



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

mHealth Application Based on Artificial Intelligence for Decision Support System

Ana Beatriz da Cruz Baptista Galamba Vieira

Master's in integrated Decision Support Systems

Supervisor:

PhD José Miguel de Oliveira Monteiro Sales Dias, Full Professor,
Departamento de Tecnologias Digitais Aplicadas

Co-Supervisor:

MSc Luís Manuel Nobre de Brito Elvas, Invited Assistant,
Departamento de Ciências e Tecnologias da Informação

September, 2023



TECNOLOGIAS
E ARQUITETURA

ISTA – School of Technology and Architecture

mHealth Application Based on Artificial Intelligence for Decision Support System

Ana Beatriz da Cruz Baptista Galamba Vieira

Master's in integrated Decision Support Systems

Supervisor:

PhD José Miguel de Oliveira Monteiro Sales Dias, Full Professor,
Departamento de Tecnologias Digitais Aplicadas

Co-Supervisor:

MSc Luís Manuel Nobre de Brito Elvas, Invited Assistant,
Departamento de Ciências e Tecnologias da Informação

September, 2023

Acknowledgements

I would like to express my heartfelt appreciation to my supervisors, Professor José Miguel Dias, and Professor Luís Brito Elvas, for granting me the privilege of participating in the AIMHealth project and for guiding and supporting me throughout the course of this dissertation.

I would also like to express my gratitude to Professor João Carlos Ferreira for his knowledge-sharing and unwavering availability throughout the past two years. To Professor Luís Rosário, Dr. Elsa Fragoso and Dr. Paula Esteves for their invaluable contributions to the app. Their testing efforts and insights provided a valuable perspective from the hospital environment, enriching the development process.

To my family, whose warmth, support, and inspiration have been instrumental in helping me achieve this personal milestone. To my friends, whose presence and joy they bring have expanded and enriched my world. To my friend Beatriz Guedes for being a constant presence throughout my journey over the past five years and particularly during this dissertation.

I would also like to extend my gratitude to my former AIMHealth colleagues, Matilde Cascalho and Riccardo Zichittella, for their assistance and guidance in helping me navigate the intricacies of this project.

Lastly, I would like to express my appreciation to my pets, Beanie, Dobby, Luna, and Woskie, for keeping me company and bringing joy to this journey.

My sincere gratitude to all those who provided their support during this phase of my life. Thank you!

Resumo

A monitorização remota surgiu como uma solução valiosa, particularmente no contexto da Covid-19, uma vez que permite oferecer serviços de saúde acessíveis a uma população mais ampla. Através da recolha de dados fisiológicos e informações de saúde do utilizador, a tecnologia mHealth pode ser utilizada para monitorizar pacientes com doenças crónicas, detetar anomalias e prever eventos. O objetivo principal desta dissertação foi desenvolver mais uma aplicação de mHealth baseada em Inteligência Artificial, a aplicação AIMHealth. Para cumprir este objetivo, foram criados dois artefactos distintos: "Artefacto 1 - AIMHealth App para Recolha de Dados" e "Artefacto 2 - Análise de Dados para a AIMHealth App", cada um abordando uma pergunta de investigação diferente. O primeiro artefacto, que se concentrou em melhorar a experiência do utilizador na aplicação, através do emprego de princípios de design centrados no utilizador, recebeu avaliações positivas por parte dos médicos. Em contraste, o segundo artefacto que visou identificar padrões nos dados de saúde, obteve resultados inconclusivos.

Palavras-chave: Saúde Móvel, Aprendizagem Automática, Desenvolvimento de Aplicações Móveis, Monitorização Remota de Pacientes, Análise de Dados, Experiência de Utilizador

Abstract

Remote monitoring has emerged as a valuable solution, particularly in the context of the Covid-19, pandemic due to its capacity to provide accessible healthcare services to a larger population. By collecting user's physiological data and health information, mHealth technology can be utilized to monitor patients with chronic conditions, detect anomalies and forecast events. The primary objective of this dissertation was to further develop an Artificial Intelligence-based mHealth application, called the AIMHealth app. To accomplish this objective, two distinct artifacts were created: "Artifact 1 - AIMHealth App for Data Collection" and "Artifact 2 - Data Mining for the AIMHealth App", each addressing a different research question. The first artifact, which focused on improving the user experience within the app, through the application of user-centered design principles, received positive evaluations by physicians. In contrast, the second artifact aimed to identify meaningful patterns in health data and followed the principles of the cross-industry standard process for data mining, however the results obtained were inconclusive.

Keywords: mHealth, Machine Learning, Mobile Application Development, Remote Patient Monitoring, Data Mining, User Experience

Contents

Acknowledgements	v
Resumo	vii
Abstract	ix
List of Abbreviations and Acronyms	xv
List of Figures	xvii
List of Tables	xix
Chapter 1. Introduction	1
1.1. Topic Context	1
1.2. Motivation	2
1.3. Problem	2
1.4. Research Questions	3
1.5. Objectives	3
1.6. Methodologies	3
1.7. Outline of the Dissertation	4
Chapter 2. State of the Art	7
2.1. Search Strategy and Inclusion Criteria	7
2.2. Study Selection	7
2.3. Data Extraction and Synthesis	7
2.4. Results	8
Chapter 3. Artifact 1- AIMHealth App for Data Collection	13
3.1. Introduction	13

3.2.	UX Research	13
3.3.	Prototyping	19
3.4.	App Development and Improvements	22
Chapter 4.	Artifact 2- Data Mining for the AIMHealth App	25
4.1.	Introduction	25
4.2.	Data Mining with CRISP-DM	25
4.3.	Business Understanding	26
4.4.	Data Understanding	26
4.4.1.	Pneumonia Dataset	28
4.4.2.	Myocarditis Dataset	30
4.5.	Data Preparation	30
4.5.1.	Pneumonia Dataset	31
4.5.2.	Myocarditis Dataset	39
4.6.	Data Modeling	41
4.6.1.	Pneumonia Dataset	42
4.6.2.	Myocarditis Dataset	43
4.6.3.	Pneumonia and Myocarditis Dataset	44
Chapter 5.	Evaluation	47
5.1.	Introduction	47
5.2.	Artifact 1- AIMHealth App for Data Collection	47

5.3.	Artifact 2- Data Mining for the AIMHealth App	51
Chapter 6.	Conclusions	53
6.1.	Discussion and Conclusions	53
6.2.	Research Limitations	54
6.3.	Contributions	54
6.4.	Future Work	54
References		57
Annex A		63
Annex B		64

List of Abbreviations and Acronyms

AI Artificial Intelligence

ML Machine Learning

DL Deep Learning

RQ Research Question

UCD User-Centered Design

DSRM Design Science Research Methodology

CRISP-DM Cross-Industry Standard Process for Data Mining

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analysis

WoS Web of Science

EHR Electronic Health Records

cbPPG Camera-Based Photoplethysmography

UX User Experience

HSM Hospital de Santa Maria

DBSCAN Density-Based Clustering of Applications with Noise

HDBSCAN High Density-Based Clustering of Applications with Noise

AGNES Agglomerative Nesting

List of Figures

- 1.1 Interaction Between the Proposed Artifacts
- 1.2 DSRM Process Model [1]
- 2.1 PRISMA Workflow Diagram
- 2.2 Evolution of the Eligible Studies by Year
- 2.3 Main Topics from the Literature Review
- 3.1 Screenshots of the Initial Version of the AIMHealth App
- 3.2 User Persona
- 3.3 User Flow
- 3.4 Low-fidelity Prototype
- 3.5 High-fidelity Prototype
- 3.6 Screenshots of the New AIMHealth App Version
- 3.7 Examples of the Results Stored in the Appwrite Platform
- 4.1 CRISP-DM Methodology
- 4.2 Data Relationship Diagram
- 4.3 Patient and Record Distribution by Year
- 4.4 Total Patient Number by Age Group and Sex
- 4.5 Top 10 Pneumonia Diagnosis
- 4.6 Top 10 Length of Stay
- 4.7 Patient Distribution by Season
- 4.8 Anomaly Distribution by Month
- 4.9 Heart Rate Comparison from Patient 1 (“Senior”) and Patient 2 (“Child”)
- 4.10 Patient and Record Distribution by Year
- 4.11 Total Patient Number by Age Group and Sex
- 4.12 Patient with the Longest Length of Stay
- 4.13 Pneumonia Clusters Identified by AGNES Clustering Model
- 4.14 Pneumonia Clusters’ Variables Distribution
- 4.15 Myocarditis Clusters Identified by AGNES Clustering Model
- 4.16 Myocarditis Clusters’ Variables Distribution

- 4.17 Pneumonia and Myocarditis Clusters Identified by AGNES Clustering Model
- 4.18 Myocarditis and Pneumonia Clusters' Variables Distribution
- 5.1 Information System Artifact Evaluation Criteria Hierarchy and Selected Criteria

List of Tables

2.1 Keyword Selection

2.2 Studies by Topics

3.1 Demographic Data of the Five People Interviewed

3.2 Functional and Non-functional App Requirements

3.3 User Tasks

4.1 Description of Selected Tables and Views

4.2 Pneumonia Dataset

4.3 Problems Identified, Variable Selection and Data Cleaning Procedures Applied

4.4 New Variables

4.5 Anomaly Count by Diagnosis

5.1 Capabilities, Dimensions, Criteria and Objective Statements used in the DSRM Evaluation

5.2 ISO 15504 Four-Level NPLF Scale

5.3 App Evaluation Criteria, Questions, Objective Statements and Results

Introduction

1.1. Topic Context

A public health crisis began on 11th March 2020, when the World Health Organization declared Corona Virus Disease 2019 (Covid-19) a global pandemic [2]. During the outbreak, numerous regions across the world experienced significant challenges with their existing public healthcare systems, which were overwhelmed and underprepared to handle the scale and impact of the pandemic [3]. In order to mitigate the spread of the disease and minimize physical contact and hospitalization, rapid diagnosis, prognosis, remote monitoring became extremely crucial [4].

Smartphone applications, particularly mobile health (mHealth) applications, gained recognition as rapid and reliable solutions for diagnosing and monitoring patients, particularly with Covid-19 [5]. The adoption of mHealth applications offers several benefits, including reduced hospitalizations and in-person clinic visits, which is of immediate value in response to Covid-19, when healthcare systems strive to minimize face-to-face interactions and optimize the availability of hospital beds [6].

In addition to the aforementioned impacts on the healthcare system, Covid-19 disrupted the delivery of regular medical care, leading to patients postponing or completely forgoing necessary medical treatments [7]. The disruption and avoidance of regular medical care during the Covid-19 pandemic present significant challenges for patients with chronic diseases, as they risk receiving substandard care for their ongoing health conditions [7].

In Europe, non-communicable diseases continue to be the biggest healthcare capacity and cost burden, set to increase with an ageing society [6]. Remote monitoring can offer significant benefits by providing accessible healthcare services to a larger population, which can be particularly advantageous for patients who face challenges in receiving specialized care due to geographical constraints, limited transportation options, or physical limitations [8], [9].

Health remote monitoring systems have evolved to support easier communication between healthcare givers and patients for close monitoring, measurement of vital health parameters and routine consultation [10], [11]. This type of monitoring systems have also allowed the patient load at hospitals and health centers to be reduced [11]. Particularly, remote collection of vital signs (e.g., respiratory rate, heart rate, body temperature, blood oxygen saturation), as these are the first signs to show when something is alarming in several disease conditions [3], has allowed both patients and healthcare providers to have reduced exposure to Covid-19 [12].

By collecting user's physiological data and health information, mHealth technology can be utilized to monitor patients with chronic conditions, detect anomalies and forecast events, for instance, changes in the heart beat at rest, can be a sign of infection [13]. mHealth applications are thus promoting patient self-management, thus mitigating the necessity for urgent care and consequently reducing the number of hospitalizations [14].

In addition to what's been said, recent breakthroughs in Artificial Intelligence (AI) and Machine Learning (ML), have increased the capacity of remotely conduct a variety of tasks that previously required the physical presence of a medical professional, such as surveillance of vital signs of infected as well as suspected individuals [3], [15]. Using healthcare data, AI and ML have great potential in disease prediction [16]. By harnessing the capabilities of AI and ML, mHealth can deliver health services that are highly effective, personalized, and conducive to improved patient care [17].

This dissertation further develops an mHealth application, the AIMHealth app. The AIMHealth app is a part of the AIMHealth Secure Platform, seen in the Annex A, and was developed within the AIMHealth – AI-based Mobile Applications for Public Health Response project funded by Fundação para a Ciência e Tecnologia (FCT) [18] under DSAIPA/AI/0122/2020, coordinated by the Information Sciences, Technologies, and Architecture Research Center (ISTAR-Iscte).

1.2. Motivation

The motivation behind this research stems from the growing necessity for efficient and accessible disease remote monitoring solutions [19], particularly in response to the Covid-19 pandemic, that aim to mitigate the pandemic's impact and reduce the need for in-person hospital visits. Conventional disease monitoring approaches often rely on manual data collection and limited real-time insights, resulting in delays in identifying potential health complications [20]. The AIMHealth app aims to offer automated data collection and individualized disease monitoring, by utilizing AI, leading to improved patient outcomes and reduced healthcare costs.

1.3. Problem

The main problem lies in the lack of advanced mHealth applications that can efficiently collect health data, monitor, diagnose, and provide personalized healthcare services, particularly during pandemics such as Covid-19, while also catering to the needs of patients with chronic diseases and overcoming geographical and accessibility limitations.

1.4. Research Questions

The research questions that were addressed in this dissertation are as follows:

RQ 1: How can we optimize the user experience of the AIMHealth mHealth app to ensure ease of use and increased patient engagement during disease monitoring?

RQ 2: Can we identify meaningful patterns in health data using data mining techniques and machine learning algorithms, within the context of the AIMHealth app's objectives?

1.5. Objectives

The primary objective of this dissertation is to address the research questions previously mentioned in the section “Research Questions” by further developing an AI-based mHealth application- AIMHealth app- capable of monitoring patients who have one or more health conditions.

With this objective in mind, the main intent of this dissertation is to create two artifacts, “Artifact 1 - AIMHealth App for Data Collection” and “Artifact 2 - Data Mining for the AIMHealth App”, responding to RQ 1 and RQ 2, respectively. Figure 1.1 depicts the interaction between the two artifacts: the first artifact collects the user’s health data, encompassing parameters such as heart rate, blood pressure, body temperature and blood oxygen levels, for example, and is intended to supply data for the subsequent data mining process in the second artifact.



Figure 1.1. Interaction Between the Proposed Artifacts

1.6. Methodologies

To accomplish the research objectives, several methodologies were adopted. Initially, a state of the art review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [21] methodology. Subsequently, the dissertation adhered to the Design Science Research Methodology (DSRM) [1] and the seven guidelines proposed by the authors [1], involving the development of two distinct artifacts. This approach has its origins in engineering and artificial sciences, and its main objective is to create relevant artifacts in order

to add value to the fields they are applied to. Figure 1.2 depicts the nominal sequence of the six activities that describe the DSRM process, according to its authors [1].

