Lessons Learned from mHealth Monitoring in the Wild

Pedro Almir M. Oliveira[1,2][a], Rossana M. C. Andrade[1][b] and Pedro A. Santos Neto[3][c]

[1]*Group of Computer Networks, Software Engineering and Systems (GREat), Federal University of Ceará, Ceará, Brazil*

[2]*Federal Institute of Maranhão (IFMA), Pedreiras, Maranhão, Brazil*

[3]*Laboratory of Software Optimization and Testing (LOST), Federal University of Piauí, Piauí, Brazil*

Keywords: Health Monitoring, Self-reported Quality of Life, Practical Report, Lessons Learned.

Abstract: In the modern world, it is no overstatement to say that "*our devices know us better than we know ourselves*". In this sense, the vast amount of data generated by wearables, mobile devices, and environmental sensors has enabled the development of increasingly personalized and intelligent services. Among them, there is a growing interest in the delivery of medical practice using mobile devices (*i.e.*, mobile health or mHealth). mHealth makes it possible to optimize healthcare systems based on continuous and transparent health monitoring, aiming to detect the emergence of diseases. However, mHealth monitoring in the real world (*i.e.*, uncontrolled environment or, as labeled in this paper, "in the wild") has many challenges. Therefore, this practical report discusses ten lessons learned from the Quality of Life (QoL) monitoring of twenty-one volunteers over three months. The main objective of this QoL monitoring was to collect data capable of training Machine Learning algorithms to infer users' Quality of Life using the WHOQOL-BREF as a reference. During this period, our research team systematically recorded the problems faced and the strategies to overcome them. Such lessons can support researchers and practitioners in planning future studies to avoid or mitigate similar issues. In addition, we present strategies for dealing with each challenge using the 5W1H model.

1 INTRODUCTION

Our world is becoming mobile (Palos-Sanchez et al., 2021). As a benchmark, 67.1% of the world's population uses smartphones, which means 5.31 billion of unique users by the start of 2022[1]. In addition, a similar percentage – 62.5% of the world's population – has Internet access. This outstanding diffusion associated with advances in hardware (such as, cost reduction, sensor miniaturization, and expansion of processing power) has enabled a massive transformation in access to a variety of healthcare services, especially in the area called mobile health (mHealth, for short) (Bravo et al., 2018).

mHealth can be defined as the delivery of medical practice by mobile devices, including smartphones, tablets, or wearable monitoring devices (Bostrom et al., 2020). mHealth apps facilitate the collection and sharing of health data and the delivery of health services (Qudah and Luetsch, 2019).

Unique features such as accessibility, context awareness, and personalized solutions have made the use of mHealth attractive for the healthcare industry (Akter and Ray, 2010). The mobile health market was valued at USD 63 billion in 2021 and is projected to reach more than 230 billion by 2027[2]. Furthermore, mHealth has emerged as an opportunity to optimize health systems resources, promoting high-quality at a low-cost (Islam et al., 2015).

As stated by Varshney (2014), mobile health is more than just some healthcare applications on a mobile phone. Mobile health makes possible many kinds of applications such as non-intrusive Quality of Life (QoL) monitoring (Oliveira et al., 2022c), older adults fall detection (Araújo et al., 2022), gait and posture analysis (Junior et al., 2021).

The analytical model of mobile health generally includes applications that assist patients (*i.e.*, users) during their treatment. For example, the patient can be a child, an adult, or an older person. They can also have chronic or acute illnesses, and these health issues make them dependent or independent.

In addition, healthcare professionals follow up

[a] https://orcid.org/0000-0002-3067-3076
[b] https://orcid.org/0000-0002-0186-2994
[c] https://orcid.org/0000-0002-1554-8445

[1]Digital 2022 Global Report: wearesocial.com.

[2]mordorintelligence.com/industry-reports.

Oliveira, P., Andrade, R. and Santos Neto, P.
Lessons Learned from mHealth Monitoring in the Wild.
DOI: 10.5220/0011689600003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 5: HEALTHINF*, pages 155-166
ISBN: 978-989-758-631-6; ISSN: 2184-4305

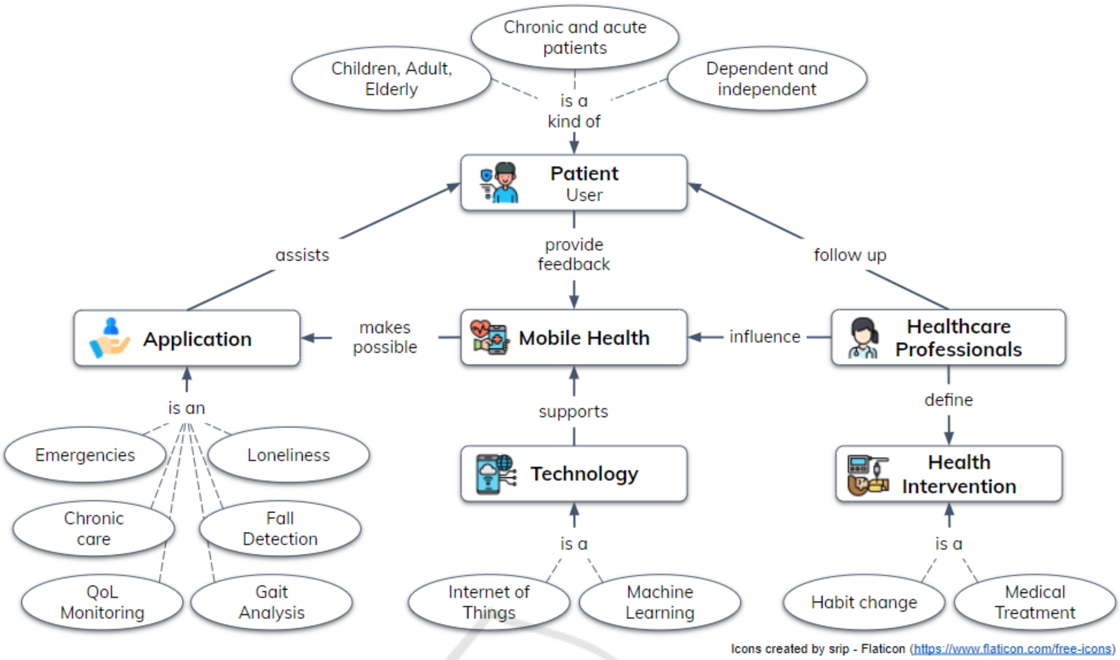

Figure 1: Mobile Health analytical model. Adapted from (Varshney, 2014).

with the patient and define health interventions (*e.g.*, habit change). Figure 1 reinforces this analytical model, highlighting that the Internet of Things (Rodrigues et al., 2018) and Machine Learning (Ian and Eibe, 2005) are technologies applied to support mHealth.

