

Replace instruction (in italic style), You copy this document and share it openly so that everyone with the link can comment, not just within the organisation.

Student Project Proposal (SPP)

This document is intended to guide students to describe their SPPs for the 2dv50e course so that course management gets an idea of if it is suitable, or not. It should also guide the assignment of supervisors. This is not a contract, what you write here will most probably change.

We want to have projects that are:

- Clearly and precisely defined, that can be understood.
 - Written in English
 - Spellcheck and check grammar.
 - Proofread it! Read it aloud to check your language.
 - Let another human read it.
 - Practice your pitch to several people and ask them to explain your project back to you afterwards.
 - Watch out for prepositions.
- Projects within Computer Science
 - Builds on previous research (Related Work) within computer science.
 - It's research questions are interesting to the CS community
- Realistic and manageable projects
 - Can be done by one/two students within the timeframe of the thesis
 - Work to reduced risk of failure
 - Access to resources needed
- Meaningful and motivated
 - contribution to research
 - contribution to society
 - ethically motivated
 - has a target audience
- Suitable for your program profile
 - builds on courses you have taken

Your name(s): Gustav Näslund & Duc Anh Pham

LNU Supervisor:

Supervision status according to student

Check one of these:

☐ Currently working with (Ongoing project)

☐ Agreed but not started

☐ Wish to work with (or unclear of status)

☒ No preference, help me find one

Cooperative partners (Optional):

Company: Neurawave

Status: Two video meetings and follow up e-mails.

Preliminary Title:

Designing a Crossplatform Mobile Framework for Normalizing and Integrating Wearable Health Data to Support Machine Learning Applications in Migraine Analysis.

Elevator pitch:

The proliferation of wearable health devices has created a fragmented data landscape where each platform uses different formats and APIs [1], making it challenging for developers to create cross-platform applications that can effectively utilize health data for migraine monitoring and prediction [2]. While solutions like Shimmer exist for server-based data aggregation [3], modern health applications require real-time data access and direct platform integration, particularly for continuous monitoring of migraine-related physiological signals. I intend to develop a mobile framework that standardizes the integration of migraine-relevant health data from multiple wearable platforms into the Open mHealth format [4], providing developers with a unified interface for accessing normalized wearable data suitable for migraine analysis and prediction. The framework will be evaluated using Design Science methodology, including performance benchmarks and validation of data normalization accuracy across different platforms, with specific focus on data types relevant to migraine monitoring.

Steps/Milestones/Actions

- Research and requirements gathering
 - Investigate wearable APIs (e.g Apple Healthkit, Fitbit API etc) and their data formats.
 - Study existing frameworks for data normalization and integration.
 - Define the functional and technical requirements for framework/app.
 - Identify key migraine-related data points to prioritize.
- Design the framework/app architecture
 - Plan the overall architecture, including:
 - Front end (client side API for collection)
 - Back end (data processing)
 - Data pipeline (normalization and integration)
 - Define the data schema for the unified data model.
- Develop the data collection mechanism
 - Implement authentication to access wearable APIs.

- Build functionality to fetch data from at least one wearable device.
- Store raw data temporarily in the app or a local database.
- Implement data normalization and integration
 - Develop a pipeline to normalize raw data
 - Handle missing or inconsistent data
 - Integrate data from multiple wearable devices into a unified format.
 - Ensure the data is structured for machine learning
- Test and validate the framework/app
 - Test end-to-end functionality with sample data from wearable devices.
 - Validate the quality of the normalized and integrated data.
 - Ensure the data is machine learning ready.
- Optional milestone
 - Add support for additional wearable devices
 - Demonstrate how processed data could be used for machine learning, e.g by generating a sample dataset with features and labels.

Risks

- Lack of access to wearable data.

It might be challenging to access real world data from wearable devices due to API limitations (e.g restricted access, rate limit), privacy concerns and regulatory hurdles or difficulty recruiting users to share their wearable data. Can be mitigated by using publicly available datasets for prototyping. Synthetic datasets are also an alternative for mimicking real world data.

- Complexity of data normalization and integration

Due to the fact that wearable devices generate data in different formats, schemas, sampling rates which can make normalization and integration complex and time consuming. Can be mitigated by starting with one wearable device to simplify the initial implementation. The usage of existing libraries and tools can also aid in handling common normalization tasks.