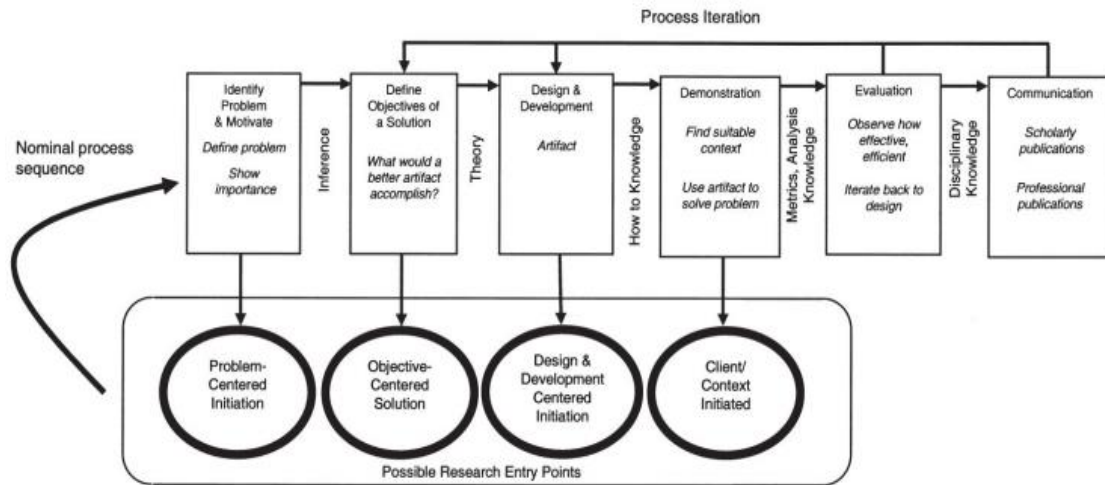


Figure 1.2. DSRM Process Model [1]

The DSRM has four entry points, and for this dissertation, the chosen entry point was “Problem-Centered Initiation” to begin the research process.

Even though the dissertation follows the DSRM principles, each artifact’s design and development follows a distinct methodology. The first artifact, dedicated to addressing the first research question concerning the enhancement of the app's user experience, was designed and developed in accordance to the principles of User-Centered Design (UCD) [22]. This involved user interviews, iterative design and prototyping to ensure user-friendly and intuitive interfaces. The design and development of the second artifact, which refers to the second research question concerning the data mining phase, followed the Cross-Industry Standard Process for Data Mining Methodology (CRISP-DM) [23].

1.7. Outline of the Dissertation

This dissertation is organized into six chapters. The first chapter is the "Introduction" where the context, objectives, and outline of the dissertation are presented.

The second chapter is titled the "State of the Art", and it provides a comprehensive and systematic review of the current state of the art, using the PRISMA methodology.

The third chapter is dedicated to the first artifact, the "AIMHealth App for Data Collection". This chapter outlines its objectives, features, and functionalities, and also covers the conducted User Experience (UX) research, prototyping, and the final interface modifications undertaken during the app's development.

The fourth chapter, concerning the second artifact, "Data Mining", details the data mining process conducted using the CRISP-DM methodology on hospital collected clinical data.

In the fifth chapter, "Evaluation", an assessment of both artifacts is carried out.

Lastly, the final chapter, "Conclusions", summarizes the key findings and conclusions derived from the research conducted in this dissertation.

State of the Art

2.1. Search Strategy and Inclusion Criteria

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Methodology [21] and after defining a research question “What is the state of the art on Artificial Intelligence-based mHealth applications for Disease Monitoring?”, a systematic review of the literature was conducted. This question matches the research aims.

The Scopus and Web of Science (WoS) databases were searched, and the research was carried out through September 12th, 2022. Only articles, original papers or reviews that had been published in journals between 2017-2022, written in English and related to Computer Science, Decision Science, Engineering and Mathematics were selected.

The search strategy was built around a query with a specific research focus, which will be addressed in section 2.4. This way it was possible to evaluate the number of articles that were present in both databases while taking the concept, context, and population under investigation into account.

2.2. Study Selection

After the database search and removal of duplicates, the title, abstract and keywords, were used to screen the papers selected for subsequent full text review, and in some cases the entire document was examined if those pieces of information weren't sufficient.

2.3. Data Extraction and Synthesis

The data, in this case the title, author, year, journal, subject area, keywords and abstract, was handled and stored with the support of Zotero and Microsoft Excel tools. Based on the results presented above, a qualitative assessment was conducted for data mining and synthesis. The Scopus and WoS databases were thoroughly searched for published work relating to the concepts “mHealth” or “Telemonitoring” or “Health App” or “Electronic Health Records”, the target population “Comorbidities” or “Medical Condition”, and within a “Data Mining” or “Machine Learning” or “Data Analysis” context of the study.

2.4. Results

The search results are depicted in Table 2.1. The query was made in both Scopus and at the WoS databases after applying the same restrictions and filters.

Table 2.1. Keyword Selection

Concept	Population	Context	Limitations
"mHealth" "Telemonitoring" "Health App" "Electronic Health Records"	"Comorbidities" "Medical Condition"	"Machine Learning" "Data Mining" "Data Analysis"	2017-2022 Only journal papers, articles, and reviews
6931 Documents			
178 Documents			
73 Documents			

As seen in the above table, the query was made using the keywords from each column (Concept AND Population AND Context AND Limitations) resulting in 73 documents. Following the PRISMA methodology, 24 documents were collected after removing duplicates, screening titles, abstracts, and keywords and after full assessment (Figure 2.1). The excluded documents did not align with the context of the study.

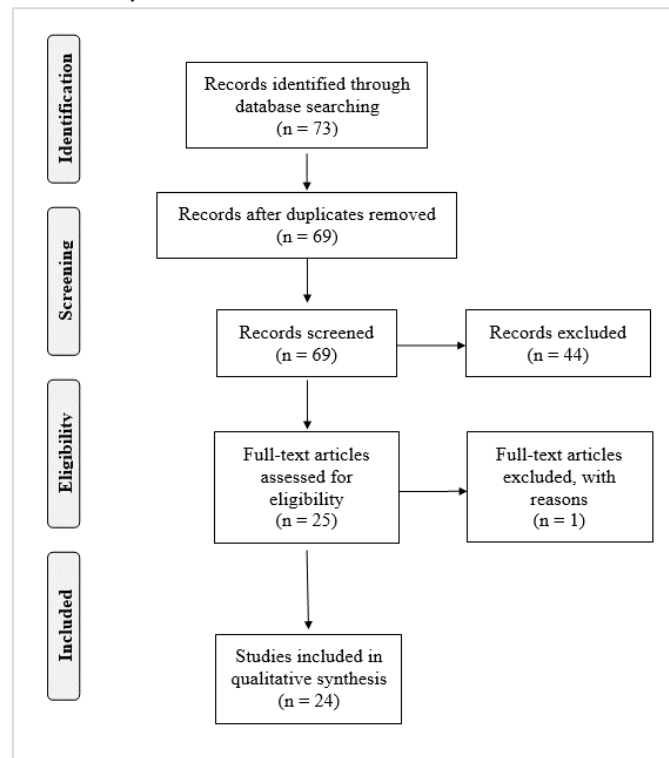


Figure 2.1. PRISMA Workflow Diagram

As depicted in Figure 2.2, the trend line shows a growth on the topic being studied, showing evidence of its pertinence.

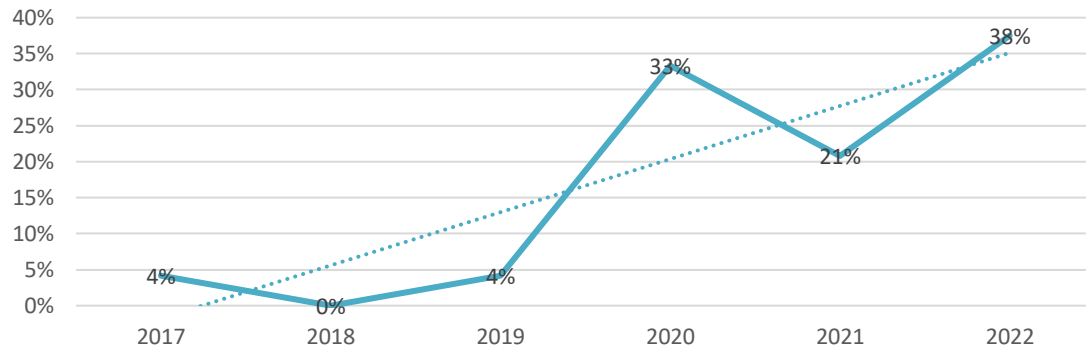


Figure 2.2. Evolution of the Eligible Studies by Year

A list of the main topics discussed on each of the reviewed articles is described in Figure 2.3, where it is visible that data mining and ML can be utilized to monitor and manage patients with comorbidities.

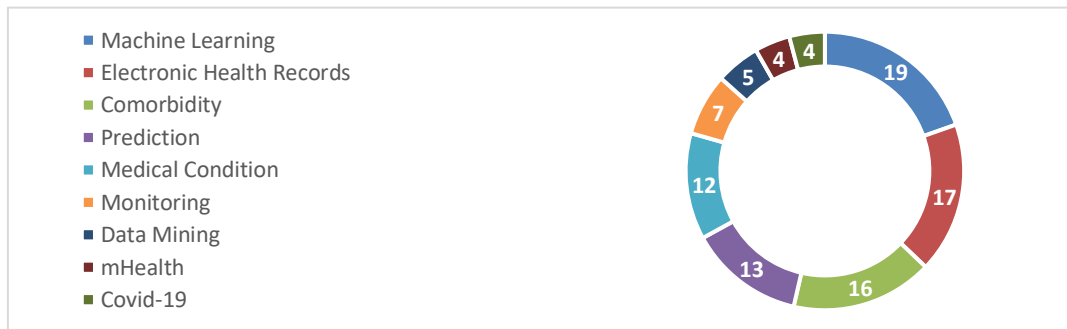


Figure 2.3. Main Topics from the Literature Review

A more thorough analysis of this review is summarized in Table 2.2, with a description of the main topics addressed by the selected papers. It is important to note that a single paper may encompass multiple topics simultaneously.

Table 2.2. Studies by Topics

Main Topics	Reference	Number of Documents
Machine learning	[24]–[43]	19
Electronic Health Records	[24], [26], [31], [33]–[41], [43]–[47]	17
Comorbidity	[24]–[26], [31]–[33], [35], [37]–[43], [46], [47]	16
Prediction	[27], [29], [33], [35], [36], [38], [40], [42], [44]–[47]	13
Medical Condition	[28], [30]–[37], [45]–[47]	12
Monitoring	[25], [27]–[30], [37], [44]	7
Data Mining	[24]–[26], [44], [45]	5

mHealth	[27]–[30]	4
Covid-19	[25], [41], [43], [47]	4

As seen in Figure 2.3, the most popular topics are “Machine Learning”, “Electronic Health Records” and “Comorbidity”.

Using ML, authors in [24] propose a point-of-care periodontitis risk assessment model, by comparing the performance of five ML methods: Naïve Bayes, Logistic Regression, Support Vector Machine, Artificial Neural Network, and Decision Tree. Authors from [36] propose a multiple instance learning boosting algorithm, able to extract hidden patterns from past EHR temporal data, for early prediction of Type 2 diabetes. In [40] a DL system that combines a convolutional neural network and a long short-term memory network is introduced for the early detection of sepsis. The authors in [25] conduct an explorative data mining of symptoms and comorbidities, comparing ML algorithms, such as Naïve Bayes, Decision Tree, Random Forest, K-Nearest Neighbor, Logistic Regression, Gradient Boosting Classifier, Support Vector Machines and Neural Networks, for Covid-19 severity detection. [41] using Extreme Gradient Boosting models, the researchers developed a method to identify potential patients with long-term effects of Covid-19. In [26] a cluster analysis, applying the K-Means algorithm, is conducted using Electronic Health Records (EHR) to identify the associations between frailty and clinical phenotypes. [43] developed models based on recurrent neural networks to predict Covid-19 patient outcomes. Researchers in [33] extracted features from EHR data to develop a Extreme Gradient Boosting model for predicting sepsis onset six hours in advance. [29] enhanced mHealth tools with decision-making capabilities, focusing on predicting surgical site infections by seamlessly integrating multiple ML models, such as K-Nearest Neighbor, Linear Regression, Support Vector Regression. In [42] five ML approaches (Category Boost, Random Forest, Extreme Gradient Boosting, Logistic Regression and Stacking Classifier) were employed to predict the likelihood of 30-day unplanned readmission for elderly patients.

Going through the studies performed using EHR, the authors from [31] investigated the utilization of unsupervised ML models (Latent Dirichlet Allocation and Poisson Dirichlet Model) to uncover latent disease clusters and patient subgroups. [32] developed a DL approach with K-means, Deep Clustering Network and Self-Organizing Map-Variational Autoencoder for clustering time-series data in EHRs to create clusters with patients who exhibited similar future outcomes of interest, such as adverse events or the onset of comorbidities. [34] highlights the variability in results from different clustering approaches (K-means, Kernel K-means, Affinity Propagation, Latent Class Analysis) for Alzheimer’s disease subtypes in EHRs. In [38] a novel framework is presented for continuous acute kidney injury prediction utilizing EHR data. [39]

present an unsupervised framework based on Word Embeddings, Convolutional Neural Networks, and Autoencoders for patient stratification using EHR. [44] proposed a data-driven model for accurate prediction of disease trajectories by leveraging EHRs. [46] assessed the effectiveness of a novel low-dimensional embedding model, disease2disease, learned from a large-scale EHRs dataset, in clustering the causes of kidney diseases and comorbidities. [47] presents a prognostic model for predicting critical states within 28 days following a Covid-19 diagnosis using EHR data and Extreme Gradient Boosting models. [45] developed methods based on tensor factorization to predict the occurrence of new chronic diseases, incorporating comorbidity patterns along with clinical and sequential factors from EHRs.

Paper [27] describes an AI aided home-based system that allows pregnant women to monitor their pregnancy on a daily basis and accurately predict their delivery date by the pattern analysis of their salivary crystals. [37] investigated chronic kidney disease progression with a hybrid semi mechanistic modeling methodology developed and applied to real world data. On [35] the construction of a patient graph structure to identify the appropriate therapies for each patient is described using basic patient information such as age and gender, along with diagnosis information. For disease management, in [28] a decision support system has been developed with the purpose of assisting individual patients in adhering to well-established clinical guidelines. [30] demonstrated the potential of smartphone-based tasks as objective measures for diagnosing and monitoring Parkinson's disease.

In summary, the state of the art results reveal that while AI has been widely applied in healthcare for various purposes, there is a notable absence of an AI-based mHealth app designed specifically for remote monitoring of patients with multiple health conditions.

Artifact 1- AIMHealth App for Data Collection

3.1. Introduction

This chapter revolves around the improvement of the AIMHealth app, guided by the principles of User-Centered Design (UCD) [22]. Its primary objective is to enhance the User Experience (UX) while addressing the first research question. Additionally, this chapter delves into the UX research conducted, the prototyping process, and the final interface enhancements that were implemented throughout the app's development process.

3.2. UX Research

Initially, the AIMHealth app had been developed in the scope of the AIMHealth project for physiological data collection, as proposed in [48], using Flutter, an open-source framework by Google used for app development [49]. Figure 3.1 displays multiple screenshots depicting the user experience before any modifications were made to the app. These screenshots showcase the app's login, home, and settings/profile pages in their original state.

