2. Method

The applied method for this thesis is Design Science Research (DSR). DSR was chosen due to its natural fit for projects involved in creation of a software artifact. It is described by vom Brocke et al [1] as a paradigm that seeks to enhance knowledge through the creation of an innovative artifacts. They further explain that the aim of DSR is to generate knowledge about how things can and should be constructed, knowledge that is referred to as design knowledge within the DSR paradigm. Within the scope of this thesis, several research methods are used in combination. These include literature review to understand the theoretical background, artifact construction through iterative development and experimentation to evaluate the artifacts ability to meet the defined objectives. Together these methods form a multimethod approach aligned with the DSR process model.

2.1 Research project

The company Neurawave [2] presented a need for collecting health metrics that could be used for migraine prediction with machine learning. In this work, we follow the DSR methodology outlined by Peffers et al [3].

Phase 1 - Problem identification and insight gathering

We began our research project by conducting open-ended interviews with the founders of neurawave [2]. This approach helps us build a deeper understanding of both stakeholder needs and the problem domain. One of the key requirements was that the data collection mechanism had to be integrated into their existing cross-platform mobile application. The main reason for this requirement is the ability for the existing application to act upon realtime health data. Additionally, we conducted literature review to gather insights regarding the state of the art available solutions, and which health data is commonly used in migraine prediction.

Phase 2 - Objective formulation

Based on the analysis of identified requirements gather in open-ended interviews and resrach gap found in the literature review, we have defined the projects objectives.

Phase 3 - Design and implementation

Based on the formulated objectives, the software artifact which is the data collection framework for mobile platforms was developed and refined iteratively.

Phase 4 - Demonstrating efficacy

Once the artifact has been developed, we will demonstrate its capabilities in addressing the defined problem through system testing. This will be followed by rigarous testing of how well the artifact meets the objectives, likely by comparing the implemented functionality with the stakeholder-defined goals.

Phase 5 - Critical evaluation and feedback integration

The quantitative assessments using controlled experiment is applied to measure the accuracy and reliability of the data collection framework in different scenarios or use-cases.

Phase 6 - Dissemination and community engagement

Our findings and its implications are articulated in this thesis work for dissemination among AI researchers and developers.

The next step in the project involved defining the objectives of a solution, which had to be closely aligned with the problem identified and the needs expressed by the stakeholders. During this phase, we will also conduct a literature review to gather insights into existing solutions and to better understand the current state of the problem domain.

Subsequently, we will proceed to develop the artifact, in our case the data collection component for mobile platforms. The development will be carried out in an iterative manner, including both the creation of the software and the refinement of its functionality based on the defined objectives. Once the artifact has been developed, we will demonstrate its capabilities in addressing aspects of the defined problem through experimentation or simulation. This will be followed by rigorous evaluation of how well the artifact meets the objectives, likely by comparing the implemented functionality with the stakeholder-defined goals.

2.2 Research methods

We began our research project by conducting open-ended interviews with the founders of Neurawave [2]. Open-ended interviewing allows for a flexible, free-flowing conversation and is considered a qualitative method of data collection [4]. This method was chosen because the stakeholders possessed significantly more domain knowledge than us, making them essential in framing the problem accurately. As described by Alshenqeeti [4], the purpose of open-ended interviews is to "broaden the scope of understanding of investigated phenomena". We believed this approach would help us build a deeper understanding of both stakeholder needs and the problem domain.

Open-ended interviews fall within the broader category of qualitative methods, which are generally holistic and aimed at answering "what" questions. In contrast, structured interviews are more quantitative, relying on closed-ended questions, such as those with yes/no responses [4]. As Lakshman et al. [5] explain, quantitative methods focus on examining the effect of an independent variable on a dependent variable in a measurable, numerical way. However, at this early stage in the project, a qualitative approach was better suited to developing our understanding and framing our objectives.

Following the interviews, we conducted litterature research to expand our contextual understanding of the problem area. Literature research involves exploring a broad body of work related to a topic and is typically less rigid than a systematic literature review [6]. A systematic review, by contrast, targets a specific research question and is intended to gather empirical evidence to support conclusions. We chose a narrative-style literature review to ground our work theoretically and better understand the current landscape of data collection for health-related machine learning applications.

Alternative approaches, such as expert interviews could also have been employed to gather real-world insights. As described by Döringer [7], expert interviews are valuable for problem-centered exploration and knowledge gathering. However, due to the time and resource constraints, this method was not feasible within the scope of this thesis. Nonetheless, it represents a promising avenue for future research and validation.

During the design and development phase, we followed the artifact creation process outlined in the Design Science Research methodology. While this is not an empirical method in the classical sense, it is a core research activity in DSR, where existing knowledge and stakeholder input are synthesized into a functional solution. We began by defining application requirements that map directly to stakeholder requirements and then iteratively developed the plugin aimed at fulfilling those needs. No specific structured development framework as followed, as our team had prior experience in collaborative software development.