In this scenario, monitoring personal QoL using mobile health applications has attracted the interest of many researchers (Oliveira et al., 2022b) due to the ability of these technologies to get data capable of understanding human behavior. Furthermore, this kind of monitoring is relevant due to the health benefits that can be achieved from an accurate QoL analysis, such as disease detection and early healthcare interventions (Oliveira et al., 2022c). Dohr *et al.* (2010) also reinforces that these benefits have individual impacts by increasing well-being, economic impacts by improving the cost-effectiveness of healthcare resources, and social impacts by promoting better living conditions.

The history of the Quality of Life term began a long time ago (Elkinton, 1966). Even so, despite being discussed a lot, this term can be observed from many perspectives (Karimi and Brazier, 2016). For example, the QoL can be related to the absence of chronic diseases, perception of loneliness, physical well-being, and understanding of the aging process.

The World Health Organization's Quality of Life definition is the primary reference in this work. Thus, QoL can be defined as the individual perception of life in a sociocultural context (Orley and Kuyken, 1994). Based on this definition, many instruments to assess QoL have been proposed, such as the WHOQOL-BREF questionnaire, SF-36 health survey, and KIDSCREEN-52 for children.

Unfortunately, the continuous application of this kind of questionnaire is tedious, bothersome (Sanchez et al., 2015), and can also include a bias as the patient needs to actively provide the information, which makes it challenging to engage the participants (Hao et al., 2017). Therefore, QoL continuous monitoring is still an open problem due to the complexity of the measuring instruments and the invasive approaches that do not preserve privacy (Oliveira et al., 2022b).

Motivated by this context, we decided to start – in a previous work (Oliveira et al., 2022c) – an investigation about the use of the Internet of Health Things (IoHT) to collect data from Smart Environments and apply Machine Learning to infer QoL measures. Then, to evaluate this proposal, we conducted a longitudinal study in which twenty-one (21) participants were monitored "in the wild" for three months. This expression "in the wild" reinforces the inherent complexity of monitoring health data outside a controlled environment such as a laboratory or hospital.

Thus, the main goal of this paper is to present and discuss the lessons learned during this longitudinal health monitoring. The systematization of these lessons contributes to researchers and practitioners anticipating possible issues and highlighting some strategies to overcome them.

This paper is outlined as follows: Section 2 discusses similar studies focused on lessons learned from mHealth data monitoring; Section 3 details the longitudinal study design; Section 4 briefly exposes the results obtained by the Machine Learning regressors; then, Section 5 discusses the volunteers' perceptions using the Technology Acceptance Model; Section 6 presents all challenges and limitations faced in the study; then, Section 7 summarizes lessons learned and strategies to address possible challenges related to health monitoring "in the wild"; and, finally, the Section 8 brings final remarks and future work.

2 RELATED WORK

To compose our related work, we employed Google Scholar to search for relevant publications using the terms "*lessons learned*" and "*mobile health monitoring*". Though this search strategy does not cover a wide range of terms, it found papers appropriate to situate the reader about what has been developed in this area. The similarity to our proposal was applied as the primary filter, and the papers were sorted by Google Scholar relevance metric.

Aranki *et al.* (2016) present a physical activity monitoring system for patients with chronic heart failure. Similar to our work, they conducted a pilot study with 15 participants in the real world. Among the main lessons learned, the authors highlight that the behavior of patients is neither static nor uniform and that patients tend to suffer fatigue in using technology. In addition, they discuss aspects related to battery consumption and the privacy of sensitive data. The main difference between this study and ours is that the data were collected only from smartphones that should be located on the right hip at the waistline level, which is not typical for users.

Bravo *et al.* (2018) describe mobile health as an emerging field capable of transforming how people manage their health. In this work, the authors discuss lessons from the experiences obtained from mHealth development by the MAmI Research Lab. However, unlike our work, the lessons focus on developing and representing data in mHealth systems. Also, the experiences are diluted throughout the sections.

L'Hommedieu *et al.* (2019) provide recommendations for conducting longitudinal sensor-based research using both environmental sensors and wearables in healthcare settings. Among the recommendations, it is possible to highlight the need to build trust with the key stakeholders and volunteers and monitor the data collected to identify possible issues in the sensors. Although this work is similar to ours, the recommendations presented in this paper are complementary and could compose a more comprehensive set of recommendations.

Finally, Gjoreski *et al.* (2021) systematically compare machine learning approaches when applied to cognitive load monitoring with wearables and summarize the learnings related to a machine learning challenge. The recommendations presented by the authors are relevant since there is a trend in using intelligent algorithms to provide mHealth services.

3 STUDY DESIGN

In order to understand the lessons presented as the result of this paper, it is essential to figure out how our longitudinal study was conducted.

This evaluation aimed to analyze the QoL inference process in physical and psychological domains, using data collected from smartphones and commercial wearables. These two QoL domains were chosen from the observation that a large amount of data collected by mobile devices can provide insights into the users' QoL (Ghosh et al., 2022). The physical domain assesses motor facets such as daily activities, medicines dependence, mobility, sleep quality, and work capacity. The psychological domain is related to body image, negative and positive feelings, self-esteem, and other mental health aspects (Orley and Kuyken, 1994).

The evaluation was conducted to assess the feasibility of Quality of Life inference concerning errors (Mean Absolute Error and Root Mean Squared Error) obtained by the machine learning regressors using as a reference value the WHOQOL-BREF in the context of independent adults.

3.1 Participants

Thirty adults were invited to participate as volunteers given the following criteria:

a. age between 18 and 65 years;

b. prior knowledge in the use of smartphones;

c. availability for continuous use of wearables.

However, only 21 completed the study. Seven accepted our terms but did not start due to lack of availability or devices' incompatibility (*e.g.*, iOS devices). In addition, one participant dropped out after the initial setup reporting that he/she could not use the wearable continuously, and another dropped out in the middle of the study because he/she had a wrist allergy.

The participants' invitation prioritized members of our research laboratory (due to COVID-19 restrictions) and those who had a smart band or smartwatch. This last criterion was essential to reduce costs. Therefore, after accepting the invitation, the procedure for starting the study had six steps:

1. agreeing to the informed consent form;
2. answering the WHOQOL-BREF supported by the responsible researcher to clarify possible issues;
3. configuring the wearable to sync data;
4. installing the QoL Monitor app;
5. granting permissions to monitor health data;
6. effectively initiates monitoring.

