Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240391

Using One App Only – Collecting a Comprehensive Set of Health-Related Data for Prevention of Chronic Conditions

Eirik ÅRSAND[a,1], Miroslav MUZNY[a,b], Tina RISHAUG[a], Anna M. WÄGNER[g], Carmelo BETANCORT[g], Conceição GRANJA[b,c], Cristina SOGUERO-RUIZ[d], Gunnar HARTVIGSEN[a], Mihai NAMOLOSANU[e], Mikael RINNETMÄKI[f], André HENRIKSEN[a] and WARIFA CONSORTIUM[h]
[a] *UiT The Arctic University of Norway, Tromsø, Norway*
[b] *Norwegian Centre for E-health Research, Tromsø, Norway*
[c] *Nord University, Faculty of Nursing and Health Sciences, Bodø, Norway*
[d] *Rey Juan Carlos University Fuenlabrada, Madrid, Spain*
[e] *NetSun Services SRL, Bucharest, Romania*
[f] *Sensotrend Oy, Tampere, Finland*
[g] *Complejo Hospitalario Univ. Insular Materno-Infantil de Gran Canaria, Spain*
[h] *See Acknowledgements, Tromsø, Norway*
ORCiD ID: Eirik Årsand https://orcid.org/0000-0002-9520-1408

Abstract. This paper presents the design, implementation and early tests of an app that collects a comprehensive set of health-related data, as part of the EU-project WARIFA. To achieve the main aim of the project – using AI to prevent chronic conditions – a wide range of data needs to be collected and stored at a backend server for processing. The methods and elements for creating this system are presented, as well as results from the co-creation process and early user-tests. Challenges regarding complexity, security, and privacy are discussed, as well as the needs and prospectives for easier ways of collecting health-related data.

Keywords. data collection, mHealth, eHealth, disease detection, AI

1. Introduction

Health apps are typically focusing on a few parameters and a single health challenge or disease. The western population is aging, environmental factors are affecting our health negatively, and health challenges are getting more comprehensive and complex. Systems that handle many parameters at once might help prevent health issues. The presented work describes an app designed to collect a comprehensive set of data, as part of a complete system where users receive feedback based on these data. The work is part of the project *Watching the risk factors: Artificial intelligence and the prevention of chronic conditions* (WARIFA), funded by the European Union's (EU) Horizon 2020 research and innovation programme under grant agreement No. 101017385 (Article 29.1 GA).

Non-communicable diseases (NCDs) such as cardiovascular diseases, cancers, diabetes, and chronic respiratory disease are leading causes of death in the EU [1]. These NCDs, except type 1 diabetes (T1D), are associated with lifestyle factors and can be prevented. Most Europeans have access to mobile devices like smartphones and smartwatches, collecting a

[1] Corresponding Author: Eirik Årsand; E-mail: eirik.arsand@uit.no.

variety of health data, such as physical activity, sleep, and heart rate. There is an increasing emphasis on research and innovation in artificial intelligence (AI) in the EU [2], from which the health sector could benefit greatly. When applying AI to disease prevention and management, comprehensive data collection is crucial.

2. Methods

The development of the app took place as a co-creation process with end-users, involving professionals and clinical experts (i.e., dietitians, psychologists, and physicians). Fourteen focus groups were conducted, involving 3-12 end-users per group. Participants assessed the app prototype and provided feedback. The feedback was shared with the project's technical experts and implemented. We used an agile iterative development process where new versions of the app were tested in new focus groups. Totally, 112 adults participated, including representatives from the general population and people with T1D.

To detect bugs, technical issues, and other problems within the app, several testing sessions were performed on five different smartphones (Samsung Galaxy A53 5G, Galaxy Z Fold 4, OnePlus 10 Pro 5G, Sony XQ-AT51, and Google Pixel 7 Pro). Elements tested were functionalities such as settings, home screen, data registering, questionnaires, and inclusion of UV/air pollution data. Retrospective testing, battery consumption, and navigation were investigated on all phones, including testing for negative scenarios, e.g., disabling certain functionalities to check for consistency.

As a fundamental step to comply with the General Data Protection Regulation (GDPR) [3], risk assessments were conducted for: 1) the WARIFA app; 2) welfare devices that use a cloud service for data transfer; 3) welfare devices that directly transfer data to the WARIFA app on a mobile phone; and 4) the questionnaire platform. A qualitative approach, based on ISO 27005 standard [4], was used to identify and assess risks, categorizing them according to their severity and likelihood of occurrence, in collaboration with the Chief Information Security Officer at the University Hospital of North Norway. The project was found exempt from the scope of the reginal ethical committee REC north, ref. no. 437963.

3. Results

We present the main elements of a comprehensive data-collection system, consisting of a mobile application and an associated server system. Evaluation of the system is presented in form of technical tests on different platforms, and user-feedback from focus group meetings. It is exemplified how the collected data can be utilized to identify critical risk factors using AI-based models.