To evaluate the developed solution, we will use experimentation as the primary method of validating functionality. According to Basili et al. [8], experimentation is an iterative process of hypothesizing and testing. In our case, this involves defining (or redefining) software requirements, implementing them (the hypothesis), and verifying wether those requirements are fulfilled (the test).

Additionally, we will perform validation to ensure tha the plugin's software requirements align with the original stakeholder expectation [9]. As an alternative to experimentation, we considered survey-based research to gather stakeholder opinions on desired functionality. However, we ultimately decided against this due to our limited timeframe. We also believed that the iterative loop supported by experimentation would be hindered by the time required to create, distribute and analyze a survey. In our context, experimentation offers more immediate feedback and supports rapid iteration, which we viewed as essential for effective development.

2.3 Data for migraine prediction using machine learning

During the literature review of previous studies investigating migraine prediction using machine learning, we found only two studies [10], [11]. The health metrics used in both studies are outlined below. It is worth noting that several health metrics was included in either study, such as hours of working [11] out and step count [10]. We also found some studies [12], [13], [14] that examined different triggers (internal and external factors) such as weather, diet, hormonal changes that could be valuable in predicting migraine episodes, but none of those were included in the studies that combined migraine prediction with machine learning.

Health metric	Description	Refs
Heart rate	The rate of heart beats per minute.	[10], [11]
Heart rate variability	Variation in time between heartbeats	[10]
Skin temperature	The temperature of the skin, preferably measured at finger or wrist	[10], [11]
Skin conductance	Measurement of how good the skin conducts electricity. Has been shown to increase due to sweating, as a response to stress	[10]
Respiratory rate	Measurement of the rate of breathing. Measured as breaths per minute.	[10]
Sleep time	Measurement the amount of hours slept	[10], [11]

We have decided to include heart rate and skin temperature in our data processing. The selection is based on a combination on what data is available from regular wearables and what health metrics have been shown in previous studies to be of high value.

2.3 Reliability and validity

Validity

One limitation related to validity stems from the scope of the evaluation. The developed component will be tested by two individuals in controlled, but limited, settings. As such, the generalizability of the findings to a broader user base or different use cases is uncertain. Furthermore, the component is validated exclusively on migraine related health data. While it may be adaptable to other domains within health monitoring, such applicability remains untested and therefore unknown.

Another validity concern involves the integration of the component with the datastores and/or wearable devices. The component will initially be tested using two or three specific health data providers. Its performance and compatibility with other, non integrated wearables remain

unverified, and thus it cannot be assumed that the framework will function the equally well across other health data providers.

Additionally the component rely on third party API's for data collection. Any inaccuracies or failure in those APIs could directly compromise the integrity of the data and therefore the validity of the results. This is particularly concerning given that the quality and frequency of data vary significantly between high cost devices (e.g, Apple Watch, Fitbit) and low cost alternatives (e.g, Bangle.js, EmotiBit). The component might perform well in high quality data environments but underperform in cases where the input data is sparse, noisy or unreliable.

Reliability

A significant threat to reliability is the dynamic nature of third party wearable health data provider APIs. These APIs are frequently updated, and changes in their structure or functionality may break the integration with out framework. This implies that future researchers attempting to replicate this study might need to adapt the codebase to updated APIs. To mitigate this risk, we will document the integration process thoroughly and provide implementation guides to facilitate replication.

On the other hand, the controlled experiments conducted as a part of this project will be carefully designed and well documented. This will support the reproducibility of results and allow others to validate the findings under similar conditions.

Risks and mitigations strategies

A number of general risks may impact both the reliability and validity of the project outcomes:

- Lack of access to wearable data may hinder testing or validation of the component under real world conditions. To address this, we prioritize using devices with public or open APIs and ensure local caching of test data were possible.
- Complexity of data normalization and integration may result in inconsistent behavior across devices. This will be mitigated by adopting standardized data formats and implementing preprocessing checks.
- Data quality issues, especially from low-cost or experimental devices may reduce the effectiveness of the component. We will include a validation layers to detect and handle poor input.
- Time management and scope creep pose risks to project completion. A well defined timeline and iterative planning approach are used to ensure focus and manage scope.
- Limited expertise in wearable APIs and data pipelines may slow progress. To mitigate this, we will rely on documentation and existing libraries where possible.
- Difficulty demonstrating machine learning readiness of the collected data may arise due to insufficient volume or inconsistency. We will evaluate and report data quality with descriptive statistics and document its limitations for future machine learning integration.
- API rate limits and access restrictions may impact data collection throughput. Mitigation strategies include caching responses, adhering to API usage guidelines and contacting vendors if necessary.

By identifying these threats and planning accordingly, this project strives to maintain both reliability and validity, despite the challenges inherent in working with third-party hardware and health data.