After this initial procedure, participants were instructed to follow their activities normally.

The final profile of these participants comprises 15 men and six women aged between 19 and 47. Almost half of the participants are single, and the other half are married. Most of them have postgraduate degrees and are full-time workers. Regarding income, ten (10) participants reported receiving between 2 and 4 minimum wages (Brazilian minimum wage R$ 1,100 was used as the reference), and all claimed to live in an urban area. Regarding the family arrangement, most participants live with 1 or 2 more people at home, and there are two large groups in terms of the number of children: those who do not have children (12) and those who have 1 or 2 children (9).

3.2 Data Collection

Data were collected daily and sent anonymously to the cloud (Figure 2). Weekly, the QoL Monitor app (developed for this work) warned the participant to answer the WHOQOL-BREF only with questions about the physical and psychological domains. This data was also sent anonymized to the cloud.

Figure 3 highlights the data collected. Socio-demographic and anthropometric data are needed to understand the characteristics of the users. The other raw data directly correlates with the physical and psychological QoL domains. Also, all of them can be obtained through common devices such as smart bands and smartwatches. Additionally, it is worth mentioning that the location data only stores the number of points visited throughout the day, *i.e.*, the application does not record the specific places. The same logic was applied to the WiFi Networks field. This strategy was adopted to preserve the users' privacy.

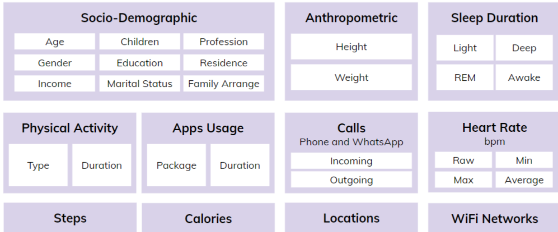

Figure 3: Raw data collected from users.

Figure 4 puts light on how the training instances are created. A sample has as predictors all data collected from 18:00 of the previous day to 17:59 of the current day. We decided to use this time slot because the last night's sleep directly impacts the current day's activities (Arora et al., 2020). The value to be predicted is obtained after answering the questionnaire on Sundays. As the user must answer this questionnaire considering the past week, we can use this value as a reference. Naturally, during the data collection, some issues can arise (*e.g.*, absence of network or battery issue). In this case, if the data is not recorded, such intervals do not generate new training instances.

3.3 Operation

After obtaining the raw data, preprocessing activities are performed to prepare our dataset. Among

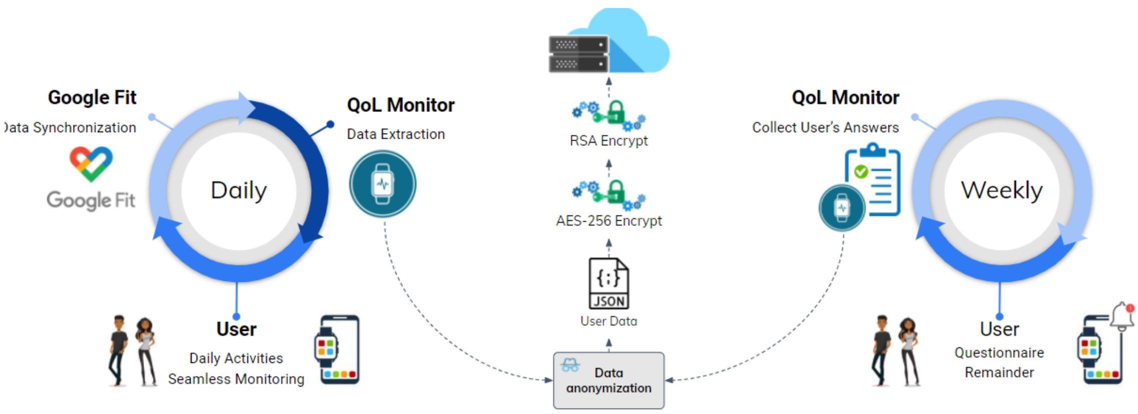

Figure 2: Data flow to collect health measures and self-reported QoL questionnaires.

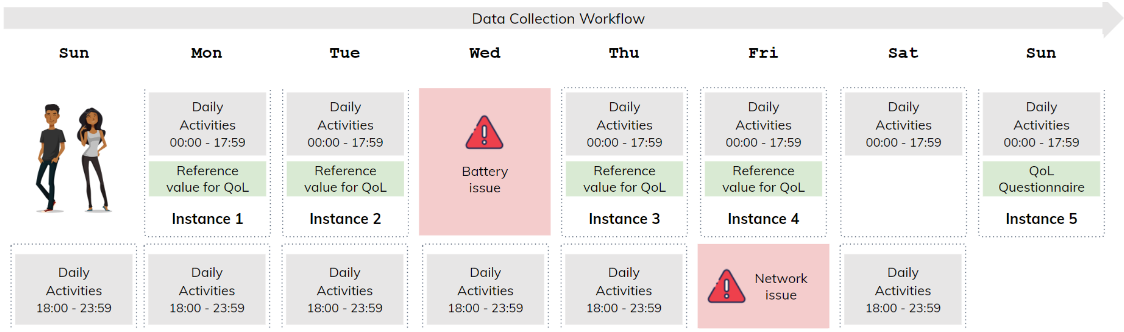

Figure 4: A representation of how the instances are created.

these activities are removing inconsistencies and outliers, data stratification, categorical variables encoding, data sync, and computation of QoL scores based on the questionnaire responses. Thus, at the end of the data preparation, two datasets are obtained: a dataset in which the last column is the QoL score for the physical domain and a similar dataset for the psychological domain. The last column changes because it is used as a reference for the learning process.

For the Machine Learning modeling, we decided to use the Scikit-learn toolbox (Pedregosa et al., 2011) due to its high acceptance in the scientific community and the consistency of its results (Tanaka et al., 2019) (Géron, 2019). Then, four algorithms were selected based on Géron (2019) guidelines: Linear Regressor, Decision Tree Regressor, Random Forest Regressor, and GBoost Regressor.

The first algorithm search for linear relationships within the dataset. It is considered a simple model and an excellent choice to start investigating regression problems (Ian and Eibe, 2005). The second algorithm is robust compared to linear regression and can find nonlinear relationships in the data. The third algorithm uses the concept of random forests to train multiple decision trees. This algorithm performs well for a wide variety of problems (Paul et al., 2018). Finally, the last algorithm uses gradient descent to minimize the error function. In addition, this investigation did not explore the hyperparameters strategy, and only the default parameters were used.