- Data quality issue.

Incomplete, inconsistent or noisy data can affect the quality of the final dataset. Can be mitigated by implementing data validation checks (e.g range checks for heart rate). Missing data can be handled through removal or by using an average value.

- Privacy and regulatory compliance

Handling sensitive data requires compliance with strict privacy regulations (e.g GDPR) which can be complex and time consuming. Can be mitigated by usage of anonymized or synthetic data. Ensuring secure data transfer protocols as well as user consent might have to be obtained before collecting data.

- Time management and scope creep.

Since the project contains multiple components (data collection, normalization, integration, app development) there is a risk of scope creep or delays. Can be mitigated through definition of clear milestones and set deadlines for each milestone. Prioritizing core functionality can also ensure that the projects can get finished, as well as using an agile development practice.

- Limited expertise in wearable APIs or data pipelines.

Unfamiliarity with wearable APIs or data pipeline development might have a steep learning curve. Can be mitigated through allocation of time for learning and experimentation with wearable APIs, as well as time allocation for thorough reading of documentation.

- Difficulty demonstrating machine learning readiness

There might be challenges associated with demonstrating machine learning readiness. Can be mitigated by the creation of a clear definition of what “machine learning-ready” means for this project.

Background and Motivations

The current state of health data holds substantial promise due to wearable devices. Wearable devices such as smart watches, smart rings and smart bracelets has evolved in the recent years and are able to give more accurate high quality data, earning an increased amount of clinically approved certifications [1]. This high quality longitudinal health monitoring provides new possibilities within health research, disease identification and treatment [5]. The usage and market of wearable devices has grown significantly, with an expected compound annual growth rate of 14.6% from 2023 to 2030 [6]. Some of the most prominent actors within the wearable device market is Alphabet Inc/Google, Apple Inc, Garmin Ltd and Samsung Electronics Co., Ltd [6].

While the growth of wearable devices brings unprecedented opportunities for health monitoring, it also introduces significant challenges. The different providers provide their own platforms for accessing and modifying data, with their own API's and restrictions, which poses a challenge for anyone who wants to aggregate data from multiple providers. The data itself differs also in structure, even if the same variable is being measured. Apple provides through its HealthKit a more general abstraction of a measurement such as HKQuantityType [7] which represents a “quantity of something” such as stepCount, while Google Health represents the same variable in a more object oriented fashion with associated details included such as start and stop time for a step metric [8]. These differences in data representation extend beyond simple metrics such as step counts to more complex physiological measurements such as heart rate variability, sleep stages and stress levels. The lack of standardized approaches for normalizing and integrating this heterogeneous health data creates a significant barrier for developers and researchers trying to build cross platform health applications, particularly for those leveraging machine learning techniques.

Migraine is an episodic and complex sensory processing disturbance with a collection of different neurological symptoms where headache is the most common and primary symptom. Migraine is the most common cause of disability among young women, and the second most common cause of disability globally overall [9]. Migraine can be classified both with or without the presence of aura, and also if the migraine is chronic or episodic [10]. The economic burden of migraine for society is profound, in the USA there has been estimated that direct costs associated with migraine is more than 17 billion USD [10].

There have been several potential triggers of migraine identified, such as stress, hormonal changes, hunger, sleep disturbance, etc [11]. Many of these triggers are associated to lifestyle factors and are essential to manage in the preventive treatment [12]. Identification of these triggers is mainly up to the patient itself, and relies on the observations and recollection of the patient. This self reporting might suffer from flaws of subjectivity and faulty perceptions as indicated by Casanova et al [13] where “even the most commonly endorsed triggers were statistically associated with in fewer than one third of individuals suspecting each trigger”. The need for more objective measurements with a more detailed analysis is a glaring.

In terms of migraine, there is a possibility that the data obtained from wearable devices could provide invaluable insight into triggers both on an individual and societal level. The current digital treatment aids for headache mainly revolves around headache calendars, instructions for relaxation and endurance sports, chatbots as well as providing structured data for the treating medical team of a patient [14]. The analysis of the captured data could be handled through machine learning to enable the processing of large amounts of data and for a greater chance of discovering new patterns and insights of the data. There have been several studies examining the use of machine learning in the process of diagnosing and classifying migraines [15], [16], [17], as an aid when prescribing medicine [2] and for predicting migraine episodes with the data from wearable devices in combination with a digital headache diary [18].