3.1. Data collection

The data collection from the developed app is based on multiple methods and data sources, such as sensors, online accounts, public databases, questionnaires, and direct questions. Figure 1 presents some of the variables collected from the app.

Sensors and online accounts: Studies show that automatic registration of health data is preferred [5], therefore, sensor access was implemented for the health parameters that were possible to record automatically and integrate into the system. This includes location data from the built-in navigation features of mobile phones, physical activity sensors, and Continuous Glucose Monitors (CGMs), to capture data such as heart rate, hours of sleep, number of steps, and blood glucose values. The data are collected from the sensors via

Bluetooth and transferred to the manufacturers' own app and the users' online accounts on the accompanying cloud services. From there, our app can download the data connected to these accounts, both through application programming interface (API) integrations and by a person downloading their data as a file and uploading it to a backend service, i.e., the Sensotrend Connect service (Sensotrend Oy, Tampere, Finland) for glucose data. The user may also connect to Google Fit through the app and use other sensors and devices that can transfer data to this cloud service, e.g., Fitbit devices.

Public databases: OpenWeatherMap.org and Yr.no, are used to retrieve weather and environmental data, e.g., air pollution and ultraviolet (UV) index. Air quality index (AQI) has been chosen as the pollution parameter, and UV index can determine the user's historic and real-time sun exposure, at the user's location.

Questionnaires: The questionnaire feature is designed to collect information on health-related parameters not addressed by the previous methods, such as user demographics and disease-specific details. It facilitates gathering of information at predetermined intervals, based on user preferences, such as favored reminder times and frequency. The questionnaires are managed through the NetSun (NetSun Services SRL, Bucharest, Romania) questionnaire engine, which operates as a component on the backend solution.

Direct questions: Direct questions are questions not included in questionnaires but are required to set up the app or tailor its usage, e.g., preferences of connected health sensors, location settings, and parameters to monitor. They may also appear as notifications at given intervals, prompting users to adjust their settings to increase their motivation and engagement with the functions of the app.

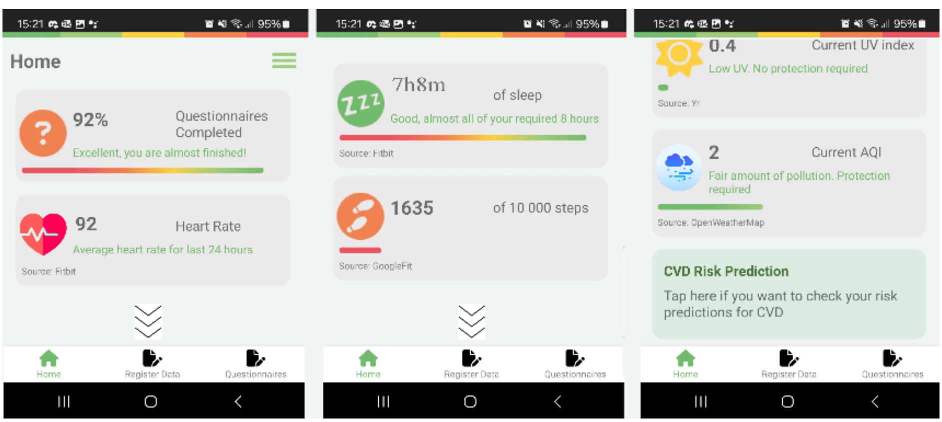

Figure 1. Some variables presented in the app, and some graphical feedback elements.

3.2. Backend system

The backend system is divided into two main areas – publicly accessible area and secure area. The secure area is used for storage and processing of sensitive data, while the publicly accessible area is used for communication with the data collection app.

Storage: The storage area is divided into the following three types – the Questionnaire Platform, the HL7 FHIR-based Research Data Management Server, and the Diabetes Data Collection platform. The underlaying infrastructure is built on top of self-managed servers hosted using the VMWare virtualization technology i.e., not using public cloud providers.

Communication: The communication between the app and the backend system is done through a custom REST API. Push notifications are delivered by the Firebase service. Researchers can query data in the secure area through a self-developed Python library.
Security: To mitigate the risk to confidentiality, measures such as user training, encryption, and authentication have been implemented. To address the risk to integrity, user awareness, authentication, and validation of input data, have been employed. Availability risks will be mitigated through user education, notification mechanisms, and regular compatibility checks.
Data processing: Machine learning models and algorithms for feedback generation and personalization use the collected data to provide feedback that is presented in the app. System requirements, including memory and CPU/GPU utilization, were considered in the infrastructure design. These algorithms are orchestrated by a Kubernetes engine, to ensure horizontal scalability and meet increasing demands, e.g., higher number of users or data processing needs.