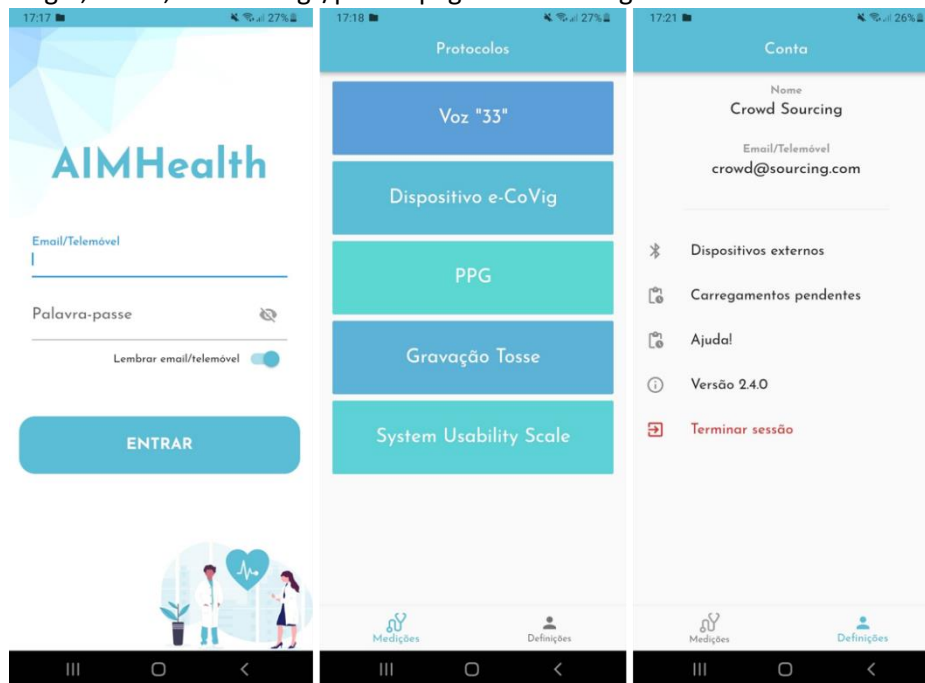


Figure 3.1. Screenshots of the Initial Version of the AIMHealth App

The initial version of the app, presented above, implemented five protocols for the evaluation of cardiac and respiratory functions, and of the app's usability, as well. These protocols were created to evaluate the status of Covid-19 disease in consultations with clinical specialists at

Hospital de Santa Maria (HSM) in Lisbon, Portugal. Two of the protocols played a role in cardiac function assessment: one in evaluation of the cardiac function resorting to camera-based photoplethysmography (cbPPG) acquisition for a period of 30 seconds - cbPPG is able to measure respiration rate and blood oxygenation [50]- and the second involved using data gathered from a wearable device connected with the smartphone, acquiring the same type of physiological data for the purpose of comparison and clinical evaluation. There were two protocols used for the respiratory function assessment, both recorded a brief 5 second audio clip through the smartphone's microphone: one for cough recordings ("Gravação Tosse") and another for voice '33' ("Voz '33'"). Finally, the System Usability Scale [51] questionnaire had been included to evaluate the usability of the application.

To enhance the development of the application, user interviews were conducted specifically targeting five potential users of the app: older individuals with one or more chronic conditions who were capable of using smartphones. Table 3.1 presents basic demographic information concerning the five individuals who were interviewed.

Table 3.1. Demographic Data of the Five People Interviewed

Age	Sex	Education	Marital Status	In Labor Force	Health Status
93	F	High school	Widowed	No	Diabetes, Arthritis, Deafness, Back problems, Hypotension
82	F	Primary school	Widowed	No	Renal Insufficiency, Hypertension
51	F	Bachelor's degree	Married	No	Hypertension
54	F	Bachelor's degree	Married	Yes	Healthy
73	M	Bachelor's degree	Married	Yes	Healthy

In the interviews, the users were asked "How do you currently monitor/track your health conditions?", "Do you maintain any kind of records? Do you have assistance from formal/informal caregivers or do you self-monitor?" and "Do you use any mHealth application?". These questions combined with the demographic data in Table 3.1, provided insights for the creation of our primary app persona. User personas play a crucial role in gaining a comprehensive understanding of the target audience [52], [53]. By grasping the expectations, concerns, and motivations of the target users represented by our primary persona, it becomes feasible to design an app that effectively fulfills their needs. Personas also assist in fostering empathy towards the end-users [53]. To enhance comprehension of user needs and ensure a

representative portrayal of the target user, the persona Carlota was crafted in alignment with the question “To whom is the app being developed?” (Figure 3.2).

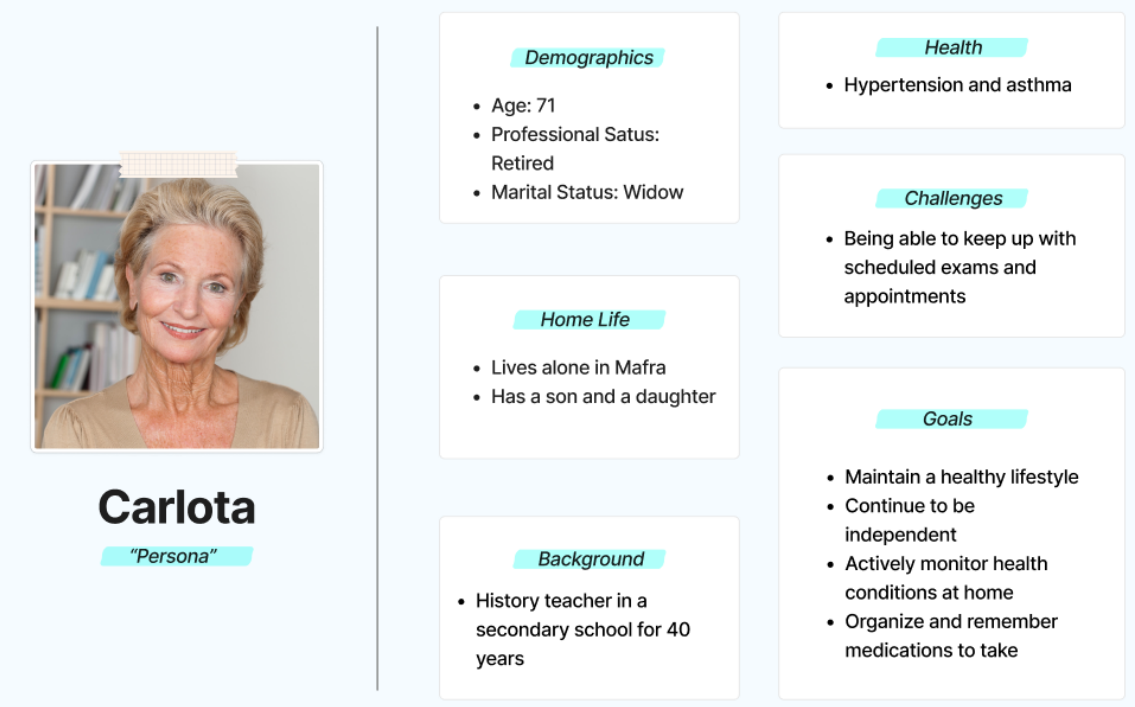


Figure 3.2. User Persona

Carlota is 71 years old and is currently living alone in an apartment in Mafra since the death of her husband and since her two children, Diogo and Liliana, left home. For 40 years she was a history teacher in a secondary school and has recently retired. For seven years she has suffered from hypertension and asthma. Despite this, she maintains a healthy lifestyle and actively monitors her health conditions by recording her oxygen saturation and blood pressure levels daily. Above all, Carlota wants to continue to have her independence by being able to manage her health at home. One of the greatest difficulties that Carlota has at the moment is being able to keep up with all the scheduled exams and appointments. Beyond this difficulty, Carlota wants to find an easier way to keep organized and remember the multiple medications she must take every day. Carlota uses a smartphone and, despite sometimes having some difficulty, she is still interested and willing to learn how to work with it. Even though she has never used them, she considers health apps a possible good solution to the problems she has been presenting: recording physiological data measurements, keeping track of exams and appointments, and organizing the prescribed medications.

Additionally, in the interviews the users were asked questions related to the features and usability of the app, such as “What would you like to see in a mHealth app?” and “What would make an mHealth app easy to use/understand?”. These made it possible to derive a set of new functional and non-functional user requirements for the app, as presented in Table 3.2.

Table 3.2. Functional and Non-functional User Requirements

Functional requirements	Non-functional requirements
System for measuring basic parameters (blood pressure, oxygen saturation, etc.)	Large font size
System for tracking the evolution of registered parameters	Use of colors and symbols
System for organizing prescribed medications	Easy access to the app via a simple code
System for organizing and monitoring scheduled exams and appointments	
System for alerting doctors in the event of major changes	
Minimal authentication required	
Clarification of the various options on the protocol/measurement screen	

Furthermore, a user flow was designed to showcase the potential interaction between the user and the application, providing insights into the user's journey within the app (Figure 3.3).

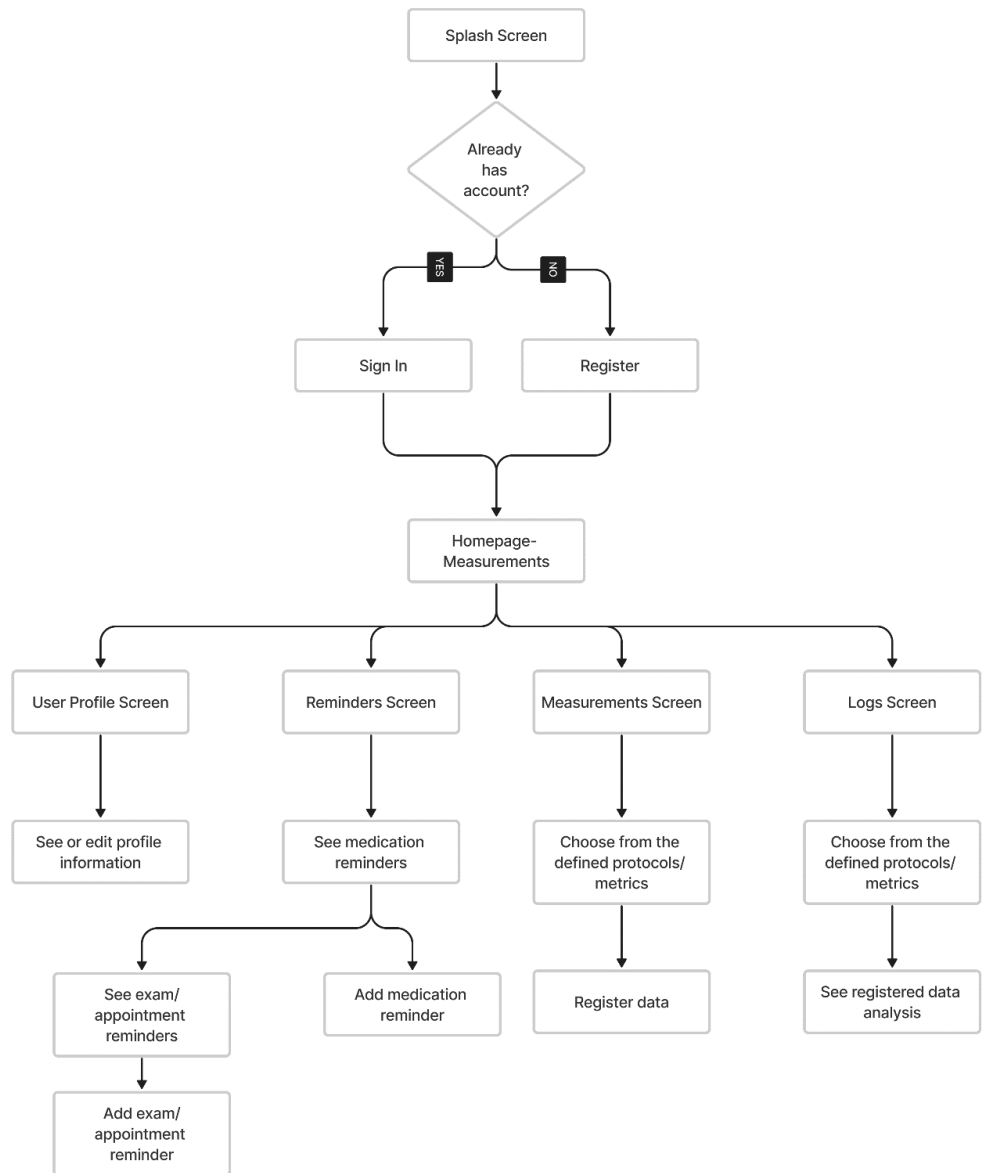


Figure 3.3. User Flow

After considering all the prior research, the following new tasks were identified as essential, to be supported by the app functionalities (Table 3.3):

Table 3.3. User Tasks

1. Sign In/Register

Upon opening the app, the user is greeted with a splash screen displaying the AIMHealth logo for a brief period. Then, a second screen appears presenting two options to the user: "Entrar" (Sign In) or "Registrar" (Register).

a. Sign In

If the user already has an account created, they will be prompted to log in using the email and password associated with their account. The app will provide input fields for the user to enter their email and password,

followed by a login button. Once the user enters their credentials and taps the login button, the app will validate the information and grant access to their account if the provided credentials are correct.

b. Register

If the user does not have an account created, they will be directed to the registration process to create a new account. During registration, the user will be prompted to enter their name, email address, and password. The app will provide input fields for the user to input their name, email address, and desired password. Once the required information is provided, the user can proceed to submit the registration form. The app will then process the registration request and create a new account associated with the provided credentials.

2. Physiological data measurement

The main screen of the application will present options displayed as squares, each representing a specific protocol or metric. These options will showcase the defined protocols or metrics that users can select. By selecting one of the options displayed on the main screen, users can initiate the recording of specific physiological data. This includes measurements such as heart rate, oxygen saturation, blood pressure, any other data that can be collected using a Bluetooth device, among others.

3. Medication, exam, and appointment reminders

On the reminders screen, users will encounter two displayed options: "Medicamentos" (Medications) or "Exames/Consultas" (Exams/Appointments).

a. Medication reminders

When the option "Medicamentos" is selected, the user can see information regarding the medication, such as the schedules, names, and quantities to be taken. Additionally, there is the possibility to add medications, by providing their name, quantity, days, time, and duration in which they must be administered, as well as adding notes, if necessary.

b. Exam and appointment reminders

By selecting the option "Exames/Consultas", a list of upcoming scheduled events will appear. As with medications, you can add exams and appointments providing information such as the type of event (appointment or exam), the specialty, the date and time it was scheduled for and the location (hospital/clinic where it will be performed). It is also possible to add notes in case this event has some prerequisite (take an exam, fasting, etc.), for example.

4. Recorded data visualization

On the logs screen, users will find the date of their last recorded measurements, along with squares similar to those on the main screen. However, these squares will display the last recorded values for each specific metric. The values will be color-coded, with green indicating values within the healthy range and red indicating alarming values that are outside the healthy range. By selecting one of these squares, users can access a detailed analysis of the collected data for that specific metric. The analysis includes an evaluation of the last recorded value on a scale, indicating whether it is below, above, or equal to the recommended/healthy value. Additionally, users can view the average, maximum, and minimum values recorded over the week and month, providing a broader understanding of their data trends and variations.