There are multiple health platforms such as Apple Health [19], Google Health [20] and Samsung Health [21]. They are all closely integrated with their respective companies products, but can also be used to collect data with compatible third-party devices. The closest thing available as to a uniform platform that combines the data of different health platforms is Shimmer [3], which is “the first and only open-source health data aggregator”. Shimmer allows for fetching data from multiple different health platforms in an Open mHealth compliant format. This uniformity of the data structure is important because semantics of complex health data makes a huge difference, such as if blood glucose is measured fasting or non fasting [4]. Shimmer [3] is built in Java and is composed of several different components: individual shims, a resource server and a console [22]. A shim is defined as “an adapter between an Open mHealth API and the API of a third-party data provider” [23]. This server architecture allows for a uniform API for the user that is extracting data, and Shimmer handles many of the pitfalls such as authentication. While Shimmer provides invaluable service for data integration, its server-based architecture presents challenges for mobile application development. Modern health applications often require real-time data access, offline functionality and direct integration with platform specific SDKs. The need for a server intermediary can introduce latency, connectivity dependencies and additional complexity in mobile applications.

Health data can be categorized as structured and unstructured. Unstructured data is data that does not follow a standardized structure and could encompass a clients own description of how they are feeling. Structured data follows a standardized structure and has both a clear structure and meaning behind it. When training machine learning models with health related data, it is generally considered to be easier to do with structured data [24] compared to with unstructured data. For clarity there has been shown that unstructured data can contain very valuable information and if handled correctly in accordance with structured data, the combination can potentially provide additional benefits [25]. Due to the fact that Open mHealth provides this structured format for health measurements performed by wearable devices, that can be considered to be a good preparation of the data for machine learning as well.

The knowledge gap lies in the combination of the gathering of the relevant health data through wearable devices, the utilization of the data for predicting migraines and the standardization of the data to a common format that could be used for machine learning as a part of the solution. The application of all of these components into a cross platform mobile framework does not exist to date. The framework is important due to the potential of enabling better health for those suffering of migraine, and as well for the potential societal impact of alleviating resources that otherwise would be consumed by migraine patients.

The target group for this thesis is health informatics developers, mobile developers and machine learning & artificial intelligence researches in health care.

The closest solutions

Name	Features	Missing features
health 12.0.1	Enables reading and writing health data to and from Apple Health and Google Health Connect.	Single provider, non standardization
React Native Health	Package to interact with Apple HealthKit for iOS.	Single provider, non standardization
React Native Health Connect	Package to interact with Health Connect for Android	Single provider, non standardization
Shimmer	Facade for multiple providers to extract normalized data.	Non mobile multiplatform
Tasrif	Facade for multiple providers to extract data with integration to ML libraries.	Standardized format, non mobile platform

Application areas: Healthcare analytics, digital health platforms, machine learning for healthcare.
Research area: Data integration and interoperability

Related work

This work positions itself in the area of Data integration and interoperability, specifically focusing on frameworks for normalizing heterogeneous health data from wearable devices for migraine monitoring applications.

Current approaches to wearable data integration can be categorized into two main types: web-based solutions and data processing frameworks. WearMerge [26], presented at IEEE International Conference on Pervasive Computing and Communications Workshops in 2022, represents a web-based approach, converting wearable data to Open mHealth schemas. While this framework addresses similar data aggregation challenges, it differs from our proposed solution in its web-based architecture, which limits its applicability for real-time mobile applications.

In the same year, Tasrif [27] was introduced as a Python-based preprocessing framework for wearable data. While it effectively handles data from commercial devices like Apple Health and provides direct integration with machine learning frameworks, it focuses on offline data processing rather than real-time mobile integration. The authors demonstrate its value in reducing barriers for clinical researchers, but like WearMerge, it's not designed for direct integration into mobile applications.

The novelty of our approach lies in addressing the specific challenges of mobile-first, cross-platform data integration, an area that remains unexplored in current literature. The recent publication dates of both WearMerge and Tasrif (2022) indicate this is an active and emerging research area within computer science, with significant opportunities for mobile-specific solutions.

