEPLOYMENT AND DATA COLLECTION

Development of the REBECCA system began in September 2021, with the first production-ready version delivered in December 2022. Patient enrolment started in February 2023. Since then, REBECCA system has supported multiple clinical studies and collected substantial amounts of RWD. The table below provides details on the *six* studies – completed and on-going – and the amount of data collected using REBECCA system as of the time of writing (May 6th, 2025).

Table 1 Overview of the studies conducted using the REBECCA system and the volume of data collected across the different modalities

Study description and pilot site	No. Days with Behavioural/ Physiological indicators	No. Days with Mobility Indicators	No. Days with Sleep Indicators	No. Self-reported questionnaires	No. Self-annotated photos captured
Observational, with focus on osteopenia/osteoporosis (Spain)	15,777	14,098	1,970	19,157	7,515
RCT studying REBECCA as intervention (Spain)	3,695	4,039	466	3,898	2,888
Observational, with focus cancer treatment-related fatigue (Norway)	13,270	18,108	6,440	4,291	9,383
RCT studying REBECCA as intervention (Norway)	9,203	7,077	1,767	1,331	9,322
Feasibility study: REBECCA for prostate patients (Norway)	1,943	1,747	412	281	439
Feasibility study: personalized lifestyle counselling (Sweden)	4,846	4,064	942	3,299	14,534

6 CONCLUSION

In this paper, we presented the REBECCA system which is a novel platform developed to support research and interventions in breast cancer. The system collects multimodal RWD, including EHR, behavioural and physiological data from wearables and the smartphone app, and online behaviour data via a web browser plugin. These data are analysed to generate behavioural aggregations and visualisations in the system's Clinical Dashboard. Overall, the goals of REBECCA system are to support clinical research and serve as a complement to RCTs, as well as to improve patient care though personalisation and prevent health deterioration through timely early warnings. To this end, we envision the system having a significant impact on studying and treating complex chronic conditions where access to real-world evidence is essential.

ACKNOWLEDMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 965231.

References

[1] C. Peng, P. Goswami, and G. Bai, "A literature review of current technologies on health data integration for patient-centered health management," *Health Informatics Journal*, vol. 26, no. 3, pp. 1926–1951, Sep. 2020.

[2] Technology Innovators, "Internet of Medical Things (IoMT): Connected Devices and Healthcare Monitoring," *Technology Innovators*, Aug. 2023.

[3] Z. Yang, A. Mitra, W. Liu, D. Berlowitz, and H. Yu, "TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records," Nature Communications, vol. 14, no. 1, article 7857, Nov. 2023.

[4] Contents.ai, "Harnessing Artificial Intelligence: A New Frontier in Adolescent Stress Management and Mental Well-being," *Contents.ai Magazine*, Nov. 2024.

[5] American Heart Association, "Assessing Patient Engagement With a Mobile Health Application for Self-Care Of Heart Failure in a Real-World Setting," *Circulation*, vol. 146, suppl. 1, Nov. 2022.

[6] E. Hill, "AI predicts adolescent mental health risk before symptoms emerge," *PsyPost*, Mar. 2025.

[7] M. A. Badgeley, K. Shameer, B. S. Glicksberg, M. S. Tomlinson, M. A. Levin, P. J. McCormick, A. Kasarskis, D. L. Reich, and J. T. Dudley, "EHDViz: clinical dashboard development using open-source technologies," *BMJ Open*, vol. 6, no. 3, e010579, Mar. 2016.

[8] C. Peng, P. Goswami, "Meaningful Integration of Data from Heterogeneous Health Services and Home Environment Based on Ontology". *Sensors* (Basel), vol. 19, no. 8, art. 1747, Apr. 2019.

[9] M. Ianculescu, C. Petean, V. Sandulescu, A. Alexandru, A.M. Vasilevschi. "Early Detection of Parkinson's Disease Using AI Techniques and Image Analysis". *Diagnostics* (Basel), vol. 14, no. 11, p. 2615, Nov. 2024.

[10] R. Monachelli, S. W. Davis, A. Barnard, M. Longmire, J. P. Docherty, and I. Oakley-Girvan, "Designing mHealth Apps to Incorporate Evidence-Based Techniques for Prolonging User Engagement," Interact. J. Med. Res., vol. 13, e51974, Mar. 26, 2024.

[11] M. Wang, S. Li, T. Zheng, N. Li, Q. Shi, X. Zhuo, R. Ding, and Y. Huang, "Big Data Health Care Platform With Multisource Heterogeneous Data Integration and Massive High-Dimensional Data Governance for Large Hospitals: Design, Development, and Application," *JMIR Medical Informatics*, vol. 10, no. 4, e36481, Apr. 13, 2022.

[12] M. Schinle, M. Dietrich, S. Stock, M. Gerdes, and W. Stork, "Model-Driven Dementia Prevention and Intervention Platform," Stud. *Health Technol. Inform.*, vol. 302, pp. 937–941, May 18, 2023.

[13] L. Pappalardo, Z. Smoreda, D. Pedreschi, and F. Giannotti, "Using big data to study the link between human mobility and socio-economic development." *2015 IEEE international conference on big data (big data)*. IEEE, 2015.

[14] J. E. Quansah, B. Engel, and G. L. Rochon, "Early warning systems: a review", Journal of Terrestrial Observation, vol. 2, no. 2, p. 5, 2010.