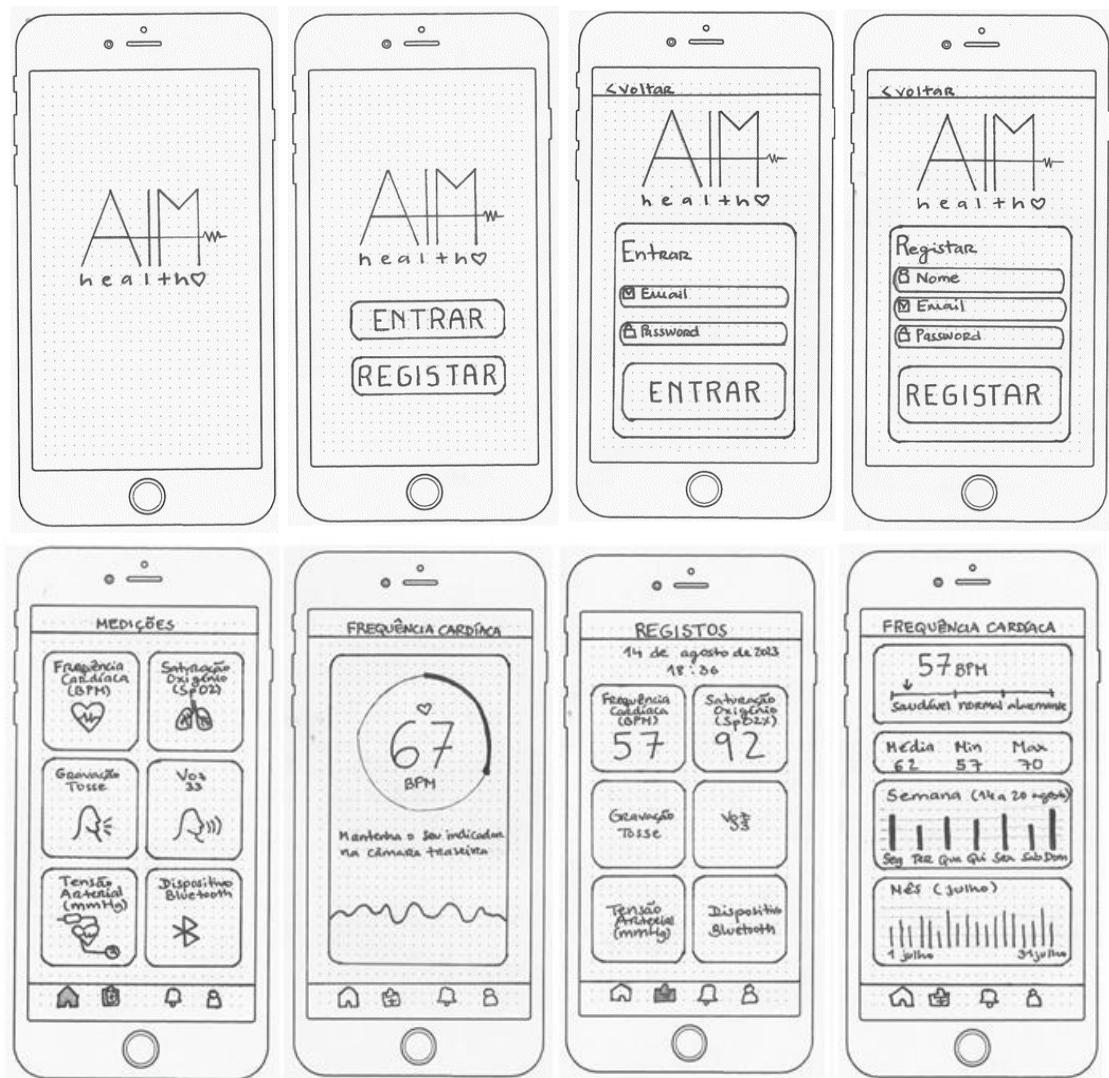
5. User profile

Within the user profile section, users have the ability to add or edit their profile information. This includes details such as their name, date of birth, weight, gender, email, and password.

3.3. Prototyping

This section introduces and presents the developed prototypes. Prototyping aids in the visualization and measurement of ideas [54]. It allows the transformation of insights and findings from the interviews into a tangible form by conceptualizing the app's functionalities, layout, and user interface [55].

A low-fidelity prototype was created to visualize the design and to conduct a usability test, with the permission of the ethics council of Iscte (Annex B) [56]. The primary objective of the low-fidelity prototype was to design a straightforward and natural to use app interface, incorporating large font sizes, easily recognizable icons, and basic displays in alignment with the previously identified user requirements of our primary persona Carlota (Figure 3.4).



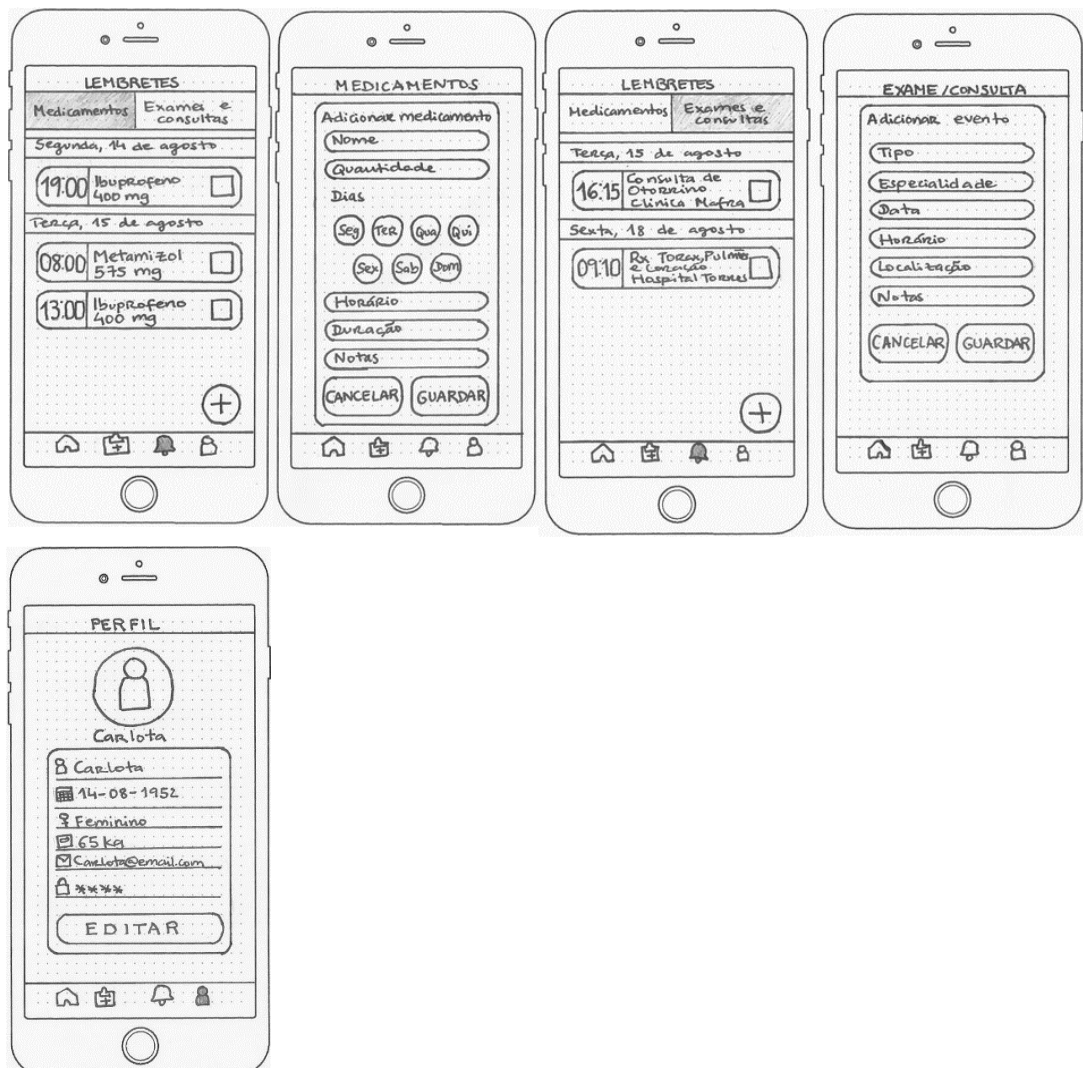


Figure 3.4. Low-fidelity Prototype

Due to the time required for the approval of the testing protocol, the low-fidelity prototype could not undergo user testing.

The prototyping process progressed by building upon the low-fidelity prototype, and since high-fidelity prototypes facilitate the demonstration and assessment of real-time interactions between users and the application [54], a high-fidelity prototype was developed using the Figma interface design tool [57], illustrated in Figure 3.5.



Figure 3.5. High-fidelity Prototype

3.4. App Development and Improvements

The AIMHealth app was further developed based on the high-fidelity prototype, utilizing the Flutter framework [49]. Figure 3.6 showcases several screenshots that illustrate the user interface of the AIMHealth app. Based on the functional and non-functional requirements defined for the app, several requirements were satisfied, these include:

- Enhancing the system for measuring health parameters – oxygen saturation, body temperature and blood pressure integrated into the “Medições” screen.
- Implementing systems to organize and monitor prescribed medications, examinations, and appointments.
- Implementing larger font sizes, distinctive colors, and symbols, to enhance user accessibility.
- Clarifying the options available on the main screen ("Medições").

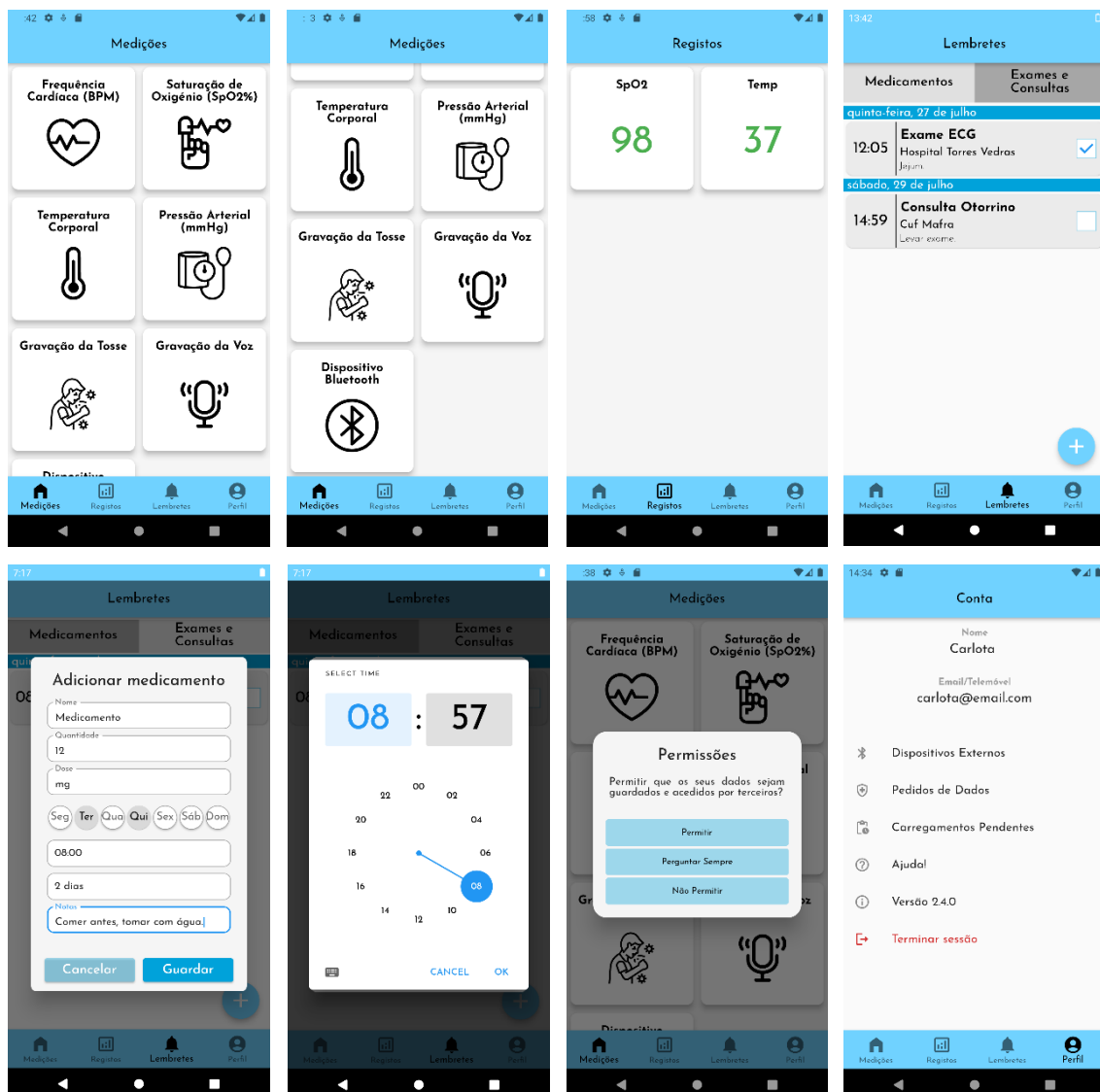


Figure 3.6. Screenshots of the New AIMHealth App Version

Currently, the AIMHealth application is being employed by physicians for the purpose of collection of patient clinical data at the Hospital de Santa Maria in Lisbon. Approval for such data collection and the use for research purposes of collected clinical data, was provided under the framework of the FCT project DSAIPA/AI/0122/2020 AIMHealth - Mobile Applications Based on Artificial Intelligence, by the Ethical Committee of the Faculty of Medicine of Lisbon, one of the project partners. The dataset is currently being accessed by Iscte and Faculty of Medicine researchers, under an NDA and in the scope of the FCT AIMHealth project. Clinical data is acquired in a pseudo-anonymous way, whereby an ID enables the discernment of the application’s users among the patients at HSM, through the app’s Bluetooth device and encompasses temperature, heart rate, and blood oxygen levels. Afterward, physicians register the same data but from medically certified devices to facilitate a comparison with the data collected through the app and ensure its validity. All this data is being stored within the Appwrite [58] platform, a backoffice system also developed within the framework of FCT AIMHealth, as depicted in Figure 3.7.

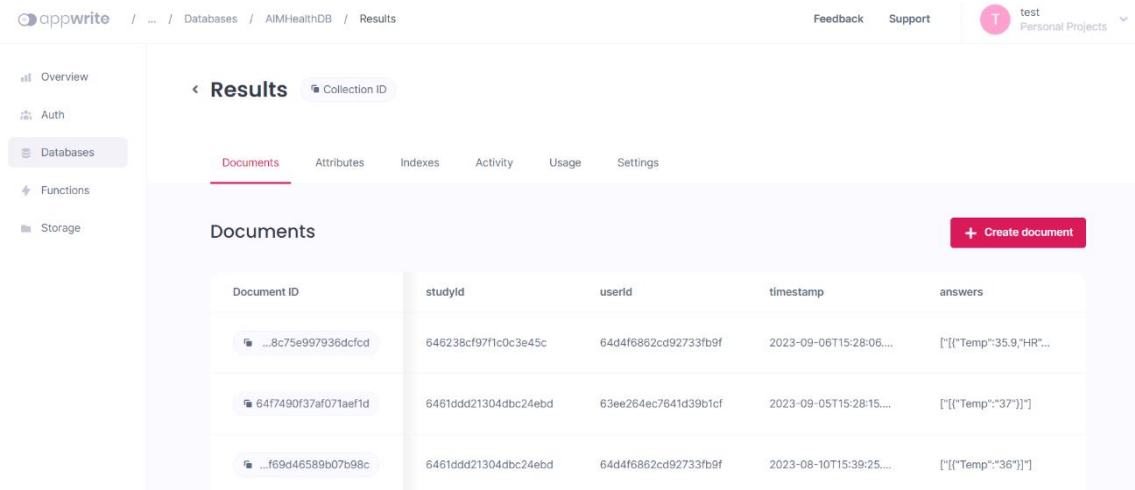


Figure 3.7. Examples of the Results Stored in the Appwrite Platform

Artifact 2- Data Mining for the AIMHealth App

4.1. Introduction

This fourth chapter addresses the second research question: “Can we identify meaningful patterns in health data using data mining techniques and machine learning algorithms, within the context of the AIMHealth app's objectives?”. This chapter details the data mining process, but because the data collected from the app was not accessible at the time of the study, it was conducted on EHR data obtained from HSM hospital, under the same approval process as defined in the previous chapter, extended to legacy data collected during the period of 01/01/2019 to 12/11/2021 in the following hospital services: Cardiology, Intensive Medicine, and Respiratory Intensive Care Unit. It includes data from 23 122 patients and contains real-time clinical signals, such as Temperature, Blood Oxygen Level (SpO2), Heart Rate, for example, equivalent to the parameters collected through the app. An additional cardiology dataset, collected in the period 2018 to 2019, was made available too. The data mining process was guided by the CRISP-DM methodology. This chapter does not encompass the evaluation, which will be addressed in chapter six, nor does it include the deployment phase, which is planned for future work.