2.4 Ethical considerations

While this project doest not involve direct interaction with human participants outside of the stakeholders in the initial interviews, it does involve collecting and processing of health related data via wearable devices. This introduces several ethical considerations, particular around privacy, data confidentiality and data ownership.

Confidentiality

The component is designed to collect health metrics that are sensitive by nature, such as heart rate, heart rate variability and skin temperature. Even though no personally identifiable information is intended to be handled, care must be taken to ensure that the processing of data is in accordance with privacy regulations such as GDPR.

To address this, all test data used during development and experimentation will be anonymized. No user names, contact information, or device IDs linked to individuals will be stored or processed in any way that could allow for re-identification. No data will be stored within the component, only processing of the data. No data will be shared externally or stored in the cloud.

Sampling bias

The testing of the component will be done with using a limited number of wearable devices and datasets related to migraine prediction. As a result, the dataset is narrow in scope which may introduce selection bias. This poses a limitation upon the generalizability of the results to other conditions such as other devices or populations. Since the primary objective is to develop and evaluate a working prototype in a focused context, this tradeoff is considered acceptable for the scope of the project.

Participation and consent

If any real user data is collected (for example, if Neurawave employees or external users voluntarily contribute data for evaluation), explicit informed consent will be obtained in written form. Participants will be made aware of what data is being collected for processing and how it will be used. Participation will be strictly voluntary, and participants will have the right to withdraw at any time without any consequence.

Risk of harm

This project poses minimal risk of harm. as it does not involve any physical or psychological intervention. However, improper handling of sensitive data could lead to privacy breaches. To minimize this risk, the component itself will not store any data.

Bibliography

[1]   in *Design Science Research Cases*, 2020, pp. 1–13.

[2]   Neurawave, "Neurawave's website." [Online]. Available: https://www.neurawave.se/

[3]   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.

[4]   Hamza Alshenqeeti, *English Linguistics Research*, vol. 3, no. 1, Mar. 2014, doi: 10.5430/elr.v3n1p39.

[5]   M. Lakshman, Leena Sinha, Moumita Biswas, Maryann Charles, and N.K. Arora, "Quantitative vs Qualitative research methods," *The Indian Journal of Pediatrics*, vol. 67, no. 1, May 2000, doi: https://doi.org/10.1007/BF02820690.

[6]   Priscilla Robinson and John Lowe, "Literature reviews vs systematic reviews," *Australian and New Zealand Journal of Public Health*, vol. 39, no. 2, Apr. 2015, doi: 10.1111/1753-6405.12393.

[7]   Stefanie Döringer, "The problem-centered expert interview. Combining qualitative interviewing approaches for investigating implicit expert knowledge," *International Journal of Social Research Methodology*, vol. 24, no. 3, May 2020, doi: 10.1080/13645579.2020.1766777.

[8] Victor R. Basili, Richard W. Selby, and David H. Hutchens, "Experimentation in software engineering," *IEEE Transactions on Software Engineering*, vol. 12, no. 7, Sep. 2012, doi: 10.1109/TSE.1986.6312975.

[9] P. Sujatha, G. V. Sankar, A. S. Rao, and T. Satyanarayana, "The Role of Software Verification and Validation in Software Development Process," *IETE Technical Review*, vol. 18, no. 1, Mar. 2015, doi: 10.1080/02564602.2001.11416938.

[10] Viroslava Kapustynska, Vytautas Abromavičius, Artūras Serackis, Šarūnas Paulikas, Kristina Ryliškienė, and Saulius Andruškevičius, "Machine Learning and Wearable Technology: Monitoring Changes in Biomedical Signal Patterns during Pre-Migraine Nights," *Healthcare (Basel)*, vol. 12, no. 17, Aug. 2024, doi: https://doi.org/10.3390/healthcare12171701.

[11] Anker Stubberud *et al.*, "Forecasting migraine with machine learning based on mobile phone diary and wearable data," *Cephalgia*, vol. 43, no. 5, May 2023, doi: https://doi.org/10.1177/03331024231169244.

[12] Antonio L Aguilar-Shea, Javier A Membrilla Md, and Javier Diaz-de-Teran, "Migraine review for general practice," *Aten Primaria*, vol. 54, no. 2, Nov. 2021, doi: 10.1016/j.aprim.2021.102208.

[13] Michael J Marmura, "Triggers, Protectors, and Predictors in Episodic Migraine," *Current pain and headache reports*, vol. 12, no. 81, Oct. 2018, doi: 10.1007/s11916-018-0734-0.

[14] Dana P Turner, Adriana D Lebowitz, Ivana Chtay, and Timothy T Houle, "Forecasting Migraine Attacks and the Utility of Identifying Triggers," *Current pain and headache reports*, vol. 22, no. 9, Jul. 2018, doi: 10.1007/s11916-018-0715-3.