Knowledge Gap/Challenge/Problem

There is a significant technical challenge in developing mobile applications that can collect and normalize wearable health data from multiple providers, particularly for continuous monitoring applications like migraine prediction. While single-vendor solutions exist, and server-based platforms like Shimmer [3] offer multi-vendor support, there is no cross-platform mobile framework that enables real-time data integration. This gap between single-vendor mobile frameworks and server-based data processing applications has been acknowledged in the context of big data healthcare applications [28], and becomes particularly critical for applications requiring continuous monitoring and real-time data access.

Knowledge Contribution/Action

This thesis aims to develop a mobile framework that standardizes the integration and normalization of migraine-relevant health data from multiple wearable platforms (Apple HealthKit, Google Health, etc.) into the Open mHealth format, with particular focus on metrics associated with migraine triggers and symptoms (such as sleep patterns, stress levels, and activity data). The framework will provide developers with a unified interface for accessing normalized wearable data across platforms while handling the complexity of different data formats and sampling rates, specifically optimized for migraine monitoring and prediction applications.

The contribution described as a set of research questions:

- How can wearable health data from different platforms be effectively normalized into a unified format suitable for migraine-focused machine learning applications?
- What are the key requirements and challenges in implementing real-time data normalization for migraine-relevant wearable data in a mobile environment?
- How effective is the Open mHealth schema as a standardized format for representing migraine-relevant physiological data from diverse wearable devices?

Empirical Evidence/Evaluation

This thesis will follow the Design Science Research methodology as outlined by Peffers et al. [29], which is suitable for developing and evaluating IT artifacts. The research will follow DSR's key phases: 1) Problem identification

- Literature review of current health data integration challenges
- Analysis of existing solutions and their limitations

2) Define objectives

- Established requirements for cross-platform data normalization
- Define success criteria for the framework

3) Design & development

- Iterative development of the framework
- Implementation of data normalization components
- Integration with health platforms

4) Demonstration

- Development of a proof of concept application for migraine monitoring
- Implementation of data collection for key migraine-relevant metrics

5) Evaluation

- Technical performance assessment (processing time, memory usage)
- Validation of data normalization accuracy for migraine-relevant metrics
- Testing with multiple data sources and sampling rates
- Verification of data consistency for machine learning preparation

Supporting methods will consist of: 1) Literature review

- To identify current challenges and solutions
- To establish state-of-the-art in health data integration for mobile applications