4.2. Data Mining with CRISP-DM

The CRISP-DM methodology [23] was used in this chapter. This methodology is divided into six stages, which include business understanding, data understanding, data preparation, modeling, evaluation, and deployment, as depicted in Figure 4.1.

The CRISP-DM model begins with business understanding, which focuses on understanding the context and core objectives for the generation of the data. The second step, data understanding, involves familiarization with the available data as well as preliminary data exploration to find data quality issues and draw the initial insights from the observed data variables. The third phase is dedicated to data preparation, covering the tasks designed to create a final dataset that is going to be used in the following phase of modelling. In this step, null, duplicate, and outlier values are going to be examined. Furthermore, new variables are going to be constructed from the original variables. The next phase, corresponding to modeling, is focused on applying data mining techniques particularly with machine learning approaches. In this phase, the activities are focused on finding patterns and consequently extracting

information. The evaluation phase primarily focuses on assessing the models developed, with metrics adopted in the literature. Lastly, the deployment phase involves integrating the data mining results into production environments.

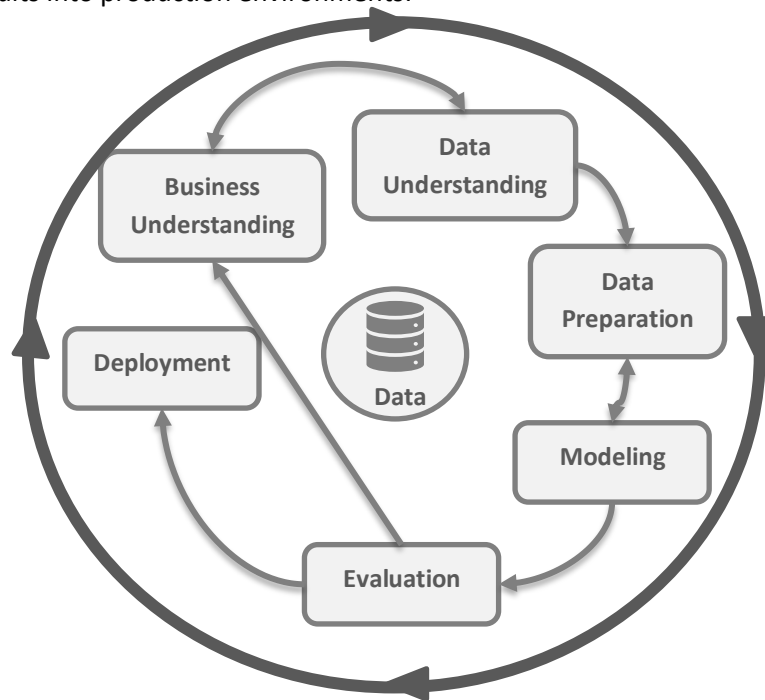


Figure 4.1. CRISP-DM Methodology

4.3. Business Understanding

The primary objective of the analysis was to identify meaningful patterns within the EHR data and then leverage these patterns and predictions to enhance the functionality of the AIMHealth app. The key goal was to empower physicians to make data-driven decisions and offer users valuable insights, proactive monitoring, and personalized interventions for better and more efficient health management.

4.4. Data Understanding

The data used for this research originates from the AimHealth database, available to this work by means of an NDA and computational means, including remote https secure access to a research server located in the ISCTE data center. We used a subset of the available data, stored in the DataGrip IDE [59], consisting of 152 tables and 13 views. Each table and view contain different types of hospital health data, such as the "V_Patients" view containing patient demographic information, the "LABRESULTS" table containing information on laboratory test results, among others. The data was extracted from "V_RTData", "PICISDATA", "ADMISSIONS", "V_Patients", "ADMISSIONDIAGNOSIS", "V_DIAGNOSIS", "LABTESTS", "ANALYSIS" and "LABRESULTS" since they provide the patients' clinical information and real-time data, therefore

being the most suitable dataset for this analysis. Due to the unavailability of access to the Covid-19 database and the presence of other health condition databases, the research focused on the examination of two specific diagnoses: pneumonia and myocarditis. These diagnoses were chosen for the research as they were specified by a physician and are indicative of health issues affecting the respiratory and cardiovascular systems, respectively.

For this process and the stages of data preparation and modeling, we employed the Python programming language [60], utilizing the JupyterLab [61] environment along with various Python libraries. Notably, the Pandas [62] library was utilized in this phase.

Figure 4.2 depicts the data relationship model diagram of the selected tables and views, and Table 4.1 provides a brief description of each table and view.

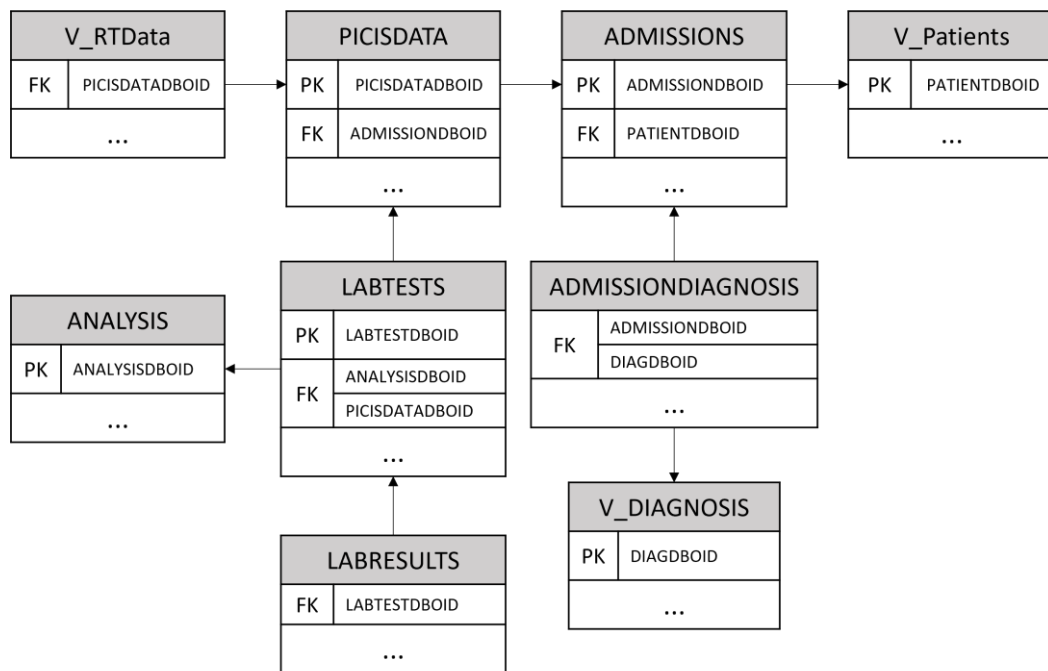


Figure 4.2. Data Relationship Diagram

Table 4.1. Description of Selected Tables and Views

Table/ View	Description	Records	Variables	Variables 100% Null
V_RTData	Contains data for physiological variables	1 015 326	93	69
PICISDATA	Links clinical data with admissions	1 159 138	4	-
ADMISSIONS	Administrative data	1 159 139	38	11
V_Patients	Demographic data for each patient	512 761	9	-
ADMISSIONDIAGNOSIS	Diagnosis data for each admission	1 348 498	16	2
V_DIAGNOSIS	Diagnosis codes, descriptions and categories defined by the hospital	12 6126	9	1

LABTESTS	Laboratory tests that have taken place	8 043 764	9	-
ANALYSIS	Names of laboratory tests	1 734	7	2
LABRESULTS	Results of laboratory tests	24 836 603	17	2

4.4.1 Pneumonia Dataset

The dataset was filtered to exclusively display data from patients diagnosed with pneumonia. Several columns from the selected tables and views contained exclusively null values, so they were excluded from the initial dataset. Additionally, certain columns represented foreign keys to tables that were not used in this analysis, while others were auto populated with their respective names or with default values. As a result, the initial dataset did not encompass every column, but was composed of 1 045 616 rows and 47 columns. The "LABTESTS", "ANALYSIS" and "LABRESULTS" tables have not been incorporated in this stage of the data mining as they require a separate filtering process to include only the relevant tests. Table 4.2 presents an examination of the data descriptions, original table or view, data types and examples in the initial dataset's columns.

Table 4.2. Pneumonia Dataset

Variable	Description	Original Table/ View	Data Type	Data Example	Null Values
RTDataDBOID	Unique numerical key to identify the RTData	V_RTData	Float64	6.381089e+19	0
RTDATADBOID2	Unique text key to identify the RTData	V_RTData	Object	63810892800411200553	0
Started	Time the physiologic variable data was read	V_RTData	Datetime64	23/02/2021 08:00:00	0
Validated	Indicates whether the physiologic variable has been validated	V_RTData	Object	F	0
UpdateTime	Date/time the variable was modified	V_RTData	Datetime64	23/02/2021 08:00:00	0
PICISDATADBOID	Unique numerical key to identify the picis data record	V_RTData	Float64	5.581083e+19	0
PICISDATADBOID2	Unique text key to identify the picis data record	V_RTData	Object	55810826474986000000	0
CREATIONDATE	Date/time the record was created	V_RTData	Object	2021-02-23 08:00:00	0
Temperatura - Registo Manual	Temperature (manual registration)	V_RTData	Float64	36.0	1033707
Respiration Rate from pCO	Respiration Rate	V_RTData	Float64	14.0	1045285

ST Segment Lead V4	Leads on an electrocardiogram used to evaluate the ST segment	V_RTData	Object	0	1043018
ST Segment Lead V3		V_RTData	Object	0	1043018
ST Segment Lead AVF		V_RTData	Object	0,2300	463326
ST Segment Lead AVR		V_RTData	Object	-0,280	463326
ST Segment Lead AVL		V_RTData	Object	0,0299	463326
ST Segment Lead III		V_RTData	Object	0,0799	455559
ST Segment Lead V2		V_RTData	Object	0,1000	1016023
ST Segment Lead V1		V_RTData	Object	-0,100	1042415
Mean Arterial Pressure 2	Mean Arterial Pressure 2	V_RTData	Float64	47.0	1044562
Diastolic Pressure (Art.) 2	Diastolic Pressure (Art.) 2	V_RTData	Float64	41.0	1044562
Systolic Pressure (Art.) 2	Systolic Pressure (Art.) 2	V_RTData	Float64	57.0	1044562
Pulso	Pulse	V_RTData	Float64	112.0	123936
Pulse Rate BP Tracing	Pulse Rate	V_RTData	Float64	80.0	1040059
Frequência Cardíaca	Heart Rate	V_RTData	Float64	113.0	17227
ADMISSIONDB OID	Unique identifier for the admission data	ADMISSIONS	Float64	4.718044e+18	0
HOSPITALSTARTED	Date/time when the patient is admitted to the hospital	ADMISSIONS	Datetime64	28/09/2020 15:16	387083
STARTED	Starting date/time of Picis admission	ADMISSIONS	Object	27/06/2011 12:21	0
ENDED	Ending date/time of Picis admission	ADMISSIONS	Object	09/11/2010 17:31	1282
PREADMISSION	Starting date/time of Picis pre-admission	ADMISSIONS	Datetime64	28/09/2020 15:18	375559
WEIGHT	Patient's weight	ADMISSIONS	Float64	70.00	3695
HEIGHT	Patient's height	ADMISSIONS	Float64	170.0	270569
ISDELETED	Determines whether the entry is logically deleted	ADMISSIONS	Object	T	0
ADMITFACILITY	Facility in which the patient is admitted	ADMISSIONS	Object	CHLN	0
HOSPITALENDED	Date/time when the patient is discharged from the hospital	ADMISSIONS	Datetime64	20/10/2018 17:00	1045616

PATIENTDBOID	Unique internal identifier for a patient	V_Patients	Object	1.795208e+18	0
BIRTHDATE	Patient's birthdate	V_Patients	Datetime64	15/11/1999 00:00	0
Age	Patient's age	V_Patients	Float64	23.5072	0
Age_Group	Patient's age group	V_Patients	Object	Jovem Adulto	0
BloodGroup	Patient's blood group	V_Patients	Object	A+	0
Sex	Patient's sex	V_Patients	Object	Feminino	0
Ethnicity	Patient ethnicity	V_Patients	Object	Caucasiano	0
EthnicityGroup	Patient ethnic group name	V_Patients	Object	Caucasiano	593976
MaritalStatus	Patient's marital status	V_Patients	Object	Solteiro	0
ADMDIAGDATE	Diagnosis date	ADMISSIONDIAGNOSIS	Datetime64	2021-11-12 00:04:12	0
DIAGID	Unique identifier for a diagnosis	V_DIAGNOSIS	Float64	7.000000e+18	0
DIAGDESC	Diagnosis description	V_DIAGNOSIS	Object	Pneumonia Nosocomial	0
DIAGCODE	Diagnosis code	V_DIAGNOSIS	Object	502	0

4.4.2 Myocarditis Dataset

This dataset shares the same variables as the previous one, with the distinction that it has been filtered to present data from patients diagnosed with myocarditis and has a significantly lower number of records (21 793).

4.5. Data Preparation

Following the completion of the data understanding stage, several issues affecting the data quality were identified. These issues primarily revolve around the high quantity of missing values, the presence of outliers and wrong data types. During this phase, the focus was on addressing these problems and carrying out the necessary activities to improve the data quality. Additionally, this phase involved creating new variables derived from existing ones, which provided further insights and enhance the analysis process.

For this process and the subsequent stage of the data mining, the following libraries were utilized: Pandas [62], NumPy [63], Matplotlib [64] and Seaborn [65].

4.5.1 Pneumonia Dataset

Table 4.3 provides an overview of the issues identified in the variables. Additionally, this table includes information about the variable selection, the reasoning behind, and the data cleaning processes done.