We selected three metrics to assess our results: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), to evaluate the error of the inference algorithm and the time in seconds to estimate the computational resource needed for training. MAE measures how far the predictions are from the correct output, and RMSE measures the square root of the square of the differences between the predicted and accurate values. The latter is one of the most used metrics to evaluate regressors (Ian and Eibe, 2005).

After data collection and processing, we performed a 10-fold cross-validation of four Machine Learning techniques using the Scikit-learn toolbox.

4 RESULTS

Tables 1 and 2 summarize the results achieved for each of the study metrics.

Table 1: Results for the physical QoL dataset.

ML Techniques	**Physical Dataset**	
	MAE	RMSE
Linear Regression	6.5866 ± 1.7582	8.8457 ± 2.9102
Decision Tree	6.1465 ± 1.6188	9.3685 ± 2.7071
Random Forest	4.9477 ± 1.5283	7.2215 ± 3.0008
GBoost	4.9569 ± 1.4472	6.9191 ± 2.6899

In both datasets, the training time grows as the classifier complexity increases. Naturally, the errors tend to decrease with more robust classifiers. However, there are exceptions. For example, the model created by Random Forest for the Psychological dataset had more minor errors than the metrics obtained by GBoost. In this case, we can conclude that the Random Forest and GBoost algorithms presented similar results but with very different computational costs. GBoost takes much longer to train. Thus, we decided to use Random Forest as our reference.

Table 2: Results for the psychological QoL dataset.

ML Techniques	**Psychological Dataset**	
	MAE	RMSE
Linear Regression	8.1918 ± 1.9133	10.6146 ± 2.4728
Decision Tree	5.8000 ± 1.7678	9.5880 ± 2.3525
Random Forest	4.6830 ± 1.2204	6.8838 ± 2.2436
GBoost	4.9707 ± 1.3524	7.0034 ± 2.2327

Using MAE and RMSE, we can state that the error obtained by the classifiers is reasonable on a scale that varies from 0 to 100. For example, considering the RMSE metric for the physical dataset, Random Forest has a mean error of 7.2215 with a standard deviation of 3.0008.

Figure 5 brings the prediction error plot for the Random Forest regressor considering both datasets. This kind of plot has the actual (represented by the x-axis) and predicted (represented by the y-axis) values generated by the model. Thus, it is possible to analyze the model variance. For example, the 45° degree gray line (identity) represents a perfect scenario where the predictor perfectly matches the actual values. In our case, the prediction errors are distributed close to this line, with few outliers. The graph also contains the black line with the best fit obtained by the regressor.

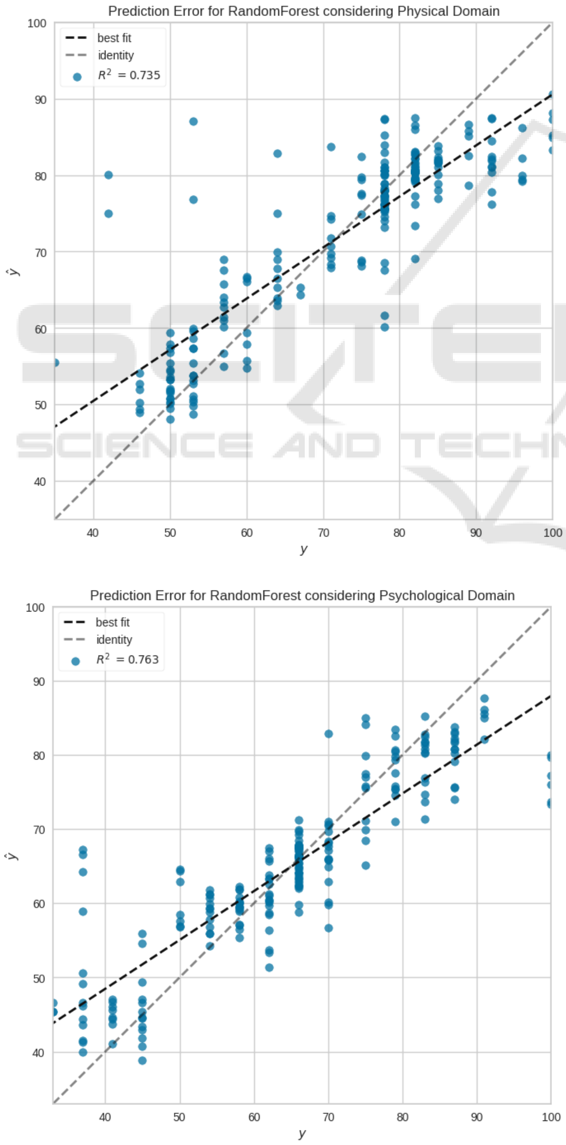

Figure 5: Prediction error plots for RandomForest.

Furthermore, it is possible to observe (in Figure 5) that the model obtained R_2 equal to 0.735. This metric evaluates the performance of regressors considering the percentage of the sum of errors concerning the mean error. In the worst case, R_2 is equal to zero, and in the best scenario, it is equal to 1. This explanation reinforces the claim that the results obtained in this investigation are satisfactory. We argue that the results should improve once we get a more robust database and algorithms with adjusted parameters.

5 TAM EVALUATION

At the end of the participant monitoring period, we decided to apply a final survey developed based on TAM3 (Technology Acceptance Model 3) (Venkatesh and Bala, 2008) to collect feedback about the study and concerning the tool used to monitor Quality of Life. TAM helps in understanding aspects related to the adoption of new technologies.

Thus, the applied questionnaire was subdivided into four groups, each exploring an aspect. The aspects analyzed were: i) perceived usefulness; ii) perceived ease of use; iii) self-efficacy when using the tool; iv) intention to use the tool. Five possible alternatives for each question were: a) I fully agree; b) I partially agree. c) neutral; d) partially disagree; e) I totally disagree. In the end, an open question was included to include perceptions about the study.

Figure 6 presents the quantitative results of the participants' answers. It is worth mentioning that the questionnaire was administered anonymously and that only 13 of the 21 participants responded.

Regarding perceived usefulness, most participants agreed that the QoL Monitor tool – previously described by Oliveira *et al.* (2022c) – is helpful for QoL monitoring. However, there was a partial disagreement regarding the reduction in monitoring cost, probably associated with the user's need to use some wearable to complement the data collected.

As for the perceived ease of use, most volunteers considered the interaction clear and did not require much mental effort. This result was expected because the app was designed to simplify user interactions.