3.3. Calculations and predictions based on the data

Using the data collected by the app, AI-based models can monitor risk indicators and provide preventive recommendations over time for the various NCD targets. The system enables users with T1D who uses a CGM sensor to see a prediction of their blood glucose levels for the upcoming 30 and 60 minutes. Similar AI-based models can be implemented and be a powerful tool in identifying critical risk factors based on the data collected from the app. This approach may be especially helpful in assessing the risk of NCDs, making it possible to identify people who would benefit from preventative care. These models can also assist in making decisions about intensifying treatments of TD1 and preventing acute events.

3.4. User- and performance testing

During the first focus groups, end-users sketched the potential app interface. An interface was designed and shown to the next set of focus groups. Functionalities were added and tested in an ongoing, iterative process, which involved end-users, clinicians, and developers. Input from the users focused on security, usability and need for interaction, visualization of progress and feedback. Feedback from the last focus group meeting stated that 4/10 participants use or have used a health app, such as: Adidas Running, Garmin Connect, Zepp, and Fitness GYM. All 10 participants preferred having all their health-related information in a single app.

Various issues were detected in the performance tests, including language errors, missing health data, missing weather data, design errors, and questionnaire-related issues. The tests and improvements were done in several rounds, where the biggest challenges were when connecting the app to external services such as Fitbit and Google Fit.

4. Discussion

Making a comprehensive data collection system is complex due to many dependencies, different sensor data standards – like we used Logic Observation Identifiers Names and Codes (LOINC) for observation coding in registrations from the app, external APIs, security, and privacy rules, and establishing seamless communication between the app and the back-end system. The Systematized Nomenclature of Medicine (SNOMED) terminology is not used, as non-numeric data are handled by the questionnaire engine. Performance testing is expected needed, as long as new services are added and new functionalities implemented. Encouraging users to use an app in the long-term to gather representative data, is challenging – as well as making the system user-friendly enough to facilitate usage. However, if we succeed in collecting the proposed comprehensive set of data, there are many possibilities to apply AI and other technologies to manage current health challenges and prevent future conditions. Managing and updating one system can be more efficient than relating to a multiple set of apps and systems. Whether this approach succeeds, is yet to be seen in the upcoming tests and pilots in the WARIFA project. Ideally, the system should support different platforms, i.e., mobile phones (e.g., Android and iOS), different health related sensors and other health data portals (e.g., Google Fit and Apple Health), as well as external and internal environmental factors. Future publications will include details about the results of these aims, and health effects using the presented system.

5. Conclusions

We have demonstrated that it is possible to make an app that gathers a lot of data from various sources, though it is complicated due to many factors, including security and privacy regulations. A main factor to succeed with this approach is maintaining close communications between backend developers, frontend developers, clinicians, patients, and data protection officers. It has been demonstrated how the diverse set of data can be used in algorithms helping to guide users to understand their health conditions, which also will be the future focus in the WARIFA project.

Acknowledgements

The members of the WARIFA Consortium collaborators in this work are as follows: Alejandro Déniz-García, Complejo Hospitalario Universitario Insular Materno-Infantil, and Instituto Universitario de Investigaciones Biomédicas y Sanitarias; Ana-Maria Forsea, Dermatology Department, Carol Davila University of Medicine and Pharmacy, Bucureşti, Romania; Giovanni Sebastiani, Istituto per le Applicazioni del Calcolo "Mauro Picone", Consiglio Nazionale delle Ricerche, Rome, Italy, Dipartimento di Matematica Guido Castelnuovo, Sapienza Università di Roma, Rome, Italy, and Department of Mathematics and Statistics, UiT The Arctic University of Norway, Tromsø, Norway; Guri Skeie, Department of Community Medicine, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway; Inger Torhild Gram, Department of Community Medicine, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway and Norwegian Centre for E-health Research, Tromsø, Norway; Maja-Lisa Løchen, Department of Clinical Medicine, UiT the Arctic University of Norway, Tromsø, Norway; Marit B. Veierød, Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway; Roald Nystad, University of Bergen, Bergen, Norway; Maryam Tayefi, Norwegian Centre for E-health Research, Tromsø, Norway; Terje Solvoll, Norwegian Centre for E-health Research, Tromsø, Norway, and Faculty of Nursing and Health Sciences, Nord University, Bodø, Norway; Thomas Roger Schopf, Norwegian Centre for E-health Research, Tromsø, Norway.

References

[1] Busse R, Blümel, M, Scheller-Kreinsen, D, Zentner, A,. Tackling Chronic Disease in Europe: World Health Organization2010.
[2] European Commission. White Paper on Artificial Intelligence: A European approach to excellence and trust. 2020.
[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), (2016).
[4] International Organization for Standardization. ISO/IEC 27005:2022. Information security, cybersecurity and privacy protection2022.
[5] Philip BJ, Abdelrazek M, Bonti A, Barnett S, Grundy J. Data Collection Mechanisms in Health and Wellness Apps: Review and Analysis. JMIR Mhealth Uhealth. 2022;10(3):e30468.