Table 4.3. Problems Identified, Variable Selection and Data Cleaning Procedures Applied

Variable	Problems Identified	Exclude?	Reason	Data Cleaning
RTDataDBOID	-	Yes	Unnecessary identifier code	-
RTDATADBOLD2	-	Yes	Unnecessary identifier code	-
Started	5 Outliers ("2242-03-16 12:56:32")	Yes	Similar meaning to "CREATIONDATE"	-
Validated	-	Yes	Has only 1 value (F)	-
UpdateTime	-	Yes	Similar meaning to "CREATIONDATE"	-
PICISDATADBOLD	-	Yes	Unnecessary identifier code	-
PICISDATADBOLD 2	-	Yes	Unnecessary identifier code	-
CREATIONDATE	Wrong data type	No	Identifies the date the record was created	Converted to Datetime64
Temperatura - Registo Manual	1033707 Missing values	No	Identifies the patient's temperature	Filled missing values with the patient's previous/next value
Respiration Rate from pCO	998295 Missing values	No	Identifies the patient's respiration rate	
ST Segment Lead V4	1043018 Missing values	No	Relevant to the analysis	Deleted missing values Converted to float64
ST Segment Lead V3	Non-numeric data type	No		
ST Segment Lead AVF	463326 Missing values	No		
ST Segment Lead AVR	Non-numeric data type	No		
ST Segment Lead AVL		No		
ST Segment Lead III	455559 Missing values Non-numeric data type	No		
ST Segment Lead V2	1016023 Missing values Non-numeric data type	No		
ST Segment Lead V1	1042415 Missing values Non-numeric data type	No		
Mean Arterial Pressure 2	1044562 Missing values	No	Identifies the patient's blood pressure	

Diastolic Pressure (Art.) 2		No		Filled missing values with the patient's previous/next value
Systolic Pressure (Art.) 2		No		Filled missing values with the patient's previous/next value or reference value "70"
Pulso	123936 Missing values	No	Identifies the patient's pulse	
Pulse Rate BP Tracing	1040059 Missing values	Yes	Not relevant to the analysis	-
Frequência Cardíaca	17227 Missing values	No	Identifies the patient's heart rate	Filled missing values with the patient's previous/next value or reference value "70"
ADMISSIONDBOIID	-	Yes	Unnecessary identifier code	-
HOSPITALSTARTEID	-	Yes		-
STARTED	-	Yes		-
ENDED	1282 Missing values	Yes	Not relevant to the analysis	-
PREADMISSION	-	Yes		-
WEIGHT	3695 Missing values	No	Identifies the patient's weight	Filled missing values with the patient's previous/next value
HEIGHT	270569 Missing values	No	Identifies the patient's height	
ISDELETED	-	Yes		-
ADMITFACILITY	-	Yes	Not relevant to the analysis	-
HOSPITALENDED	1045616 Missing values	Yes		-
PATIENTDBOIID	-	No	Necessary identifier code	-
BIRTHDATE	-	Yes	Unnecessary	-
Age	-	No	Identifies the patient's age	-
Age_Group	-	No	Identifies the patient's age group	-
BloodGroup	-	No	Identifies the patient's blood group	-
Sex	-	No	Identifies the patient's sex	-
Ethnicity	-	No	Identifies the patient's ethnicity	-
EthnicityGroup	593976 Missing values	Yes	Similar to "Ethnicity"	-
MaritalStatus	Repeated categories	Yes	Not relevant to the analysis	-
ADMDIAGDATE	Wrong data type	No	Identifies the diagnosis date	Converted to Datetime64
DIAGID	-	Yes	Unnecessary identifier code	-
DIAGDESC	-	No	Identifies the diagnosis	-

DIAGCODE	-	Yes	Unnecessary code	-
----------	---	-----	------------------	---

After the data cleaning processes mentioned in the table above were performed, some missing values remained. These missing values were replaced with 0, as they represent the absence of registered values. The decision to not utilize the mean or median for replacing these missing values was made because doing so would have had a significant impact on the results.

As part of the data preparation process, some variable names and values were modified or translated to enhance clarity and understanding. Examples of these changes include transforming "Frequência Cardíaca" to "Heart Rate" and "Temperatura - Registo Manual" to "Temperature".

To enrich the dataset, new variables were derived from the pre-existing variables. From "Creation Date", the "Season", "Length of Stay" and "Previously Admitted" variables were created. By applying specific conditions provided by a physician to the "ST Segment Lead" variables, which involve comparing ST values with the previous value, the "Anomaly" variables were determined [66]. These variables indicate whether there are any significant deviations or abnormalities in the ST segment values. Additionally, relevant laboratory test results and dates from "LABTESTS", "LABRESULTS", "ANALYSIS" were added as suggested by a physician. These tests encompass Troponin, NT-proBNP, C-Reactive Protein (CRP), Sodium, Potassium, Hemoglobin, and Blood Oxygen (SpO2). Table 4.4 depicts what these new variables represent and their relevance to the analysis.

Table 4.4. New Variables

Variable	Null Values	Description	Relevance
Season	-	Season of the Year	Seasonal patterns
Length of Stay		Duration of hospitalization for each patient	Assess disease severity and treatment response
Previously Admitted		Whether the patient has been previously hospitalized with Pneumonia	Determine if the pneumonia is recurrent
Anomaly 1		Represents a ST elevation in ST Segment Lead	Pneumonia can lead to complications affecting the cardiovascular system
Anomaly 2		Represents a ST depression in all ST Segment Lead variables	
Anomaly 3		Represents a ST elevation in ST Segment Lead V2 and V3	
Anomaly		Whether an anomaly 1, 2 or 3 happened	
TROPONIN	1045447	Patient's troponin values	

DATE_TROPONIN		Date/time the troponin test was done	
NT-PROBNP		Patient's NT-proBnP values	
DATE_NT-PROBNP	1045597	Date/time the NT-proBnP test was done	
CRP		Patient's CRP values	
DATE_CRP	1045208	Date/time the CRP test was done	
SODIUM		Patient's sodium values	
DATE_SODIUM	1045329	Date/time the sodium test was done	
POTASSIUM		Patient's potassium values	
DATE_POTASSIUM	1045340	Date/time the potassium test was done	
SPO2		Patient's blood oxygen values	
DATE_SPO2	1045630	Date/time the blood oxygen test was done	
HEMOGLOBIN		Patient's hemoglobin values	
DATE_HEMOGLOBIN	1045637	Date/time the hemoglobin test was done	

Set of relevant laboratory tests provided by a physician

Upon the creation of the laboratory results variables, it was observed that a significant number of missing values were present in these variables. This high proportion of missing values rendered these variables unusable for further analysis or interpretation. The following variables were kept: "PATIENTDBOID", "CREATIONDATE", "Temperature", "Respiration Rate from pCO", "ST Segment Lead V4", "ST Segment Lead V3", "ST Segment Lead AVF", "ST Segment Lead AVR", "ST Segment Lead AVL", "ST Segment Lead III", "ST Segment Lead V2", "ST Segment Lead V1", "Mean Arterial Pressure 2", "Diastolic Pressure (Art.) 2", "Systolic Pressure (Art.) 2", "Pulse", "Heart Rate", "WEIGHT", "HEIGHT", "Age", "Age Group", "Blood Group", "Sex", "Ethnicity", "Season", "Length of Stay", "Previously Admitted", "Anomaly 1", "Anomaly 2", "Anomaly 3", "Anomaly", "ADMDIAGDATE", "DIAGDESC".

Another issue that was identified is that a patient could have multiple pneumonia diagnoses simultaneously. This resulted in duplicated and repeated records when merging the tables. To address this, a modification was made to the "DIAGDESC" variable. For instance, cases where a patient had two identical records, one with "DIAGDESC" as "PNEUMONIAS BACTERIANAS NCOP" and the other as "Pneumonia nosocomial", were consolidated into a single record with "PNEUMONIAS BACTERIANAS NCOP + Pneumonia nosocomial" as the updated "DIAGDESC" value.

Following the data preparation procedures, the dataset underwent a reduction of 21.6%, leading to the removal of 225 625 records. As a result, the dataset now consists of 819 991 records and 33 variables.

Figure 4.3 depicts the distribution of records and patients by year. Notably, there is a greater concentration of patients in 2020, with a lower number in 2021. This observation can be attributed to the fact that the pneumonia dataset covers the period from 2018 to November 2021, with one month missing in 2021. However, it is also possible that this decrease is impacted by various factors such as the implementation of Covid-19 restrictive measures, including social lockdowns and wearing masks.

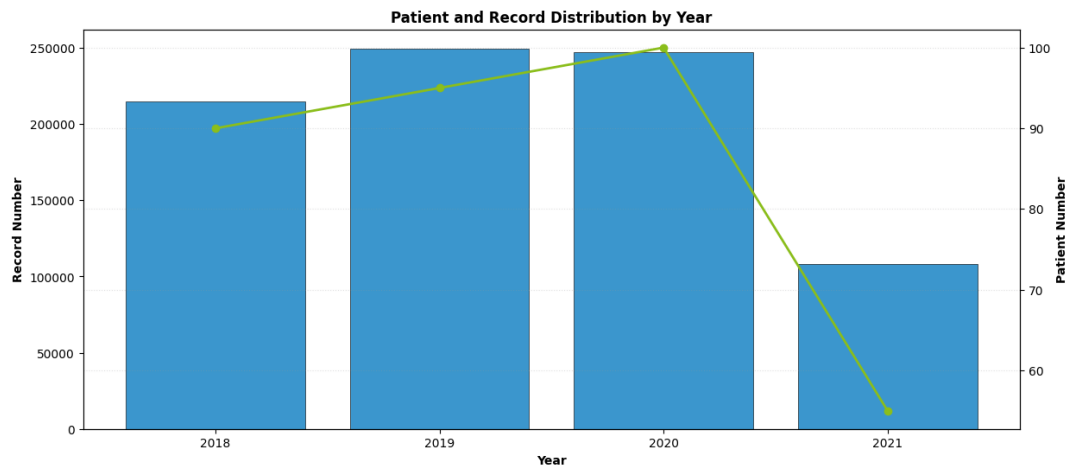


Figure 4.3. Patient and Record Distribution by Year

Figure 4.4 displays the distribution of patients based on age group and sex. As anticipated, the most affected age groups are “Child” and “Senior”, as they typically have a relatively weaker immune system compared to other age groups. Furthermore, it is evident that the male gender accounts for a larger proportion of diagnosed pneumonia cases.

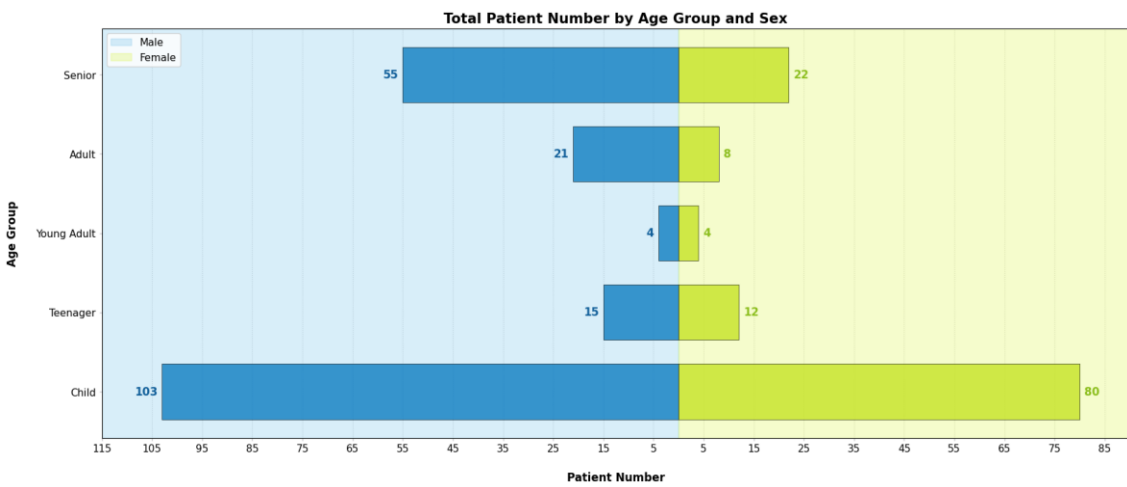


Figure 4.4. Total Patient Number by Age Group and Sex

Within the dataset, there are a total of 70 distinct variations of pneumonia, including different possible combinations. As depicted in Figure 4.5, the most diagnosed condition is “Pneumonia ou Pneumonite”, followed by “Pneumonias Bacterianas NCOP” and “Pneumonia Nosocomial”.

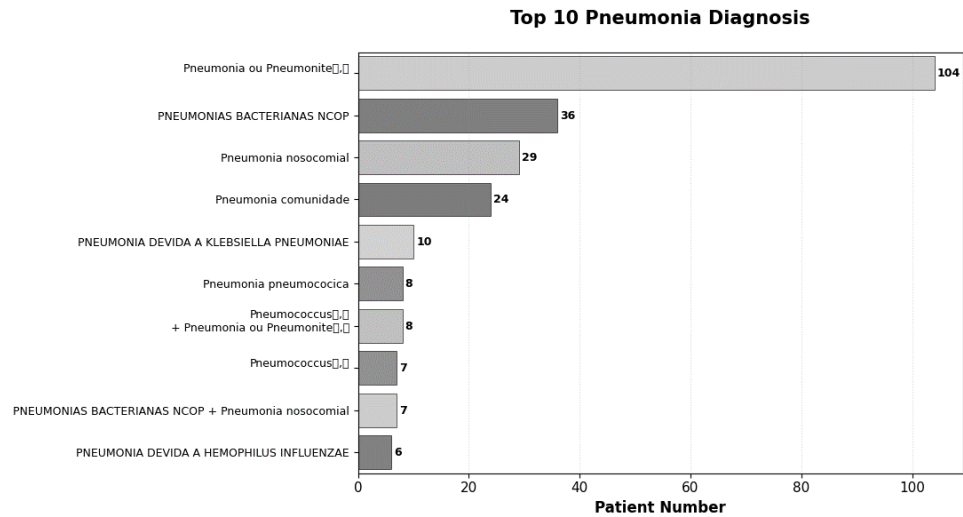


Figure 4.5. Top 10 Pneumonia Diagnosis

Regarding the length of stay, it is observed in Figure 4.6 that patients were more likely to have a hospitalization period of less than a week.

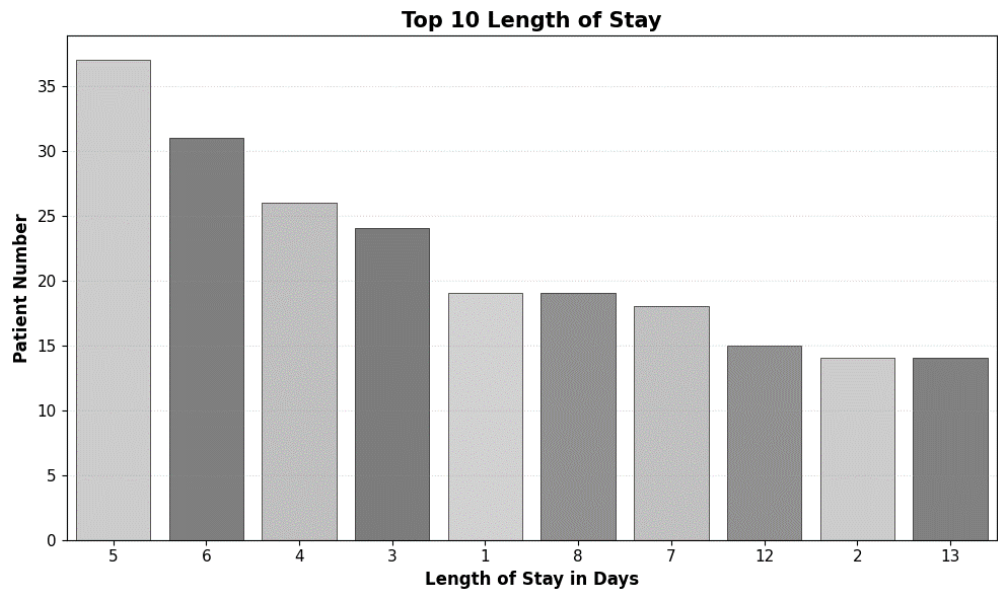


Figure 4.6. Top 10 Length of Stay

In terms of seasonal patterns, there is a peak of occurrences during the winter months, as seen in Figure 4.7. This can be attributed to various factors such as increased indoor crowding, closer contact with others, and environmental circumstances that facilitate the transmission of respiratory diseases.