The third aspect observed was self-efficacy. In this aspect, the aim was to assess the users' ability to use the tool in situations with little or no prior instruction. In the results, it was possible to observe that some users disagreed about the possibility of being able to monitor their Quality of Life only with the support of the tool or having used similar tools. Therefore, this shows that initial training is necessary for people to understand aspects related to QoL monitoring.

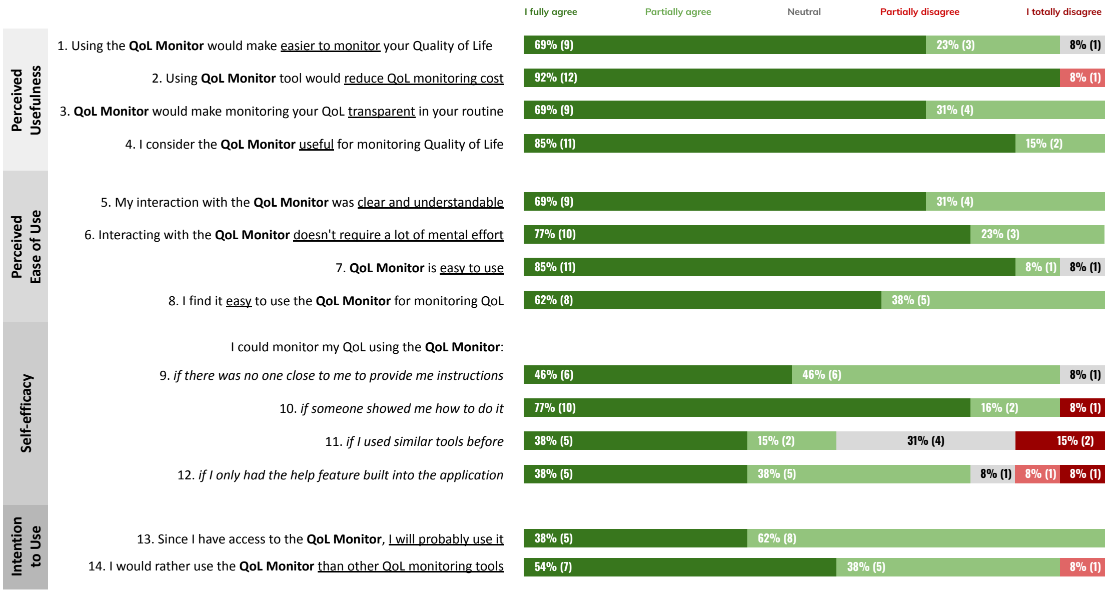

Figure 6: Results of the TAM questionnaire.

To conclude the quantitative results, most participants stated that they would use QoL Monitor again instead of other similar tools.

In addition to the quantitative results, we summarized some qualitative perceptions of the volunteers regarding the difficulties they faced throughout the study. Such perceptions were organized into three groups: discomfort, privacy, and access to data. Figures 7 and 8 present the original comment in Portuguese and its translation into English.

Regarding discomfort, the participants reported difficulties in using the device uninterruptedly and keeping the routine of filling in the surveys. In addition, on some devices, users reported issues with receiving notifications due to restrictive policies concerning background apps. For sure, this discomfort can bias the collected data. Therefore, it is essential to have strategies to reduce it.

Some participants reported that data collection was a bit invasive. This perception is probably related to the large amount of requested data and the need to grant many permissions. However, machine learning models would only perform satisfactorily with this massive data. Therefore, we have comforted participants about data privacy using anonymization and encrypted request.

Finally, some participants reported issues in extracting the data from wearables. Usually, commercial wearables do not deliver methods to access their data. Thus, we decided to extract data through the Google Fit platform. Nevertheless, some wearables apps did not allow native integration with Google Fit, requiring a third-party app to extract this information. The complexity of this process frustrated volunteers who used Samsung devices.

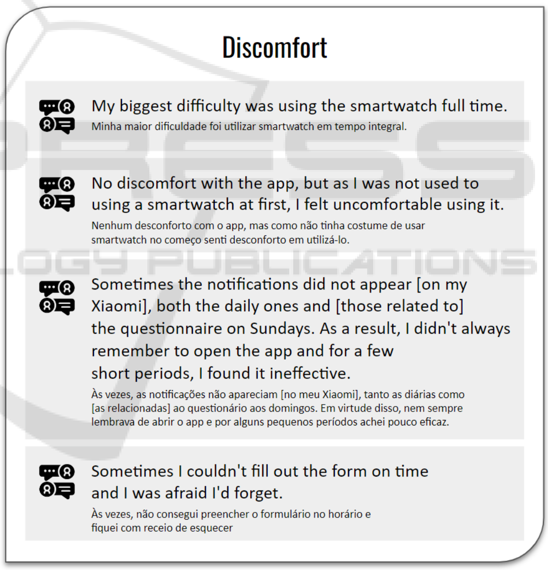

Figure 7: Participants' comments about discomfort.

6 CHALLENGES

Facing challenges and limitations are common in the empirical studies (Wohlin et al., 2012), and their discussion reinforces the work's reliability since it is presented the main issues and strategies to mitigate them. Also, this discussion represents a valuable contribu-

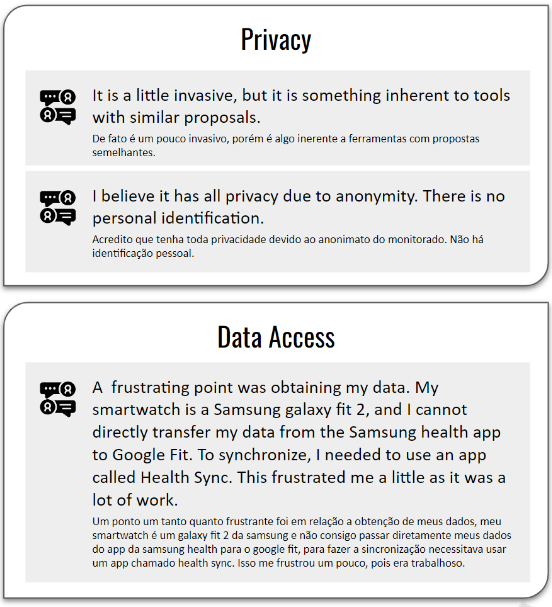

Figure 8: Participants' comments regarding privacy and data access.

tion to researchers and practitioners who work or wish to work in this investigation area. The scenario of this paper has particular value due to the high interest in developing mobile health monitoring solutions (Oliveira et al., 2022a).

Based on the challenges discussed in this section, it is possible to anticipate or even avoid issues when conducting this kind of study. Therefore, this section has organized the challenges discussion based on the evaluation phases.