2) Controlled experimentation

- To measure framework performance
- To validate data normalization accuracy

References

- [1] Sophie Huhn *et al.*, “The Impact of Wearable Technologies in Health Research: Scoping Review,” *JMIR Mhealth Uhealth*, vol. 10, no. 1, Jan. 2022.
- [2] Chia-Chun Chiang *et al.*, “Advancing toward precision migraine treatment: Predicting responses to preventive medications with machine learning models based on patient and migraine features,” *Headache The Journal of Head and Face Pain*, vol. 64, no. 9, Oct. 2024.
- [3] [Online]. Available: <https://www.openmhealth.org/documentation/#/data-providers/shimmer>
- [4] [Online]. Available: <https://www.openmhealth.org/documentation/#/schema-docs/overview>
- [5] Mohan Babu, Ziv Lautman, Xiangping Lin, Milan H B Sobota, and Michael P Snyder, “Wearable Devices: Implications for Precision Medicine and the Future of Health Care,” *Annual Review of Medicine*, vol. 1, no. 1, Nov. 2023.
- [6] Grand View Research, “Wearable Technology Market Size, Share & Trends Analysis Report By Product (Head & Eyewear, Wristwear), By Application (Consumer Electronics, Healthcare), By Region (Asia Pacific, Europe), And Segment Forecasts, 2023 - 2030,” 978–1–68038–165–8.
- [7] [Online]. Available: <https://developer.apple.com/documentation/healthkit/hkquantitytype>
- [8] [Online]. Available: <https://developer.android.com/reference/kotlin/androidx/health/connect/client/records/StepsRecord>
- [9] Antonio L Aguilar-Shea and c. Javier Diaz-de-Teran b, “Migraine review for general practice,” *Atencion Primaria*, vol. 54, no. 2, Nov. 2021.
- [10] Johra Khan *et al.*, “Genetics, pathophysiology, diagnosis, treatment, management, and prevention of migraine,” *Biomedicine & Pharmacotherapy*, vol. 139, no. 1, Jul. 2021.
- [11] Hassan Kesserwani, “Migraine Triggers: An Overview of the Pharmacology, Biochemistry, Atmospherics, and Their Effects on Neural Networks,” *Cureus*, vol. 13, no. 4, Apr. 2021.
- [12] Jason C Ray, Caroline Macindoe, Michael Ginevra, and Elspeth J Hutton, “The state of migraine: An update on current and emerging treatments,” *Australian Journal of General Practice*, vol. 50, no. 12, Dec. 2021.
- [13] Amparo Casanova, Marina Vives-Mestres, Stephen Donoghue, Alec Mian, and Paul R Martin, “An observational study of self-reported migraine triggers and prospective evaluation of the relationships with occurrence of attacks enabled by a smartphone application (App),” *Headache The Journal of Head and Face Pain*, vol. 62, no. 10, Jun. 2022.
- [14] L Neeb, R Ruscheweyh, and T Dresler, “[Digitalization in headache therapy],” *Schmerz*, vol. 34, no. 6, Dec. 2020.
- [15] Wei-Ta Chen, Cing-Yan Hsieh, Yao-Hong Liu, Pou-Leng Cheong, Yi-Min Wang, and Chia-Wei Sun, “Migraine classification by machine learning with functional near-infrared spectroscopy during the mental arithmetic task,” *Scientific Reports*, vol. 12, no. 1, Aug. 2022.
- [16] Katarina Mitrović, Igor Petrušić, Aleksandra Radojičić, Marko Daković, and Andrej Savić, “Migraine with aura detection and subtype classification using machine learning algorithms and morphometric magnetic resonance imaging data,” *Frontiers in Neurology*, vol. 14, no. 1, Jun. 2023.
- [17] Lal Khan, Moudasra Shahreen, Atika Qazi, Syed Jamil Ahmed Shah, Sabir Hussain, and Hsien-Tsung Chang, “Migraine headache (MH) classification using machine learning methods with data augmentation,” *Scientific reports*, vol. 14, no. 1, Mar. 2024.

- [18] Anker Stubberud *et al.*, “Forecasting migraine with machine learning based on mobile phone diary and wearable data,” *Cephalalgia: an international journal of headache*, vol. 43, no. 5, May 2023.
- [19] [Online]. Available: <https://www.apple.com/health/>
- [20] [Online]. Available: <https://health.google/>
- [21] [Online]. Available: <https://play.google.com/store/apps/details?id=com.sec.android.app.shealth>
- [22] [Online]. Available: <https://github.com/openmhealth/shimmer>
- [23] [Online]. Available: <https://www.openmhealth.org/documentation/#/data-providers/about-shims>
- [24] Maryam Tayefi *et al.*, “Challenges and opportunities beyond structured data in analysis of electronic health records,” *WIREs - Wiley Interdisciplinary Reviews*, vol. 13, no. 6, Feb. 2021, doi: 10.1002/wics.1549.
- [25] Dongdong Zhang, Changchang Yin, Jucheng Zeng, Xiaohui Yuan, and Ping Zhang, “Combining structured and unstructured data for predictive models: a deep learning approach,” *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, Oct. 2020, doi: 10.1186/s12911-020-01297-6.
- [26] Dimitrios Panteleimon Giakatos, Sofia Yfantidou, Stefanos Efstathiou, and Athena Vakali, “WearMerge: An Interoperable Framework for Self-tracking Data Integration and Standardization,” May 06, 2022, *IEEE*. doi: 10.1109/PerComWorkshops53856.2022.9767462.
- [27] Abdulaziz Al Homaid, Syed Hashim, Fadhil Abubaker, Ummar Abbas, Faisal Farooq, and Joao Palotti, “Tasrif: processing wearable data in Python,” May 06, 2022, *IEEE*. doi: 10.1109/PerComWorkshops53856.2022.9767286.
- [28] Sumarsono, Muhammad Anshari, and Mohammad Nabil Almunawar, “Big Data in Healthcare for Personalization & Customization of Healthcare Services,” Sep. 19, 2019, *IEEE*. doi: 10.1109/ICIMTech.2019.8843822.
- [29] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee, “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems*, vol. 24, no. 3, 2007, doi: 10.2753/MIS0742-1222240302.