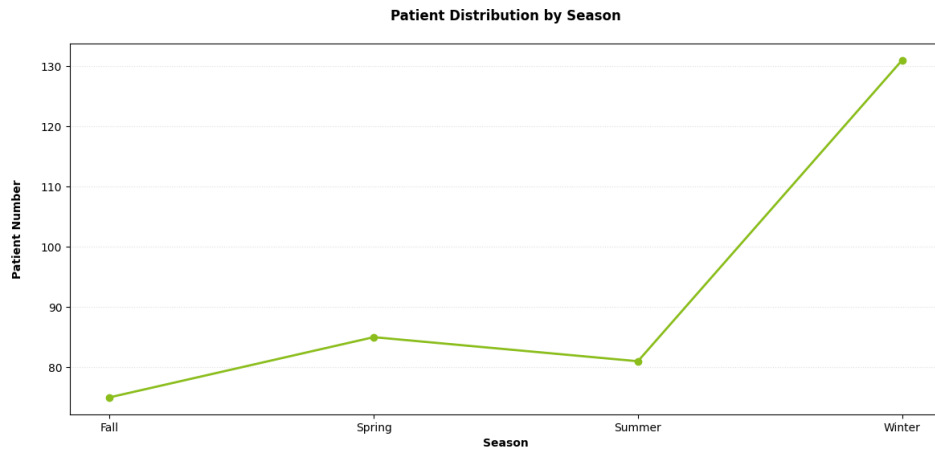


Figure 4.7. Patient Distribution by Season

From Figure 4.8, the distribution of anomalies can be observed across different months and over the years. Significant concentrations can be observed in April 2019 and June 2018. In 2021, there is a lower number of anomalies, which corresponds to the previously mentioned decrease in pneumonia patients during that year.

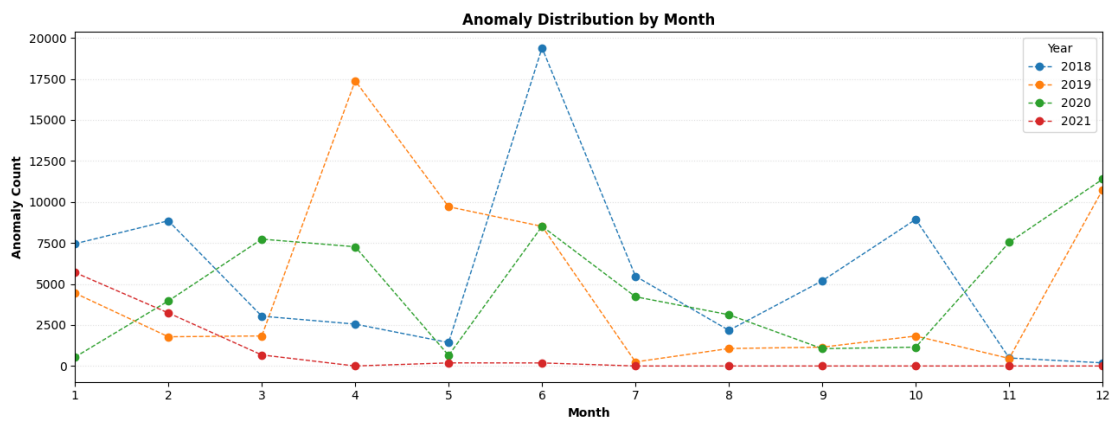


Figure 4.8. Anomaly Distribution by Month

Table 4.5 provides a breakdown of the Anomaly count by diagnosis. "Pneumonia nosocomial" has a higher likelihood of resulting in an anomaly compared to the other diagnoses.

Table 4.5. Anomaly Count by Diagnosis

Pneumonia	Anomaly Count
Pneumonia nosocomial	49 159
PNEUMONIAS BACTERIANAS NCOP	38 651
PNEUMONIA DEVIDA A KLEBSIELLA PNEUMONIAE	28 404
PNEUMONIAS BACTERIANAS NCOP + Pneumonia nosocomial	18 428
PNEUMONIA DEVIDA A PSEUDOMONAS	13 160
PNEUMONIA DEVIDA A BACTERIAS ESPECIFICADAS NCOP	9 830

PNEUMONIA DEVIDA A ESTAFILOCOCCOS AUREUS RESISTENTE A METICILINA + PNEUMONIAS BACTERIANAS NCOP	8 940
PNEUMONIA BACTERIANA NAO ESPECIFICADA + PNEUMONIA DEVIDA A KLEBSIELLA PNEUMONIAE	7 467
PNEUMONIA DEVIDA A PSEUDOMONAS + Pneumonia nosocomial	4 103
PNEUMONIA DEVIDA A HEMOPHILUS INFLUENZAE	2 204
PNEUMONIA DEVIDA A SARS	1 927
PNEUMONIA DEVIDA A ESCHERICHIA COLI E. COLI	1 888
PNEUMONIA DEVIDA A HEMOPHILUS INFLUENZAE + Pneumonia nosocomial	1 845
Pneumonia comunidade	1 084
Pneumonia C klebsiella pneumoniae + Pneumonia viral, SOE	935
Pneumonia associada a ventilador + Pneumonia nosocomial	891
PNEUMONIA DEVIDA A HEMOPHILUS INFLUENZAE + PNEUMONIA DEVIDA A STAPHYLOCOCCUS AUREUS	765
Abcesso pulmão c/pneumonia	536
PNEUMONIA DEVIDA A ESTAFILOCOCCOS AUREUS RESISTENTE A METICILINA	509
PNEUMONIA DEVIDA A STREPTOCOCCUS + PNEUMONIA DEVIDA A STREPTOCOCCUS DO GRUPO B	300
Pneumonia ou Pneumonite	118
Pneumonia pneumococica	93
Pneumonia deviva a coronavirusassoc/ ao sars	80
Pneumococcus + Pneumonia ou Pneuonite	68
PNEUMONIA DEVIDA A STAPHYLOCOCCUS + Pneumonia nosocomial	63
PNEUMONIA DEVIDA A STREPTOCOCCUS + Pneumonia aspiracao + Pneumonia comunidade	10

Two patients, each from different age groups but with the same length of stay, were selected to visualize their heart rate throughout their hospitalization period. The first patient, who is 78 years old, weighs 60 Kg, and has been diagnosed with “Pneumonia devida a pseudomonas + Pneumonia nosocomial”, was hospitalized for 10 days in 2018. The second patient, who is 1 year old, weighs 3.7 Kg, and has been diagnosed with “Pneumonia ou Pneuonite”, was hospitalized for 10 days in 2021.

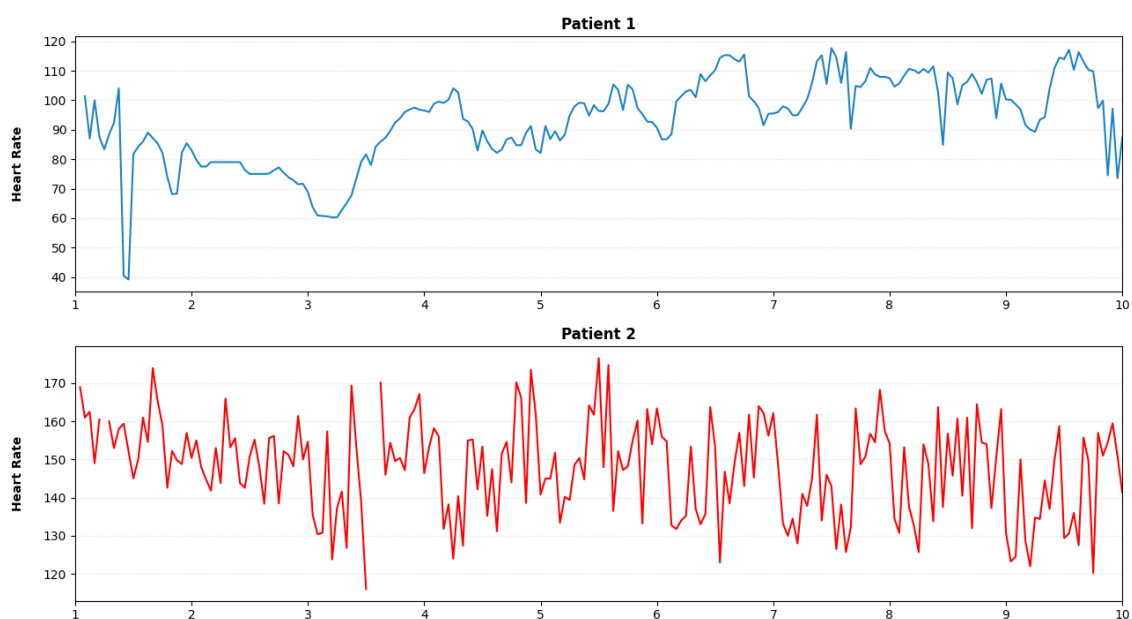


Figure 4.9. Heart Rate Comparison from Patient 1 (“Senior”) and Patient 2 (“Child”)

As observed in Figure 4.9, “Patient 1” exhibits a heart rate range of 40 to 120 beats per minute (bpm). In contrast, for "Patient 2," the heart rate range extends from approximately 120 to 180 bpm, exhibiting larger oscillations compared to “Patient 1”.

4.5.2 Myocarditis Dataset

Except for the specific issue of duplicate diagnoses observed exclusively in the pneumonia dataset and the fact that variables like "Temperature", "Respiration Rate from pCO", "ST Segment Lead V4", "ST Segment Lead V3", "ST Segment Lead V1", "Mean Arterial Pressure 2", "Diastolic Pressure (Art.) 2", "Systolic Pressure (Art.) 2" and “ANOMALY 3” were excluded from the analysis due to a significant number of missing values, the same data treatment procedures were applied to the myocarditis dataset. Variables derived from laboratory results also had a very limited number of values, becoming unfeasible for further analysis.

Following the data preparation procedures, the dataset underwent a reduction of 0.86%, leading to the removal of 189 records. As a result, the dataset now consists of 21604 records and 33 variables.

As depicted in Figure 4.10, during 2018 and 2019 there was a maximum number of four patients. This low patient count can be attributed to the rarity of the condition being studied. The higher volume of records in 2019 can be attributed to a specific patient who had a longer length of stay, 32 days, during that year.

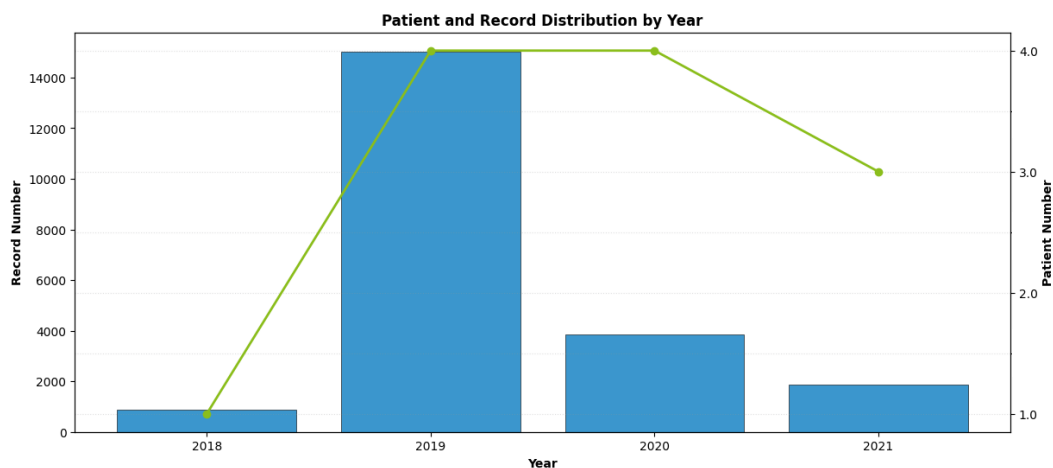


Figure 4.10. Patient and Record Distribution by Year

Unlike pneumonia, there are no “Senior” patients in this case. The distribution of patient numbers is nearly equal across both sexes (Figure 4.11).

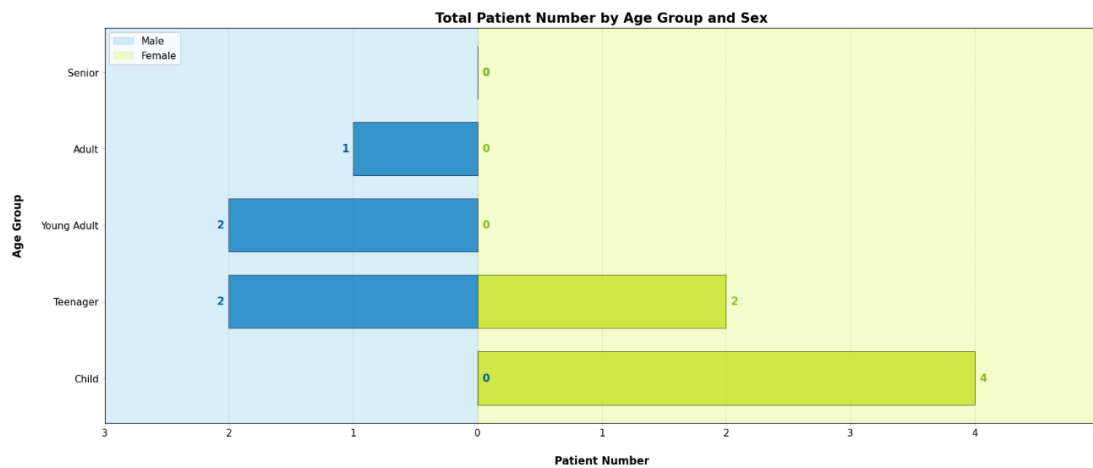


Figure 4.11. Total Patient Number by Age Group and Sex

Out of the patients in this dataset, only one individual exhibited anomalies associated to ST changes. This particular patient is a 60-year-old male and also holds the distinction of having the longest length of stay among all the patients in the dataset previously mentioned (32 days). During the patient's hospitalization, there was a noticeable change in the heart rate and SpO2 levels between the dates "2019-08-05" and "2019-08-09", as depicted in Figure 4.12. Specifically, there was a sudden decrease in heart rate and an increase in SpO2 levels during this time period, as well as a decrease in the occurrence of anomalies.