The conducting phase had many challenges. The first was related to the participants' recruitment. We decided to recruit thirty (30) adults between the ages of 18 and 65 since they usually have prior knowledge of using smartphones and smartwatches. Due to the restrictions imposed by the COVID-19 pandemic, we initially sought out members of our research group (GREat/UFC) who already own a smart band or smartwatch. Thus, the recruitment process could be completely remote. However, only six participants met such restrictions. Then, it was necessary to expand the recruitment to close people (considering our social network). Even so, that number only increased to nine participants.

Thus, purchasing some devices (Xiaomi Mi Band) and sending them to interested participants was necessary. Furthermore, the shipping and delivery logistics delayed the start of data collection and increased the study cost. Finally, some recruited participants withdrew after the initial presentation, citing lack of time and others having a smartphone incompatible with our app (*e.g.*, phones with the iOS system). Therefore, despite our efforts, this evaluation's relatively low number of participants (21) is a limitation that should be addressed in future studies.

After recruiting participants, we started collecting data. During this step, we faced many issues related to noise in data collection. For example, the absence of an Internet connection when sending data, devices without battery charge, sensors turned off, and sensors or devices with different levels of accuracy. These situations caused noise in our registry, making it difficult to clean the data.

Regarding failures in sending data to the cloud, it was necessary to implement a mechanism in order to perform retries on the connection and, after five failed attempts, internally store the data for sending in the next day. As for the disconnecting sensors, we warned the participants about the continuous use of the devices and about the need to keep at least GPS active. However, we received feedback from participants that the seamless GPS use increased battery consumption. Therefore, some participants turned it off in moments of low battery. Consequently, it was necessary to filter inconsistencies during data processing.

Regarding charge frequency, we guided the participants to charge their smartphones daily and their wearables weekly. However, when charging the devices, a data gap is created. Then, such gaps were removed from the study.

Another area for improvement in our evaluation is the non-standardization of using the same device for monitoring. In an ideal scenario, all volunteers should use the same device model to reduce inconsistencies regarding the quality of the collected data. For example, two different smart bands may differ in the number of steps recorded due to the detection algorithm. However, budget limitations prevented all participants from using the same devices. Even so, we decided to proceed with the investigation because we understand that it is impossible to guarantee that all users will use similar devices in the real world.

We have also faced difficulties in user engagement. Achieving and maintaining engagement in healthcare technologies is such a complex challenge. Because of that, many studies have been conducted to find proper strategies to keep users active in an organic way (Wang et al., 2022) (Ganesh et al., 2022).

Our study faced engagement challenges, as users had to follow a series of recurring actions, such as opening a wearable app daily to ensure data sync and weekly answering the QoL questionnaire. Even with the application's support to remind these actions, we observed that at least one-third of the participants failed to perform the questionnaire more than once.

When investigating what could be happening, some participants reported that day-to-day activities made them forget. It was also clear that despite recognizing the benefits of daily health monitoring, many participants ignored them and forgot to access the app's notifications. This challenge needs further investigation to understand the real reasons for this lack of engagement and what strategies can be used to overcome it.

We also received several reports of problems classified as "real-world issues". For example, one participant reported that he lost the smart band during a bath in the sea; two participants reported that the smart band was causing a wrist allergy and, therefore, they had to stop using it for a few days. Another participant caught chikungunya, which affected his joints, preventing him from using the wearable for a few days. These problems are inherent to "in the wild" studies, and there are few strategies to avoid such situations. We adopted a specific approach for each of them, but, in general, they all generated data gaps that were eliminated during preprocessing.

In the analysis of the results, we faced a data variability issue. This issue happens because the profile of study participants has little variability. So, many records have intermediate QoL scores and few high or low scores. Consequently, this impairs the regressors' ability to generalize.

We expect to conduct a new assessment with more participants (up to 100 members) to address this limitation, varying the subjects' profiles. Also, we explored a few algorithms (only four), as this was just a Proof-of-Concept study. However, we understand that it is necessary to expand the number of evaluated Machine Learning techniques and include many repetitions (not only k-fold validation) in the experiments to perform statistical tests.

7 LESSONS LEARNED

This section presents and discusses ten lessons learned from conducting a longitudinal investigation to monitor the Quality of Life using the Internet of Health Things and Machine Learning. The lessons were organized with a title, a short description, and alternatives to overcome it. Finally, it is presented a 5W1H table to summarize this discussion.

- **Study Design needs to be Carefully Validated**

The planning phase is crucial for adequately conducting health monitoring studies "in the wild". It is also decisive in the approval by the ethics committee. On the other hand, the absence of a rigorous planning process can invalidate the data collected and increase research costs. Thus, a possible alternative to validate the planning is to conduct pilot studies. According to Van Teijlingen *et al.* (2010), pilot studies refer to mini versions of a full-scale investigation, and they can identify potential practical problems in the research procedure.

- **Data Privacy Must be a Priority**

Currently, laws and regulations protect digital health users from mishandling data (Purtova et al., 2015). In this sense, privacy must be prioritized to create trust with the volunteers. Moreover, from the feedback collected in our qualitative assessment, it became evident that participants will be hesitant to use an invasive technology without a robust process for keeping their data secure. In this regard, a good alternative is to use data anonymization (Sneha and Asha, 2017). Another option is to avoid using data that makes it possible to re-identify the user, such as location or Internet access data.

- **Volunteer Engagement Requires Attention from Beginning to End of the Study**

Recruiting participants is not easy; keeping their engagement is even more problematic. Thus, research involving health data monitoring has the significant challenge of finding volunteers. Regarding this challenge, a helpful strategy is establishing partnerships with universities or health centers and making key people in these organizations aware of the work's relevance. These people should become ambassadors to attract volunteers. In addition, it is crucial to consider strategies to keep volunteers committed until the end of the work, for example, rewarding students who remain active or even gifting a wearable to those with high engagement.

- **The Technology Discomfort can be a Bias**

Monitoring health data requires sensors (Rodrigues et al., 2018). Such sensors can be wearable like smart bands and smart rings, personal devices such as smartphones, or even instruments fixed in the environment such as cameras and infrared sensors. During planning, the researcher should decide which sensors will be used and how to collect the data (using native apps, for example). For this decision, it must be taken into account possible discomforts for the users. For example, even commercial devices already established on the market, such as smart bands, can provoke wrist-related allergies. Thus, it is recommended to investigate whether the participants are already used to the selected technology to avoid discomfort and, consequently, bias in the data.

- **The Project Budget needs to be Considered when Selecting Devices for Monitoring**

As stated before, the researcher must select sensors for data collection during planning. Among the criteria for this selection are the number, variety, and accuracy of measurements, battery consumption, ease of use, market availability, data access, and durability. However, while conducting our case study, we realized that the project budget is a vital criterion in this selection. In general, most volunteers do not have these devices, and even for those that have, there is the problem of non-standardization since different brands and models can cause inaccurate data. In addition, as this type of study requires many participants, our strategy was to opt for a low-cost device that would allow us to obtain the necessary measurements. Thus, it would be possible to include a more significant number of participants. Therefore, we opted for the Xiaomi Mi Smart Band, which costs approximately 39 dollars in Brazil.

- **Extracting Data from Wearables is Complex**

A significant challenge for those who work with wearable devices is data extraction (Oliveira et al., 2022b). If the researcher chooses to build their own device, this new technology can face many additional issues due to the lack of maturity. On the other hand, few commercial wearables have methods for extracting data. Furthermore, those commercial wearables that share Web APIs to retrieve data tend to have a higher cost, such as smartwatches with Android Wear or Fitbit devices. An alternative is to look for devices that allow data synchronization with platforms such as Google Fit (for Android) and HealthKit (for iOS) (Oliveira et al., 2022a). Such platforms were designed to centralize users' health information and have well-documented APIs for data extraction.

- **Data Collected "in the wild" Always has Noise**

Monitoring patients in a controlled environment allows the researcher or professional to establish the required minimum parameters for the system. It is possible, for example, to guarantee that the devices will always have access to the Internet or even that there will be no lack of battery supply. On the other hand, collecting health data in real life (uncontrolled environment) implies noise in the data. For example, data gaps will be generated when removing devices for charging. Also, synchronization problems can make it impossible to record specific measures. Again, another situation that can occur in studies that include self-reported surveys is the user forgetting to answer the survey. In addition to these examples, a wide variety of other situations can occur; unfortunately, it is impossible to avoid them all. Therefore, a suitable strategy to deal with these issues is to intensify the effort dedicated to data cleaning and processing. This step is crucial to remove noise.

- **Constant Internet Access Cannot be a Premise**

As stated by Rodrigues *et al.* (2018), IoHT architecture for healthcare monitoring systems involves collecting data by sensors and sending it to robust nodes for processing and analysis. It is common for these nodes to be at the edge or in the cloud. However, the periodic sending of data cannot presuppose continuous access to the Internet. In uncontrolled environments, it is common to have unavailable access for a while, resulting in failures in sending data. In this way, it is essential to implement strategies for resending in case of failure or even temporary storage for later sending. This strategy should prevent information loss.

- **Getting Feedback Should be Uninterrupted**

After recruiting the participants, we held a session to explain the study operation, configure the devices, clarify doubts and sign the informed consent form. On this occasion, we made it clear that the participants were free to withdraw at any time and that we would be available to obtain feedback throughout the monitoring period. Unfortunately, only some volunteers kept the practice of continuous feedback. In our case, only using the final evaluation questionnaire was possible to extract qualitative data, and probably some details may have been lost. Thus, it is crucial to encourage volunteers to provide periodic feedback. For future studies, we will leave an anonymous form open from the beginning to the end of the research and ask them to keep feeding whenever they face a positive or adverse situation.

- **Unexpected Problems Should Arise**

Finally, the researcher must be prepared for unexpected issues. For example, a device being stolen from the user or even a volunteer getting sick and having to withdraw. Unfortunately, there is no specific approach to dealing with these problems. However, it is essential to keep the research team watchful to reverse them as soon as possible.

Table 3 summarizes the lessons learned in this study using the 5W1H model (in this case, who and where were suppressed because the research team always conducts activities and location is not applied).

8 FINAL REMARKS

This paper presents ten (10) practical lessons from mHealth monitoring of twenty-one (21) volunteers over three months. The main contribution of these lessons is that future studies can use this knowledge

Table 3: Summarized lessons learned.

What	When	Why	How
Study design needs to be validated	Planning	It can increase project costs	Conducting pilot studies
Data privacy must be a priority	Planning and Recruiting	It can hamper recruitment, in addition to legal issues	Anonymizing data and making privacy policies clear
Volunteer engagement requires attention	From the beginning to the end	It can lead volunteers to withdraw	Encouraging volunteer participation
The technology discomfort can be a bias	Planning, and Conduction	It can insert bias in data	Selecting usual technologies
Project budget needs to be considered when selecting devices	Planning	It can increase project costs	Weighing the cost against the device resources required by the study
Extracting data from wearables is complex	Conduction	Without data, there is no health monitoring	Using health data hub platforms like Google Fit and Health Kit
Data collected "in the wild" has noise	Conduction	Noise can lead to inaccuracies	Investing in cleaning and processing activities
Constant Internet access cannot be a premise	Planning, and Conduction	It can cause data loss	Implementing data sending retries and data staging
Getting feedback should be uninterrupted	From the beginning to the end	To avoid missing relevant feedback	Allowing continuous sending of anonymous feedback
Unexpected problems should arise	Conduction	To ensure proper study conduction	Keeping the research team on its toes

to plan how to mitigate common issues related to mHealth monitoring "in the wild" (*i.e.*, uncontrolled environments). As future work, we expect to refine this learning by replicating this study.

ETHICAL APPROVAL

Our project (nº 56153322.0.0000.5054) was approved by the coordination of the ethics committee located at the Federal University of Ceará (UFC). Furthermore, it complied with CONEP laws and followed all the international ethical standards. Finally, all volunteers signed the informed consent document.

DATA AVAILABILITY

The QoL dataset is proprietary of GREat lab and is not available for public use yet. However, the project submitted to the ethics committee, the notebooks used to process data, and the higher-resolution images are available in https://bit.ly/3SSlkqs.

ACKNOWLEDGMENTS

The authors would like to thank National Council for Scientific and Technological Development (CNPQ) for the Productivity Scholarship of professor Rossana Andrade DT-1D (N^o 306362 / 2021-0) and professor Pedro Santos Neto DT-2 (N^o 315198 / 2018-4).

REFERENCES

Akter, S. and Ray, P. (2010). mhealth-an ultimate platform to serve the unserved. *Yearbook of medical informatics*, 19(01):94–100.

Aranki, D., Kurillo, G., Yan, P., Liebovitz, D. M., and Bajcsy, R. (2016). Real-time tele-monitoring of patients with chronic heart-failure using a smartphone: lessons learned. *IEEE Transactions on Affective Computing*, 7(3):206–219.

Araújo, I., Pereira, M. B., Silva, W., Linhares, I., Marx, V., Andrade, A. M., Andrade, R. M., and de Castro, M. F. (2022). Machine learning and cloud enabled fall detection system using data from wearable devices: Deployment and evaluation. In *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 413–424. SBC.

Arora, A., Chakraborty, P., and Bhatia, M. (2020). Analysis of data from wearable sensors for sleep quality estimation and prediction using deep learning. *Arabian Journal for Science and Engineering*, 45(12):10793–10812.

Bostrom, J., Sweeney, G., Whiteson, J., and Dodson, J. A. (2020). Mobile health and cardiac rehabilitation in older adults. *Clinical Cardiology*, 43(2):118–126.

Bravo, J., Hervás, R., Fontecha, J., and González, I. (2018). M-health: lessons learned by m-experiences. *Sensors*, 18(5):1569.

Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., and Schreier, G. (2010). The internet of things for ambient assisted living. In *7th int. conf. on information technology: new generations*, pages 804–809. IEEE.

Elkinton, J. R. (1966). Medicine and the quality of life. *Annals of Internal Medicine*, 64:711–714.

Ganesh, D., Balaji, K. K., Sokkanarayanan, S., Rajan, S., and Sathiyanarayanan, M. (2022). Healthcare apps for post-covid era: Trends, challenges and potential opportunities. In *2022 IEEE Delhi Section Conference (DELCON)*, pages 1–7. IEEE.

Géron, A. (2019). *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow*. Alta Books, Rio de Janeiro.

Ghosh, S., Löchner, J., Mitra, B., and De, P. (2022). Your smartphone knows you better than you may think: Emotional assessment 'on the go'via tapsense. *Quantifying Quality of Life: Incorporating Daily Life into Medicine*, page 209.

Gjoreski, M., Mahesh, B., Kolenik, T., Uwe-Garbas, J., Seuss, D., Gjoreski, H., Luštrek, M., Gams, M., and Pejović, V. (2021). Cognitive load monitoring with wearables–lessons learned from a machine learning challenge. *IEEE Access*, 9:103325–103336.

Hao, T., Walter, K. N., Ball, M. J., Chang, H.-Y., Sun, S., and Zhu, X. (2017). Stresshacker: towards practical stress monitoring in the wild with smartwatches. In *AMIA Annual Symposium Proceedings*, volume 2017, page 830. American Medical Informatics Association.

Ian, H. W. and Eibe, F. (2005). Data mining: Practical machine learning tools and techniques.

Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., and Kwak, K.-S. (2015). The internet of things for health care: a comprehensive survey. *IEEE access*, 3:678–708.

Junior, E. C., de Castro Andrade, R. M., and Rocha, L. S. (2021). Development process for self-adaptive applications of the internet of health things based on movement patterns. In *9th International Conference on Healthcare Informatics (ICHI)*, pages 437–438. IEEE.

Karimi, M. and Brazier, J. (2016). Health, health-related quality of life, and quality of life: what is the difference? *Pharmacoeconomics*, 34(7):645–649.

L'Hommedieu, M., L'Hommedieu, J., Begay, C., Schenone, A., Dimitropoulou, L., Margolin, G., Falk, T., Ferrara, E., Lerman, K., Narayanan, S., et al. (2019). Lessons learned: Recommendations for implementing a longitudinal study using wearable and environmental sensors in a health care organization. *JMIR mHealth and uHealth*, 7(12):e13305.

Oliveira, P., Costa Junior, E., Santos, I. D. S., Andrade, R., and Santos Neto, P. d. A. (2022a). Ten years of ehealth discussions on stack overflow. *International Conference on Health Informatics (HEALTHINF1'22)*.

Oliveira, P. A. M., Andrade, R. M. C., Neto, P. S. N., and Oliveira, B. S. (2022b). Internet of health things for quality of life: Open challenges based on a systematic literature mapping. In *15th International Conference on Health Informatics (HEALTHINF)*. INSTICC.

Oliveira, P. A. M., Andrade, R. M. C., Neto, P. S. N., and Oliveira, B. S. (2022c). Towards an ioht platform to monitor qol indicators. In *15th International Conference on Health Informatics (HEALTHINF)*. INSTICC.

Orley, J. and Kuyken, W. (1994). The development of the world health organization quality of life assessment instrument (the whoqol). In *Quality of life assessment: International perspectives*, pages 41–57. Springer Berlin Heidelberg, Berlin, Heidelberg.

Palos-Sanchez, P. R., Saura, J. R., Martin, M. Á. R., and Aguayo-Camacho, M. (2021). Toward a better understanding of the intention to use mhealth apps: Exploratory study. *JMIR mHealth and uHealth*, 9(9):e27021.

Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintha, A. R., and Kundu, S. (2018). Improved random forest for classification. *IEEE Transactions on Image Processing*, 27(8):4012–4024.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Purtova, N., Kosta, E., and Koops, B.-J. (2015). Laws and regulations for digital health. In *Requirements Engineering for Digital Health*, pages 47–74. Springer.

Qudah, B. and Luetsch, K. (2019). The influence of mobile health applications on patient-healthcare provider relationships: a systematic, narrative review. *Patient education and counseling*, 102(6):1080–1089.

Rodrigues, J. J., Segundo, D. B. D. R., Junqueira, H. A., Sabino, M. H., Prince, R. M., Al-Muhtadi, J., and De Albuquerque, V. H. C. (2018). Enabling technologies for the internet of health things. *Ieee Access*, 6:13129–13141.

Sanchez, W., Martinez, A., Campos, W., Estrada, H., and Pelechano, V. (2015). Inferring loneliness levels in older adults from smartphones. *Journal of Ambient Intelligence and Smart Environments*, 7(1):85–98.

Sneha, S. and Asha, P. (2017). Privacy preserving on e-health records based on anonymization technique. *Global Journal of Pure and Applied Mathematics*, 13(7):3367–3380.

Tanaka, K., Monden, A., and Zeynep, Y. (2019). Effectiveness of auto-sklearn in software bug prediction. *Computer Software*, 36(4):46–52.

Van Teijlingen, E., Hundley, V., et al. (2010). The importance of pilot studies. *Social research update*, 35(4):49–59.

Varshney, U. (2014). Mobile health: Four emerging themes of research. *Decision Support Systems*, 66:20–35.

Venkatesh, V. and Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2):273–315.

Wang, Z., Xiong, H., Zhang, J., Yang, S., Boukhechba, M., Zhang, D., Barnes, L. E., and Dou, D. (2022). From personalized medicine to population health: A survey of mhealth sensing techniques. *IEEE Internet of Things*.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.