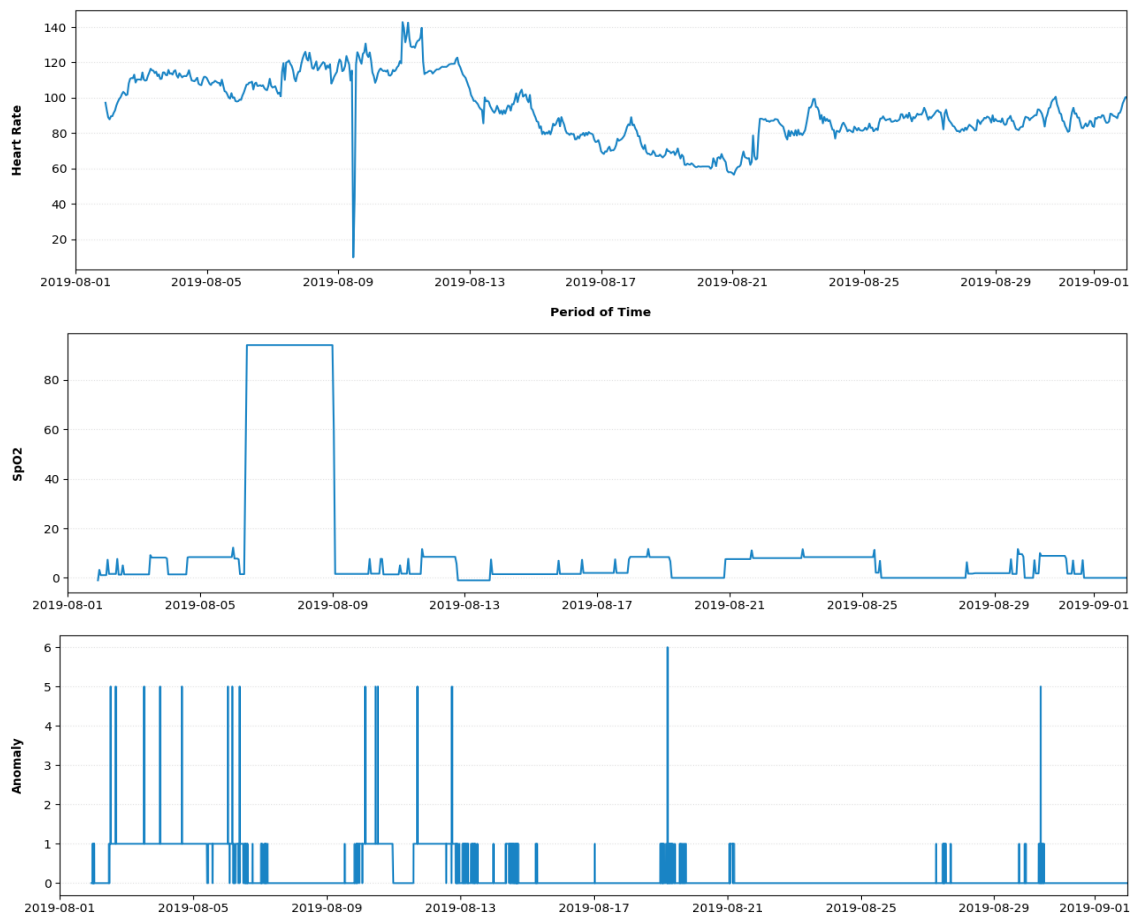


Figure 4.12. Patient with the Longest Length of Stay

4.6. Data Modeling

In this stage of the data mining, two clustering models were applied in each dataset to cluster the data into distinct groups based on their similarity. Among the algorithms initially tested, which included K-Means [67], Density-Based Clustering of Applications with Noise (DBSCAN) [68], Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [69] and Agglomerative Nesting (AGNES) [70], the DBSCAN and Agglomerative models were selected and retained for further modelling. These are algorithms that have been employed by other authors for clustering in the healthcare field, as evidenced in [71]–[80].

K-Means clustering was initially considered based on the state of the art findings. Nonetheless, it was discarded due to its requirement of choosing an appropriate number of clusters in advance and its sensitivity to outliers. The DBSCAN model was chosen due to its suitability for density-based problems. Unlike other models, the DBSCAN model does not require the specification of number of clusters and is capable of handling clusters of arbitrary shapes within spatial databases, while accounting for noise [81]. HDBSCAN was also considered as a potential clustering model, given it is an extension of DBSCAN. However, due to its tendency to produce a large number of clusters, which did not provide valuable insights for the analysis, it was ultimately rejected. The AGNES model was also chosen for the analysis. AGNES is a hierarchical clustering algorithm that constructs a hierarchy of clusters based on a similarity measure. One of the advantages of AGNES, similarly to DBSCAN, is that it does not require the number of clusters to be predetermined [81].

Given that DBSCAN is highly sensitive to its hyperparameters, the Clusteval Python package [82] was utilized to optimize the Epsilon parameter and determine the appropriate number of clusters. The same package was employed with the AGNES model to determine the optimal number of clusters based on the silhouette score. The silhouette analysis was employed as an evaluation technique in this study because the data used for clustering was unlabeled. For that reason, it was not possible to employ extrinsic evaluation methods such as homogeneity score or completeness score, which rely on the availability of ground truth labels [83]. The silhouette analysis measures the quality of clustering by assessing how well each observation is grouped within its assigned cluster and estimating the average distance between clusters [81].

To preprocess the data, the “get_dummies” [84] method was applied to the categorical variables. Additionally, feature selection was performed using a correlation matrix, the data was normalized with “MinMaxScaler” [85] and the t-Distributed Stochastic Neighbor Embedding [86] was employed to reduce the data dimensionality.

After establishing the clustering models and obtaining the results, the cluster values were added to the dataset that had not undergone preprocessing.

4.6.1 Pneumonia Clustering Model

Due to memory limitations on the utilized available server, the data used corresponded to a 12.5% sample of the original pneumonia dataset. Following the feature selection process, the variables 'Temperature', 'Respiration Rate from pCO', 'ST Segment Lead AVF', 'ST Segment Lead AVR', 'ST Segment Lead AVL', 'ST Segment Lead III', 'Mean Arterial Pressure 2', 'Diastolic Pressure (Art.) 2', 'Systolic Pressure (Art.) 2', 'Pulse', 'Heart Rate', 'WEIGHT', 'HEIGHT', 'Age', 'Sex', 'Season', 'Length of Stay', 'Previously Admitted', were selected.

The initial model developed using the DBSCAN algorithm resulted in a single cluster and a silhouette score of -0.235. A negative silhouette score might mean that the observations were assigned to the wrong cluster [81]. For that reason, the analysis of the resulting cluster is not presented here.

The second model was created with the AGNES algorithm, resulted in 3 clusters, visible in Figure 4.13. The silhouette score obtained was 0.290, which is closer to 1 than the previous model, indicating that there is a higher degree of separation between the clusters and cohesion within the clusters [81]. However, the score remains relatively low, suggesting that the clusters may not provide significant insights.

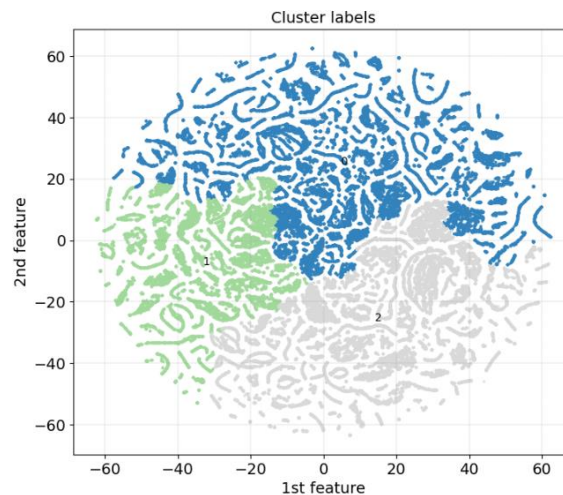


Figure 4.13. Pneumonia Clusters Identified by AGNES Clustering Model

As depicted in Figure 4.14, Cluster 2 presents the highest number of patients. The distribution of "Age" and "Temperature" exhibits similarity across the clusters. The central tendencies also demonstrate similarity among the clusters. Additionally, it is observed that "Male" prevails as the predominant sex across the clusters. Cluster 1 has 10 748 anomalies, the highest value of the 3 clusters.

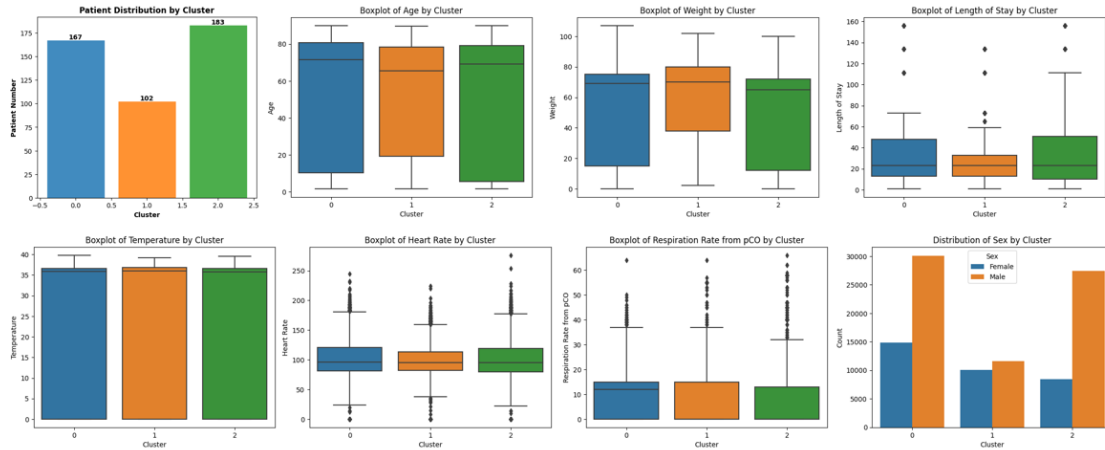


Figure 4.14. Pneumonia Clusters' Variables Distribution

4.6.2 Clustering Model - Myocarditis

Following the feature selection process, the variables selected were 'Heart Rate', 'Age', 'ST Segment Lead AVF', 'ST Segment Lead AVR', 'ST Segment Lead AVL', 'ST Segment Lead III', 'Pulse', 'WEIGHT', 'HEIGHT', 'Sex', 'Season', 'Length of Stay'. The first model, implemented using the DBSCAN algorithm, did not identify any clusters. However, the second model, built using the AGNES algorithm, successfully detected eight clusters, as illustrated in Figure 4.15. The silhouette score for this model was 0.311.

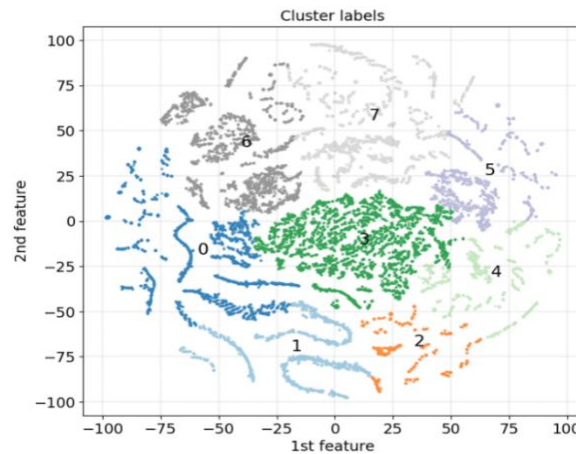


Figure 4.15. Myocarditis Clusters Identified by AGNES Clustering Model

Out of the eight clusters identified, four of them predominantly comprise female patients. In terms of age distribution, the first three clusters and the last cluster are characterized by a younger demographic.

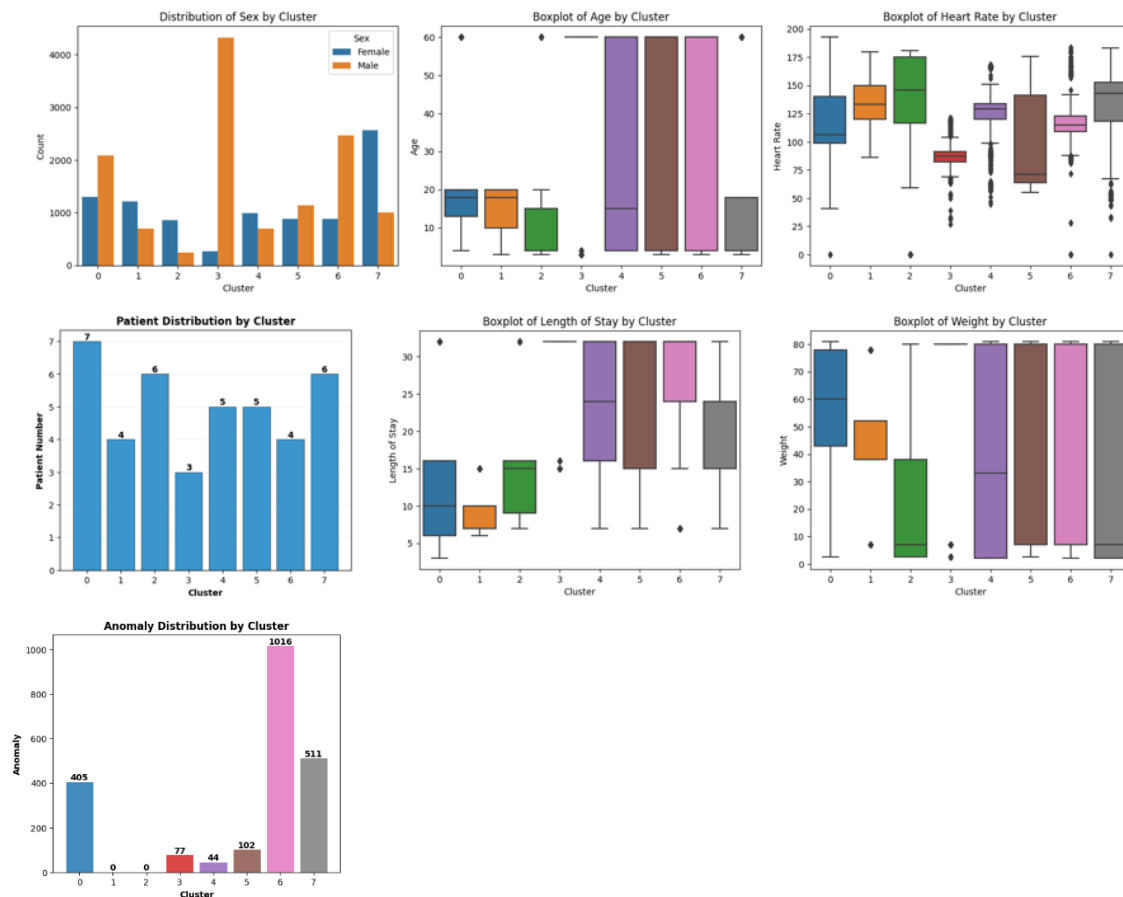


Figure 4.16. Myocarditis Clusters' Variables Distribution

4.6.3 Clustering Model - Pneumonia and Myocarditis

After multiple iterations of the previous models failed to achieve the desired results, an alternative approach was implemented. A selection process was conducted, choosing ten patients from each diagnosis, and ensuring a reasonably balanced distribution of records. Consequently, the dataset included 38521 records. The features selected were 'Heart Rate', 'Age', 'DIAGDESC', 'Pulse', 'WEIGHT', 'HEIGHT', 'Sex', 'Season', 'Length of Stay'.

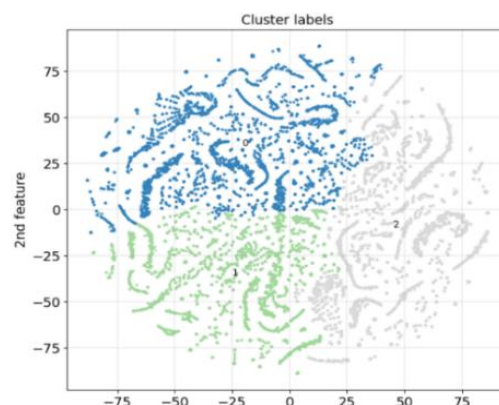


Figure 4.17. Pneumonia and Myocarditis Clusters Identified by AGNES Clustering Model

The first model, which employed the DBSCAN algorithm, did not find any cluster. On the other hand, AGNES found three clusters with a silhouette score of 0.327, visible in Figure 4.17.

As displayed in Figure 4.18, there are notable observations regarding the identified clusters. Cluster 0 and Cluster 2 predominantly consist of individuals from a younger demographic. Consequently, these clusters exhibit a comparatively greater prevalence of higher "Heart Rate" values. Regarding weight, the Cluster 1 demonstrates a higher proportion of individuals weighing over 60 Kg. Both Cluster 0 and Cluster 2 are primarily composed of female patients. Cluster 1 displays a larger quantity of anomalies and exhibits a higher proportion of myocarditis patients, suggesting a potential correlation between the two.

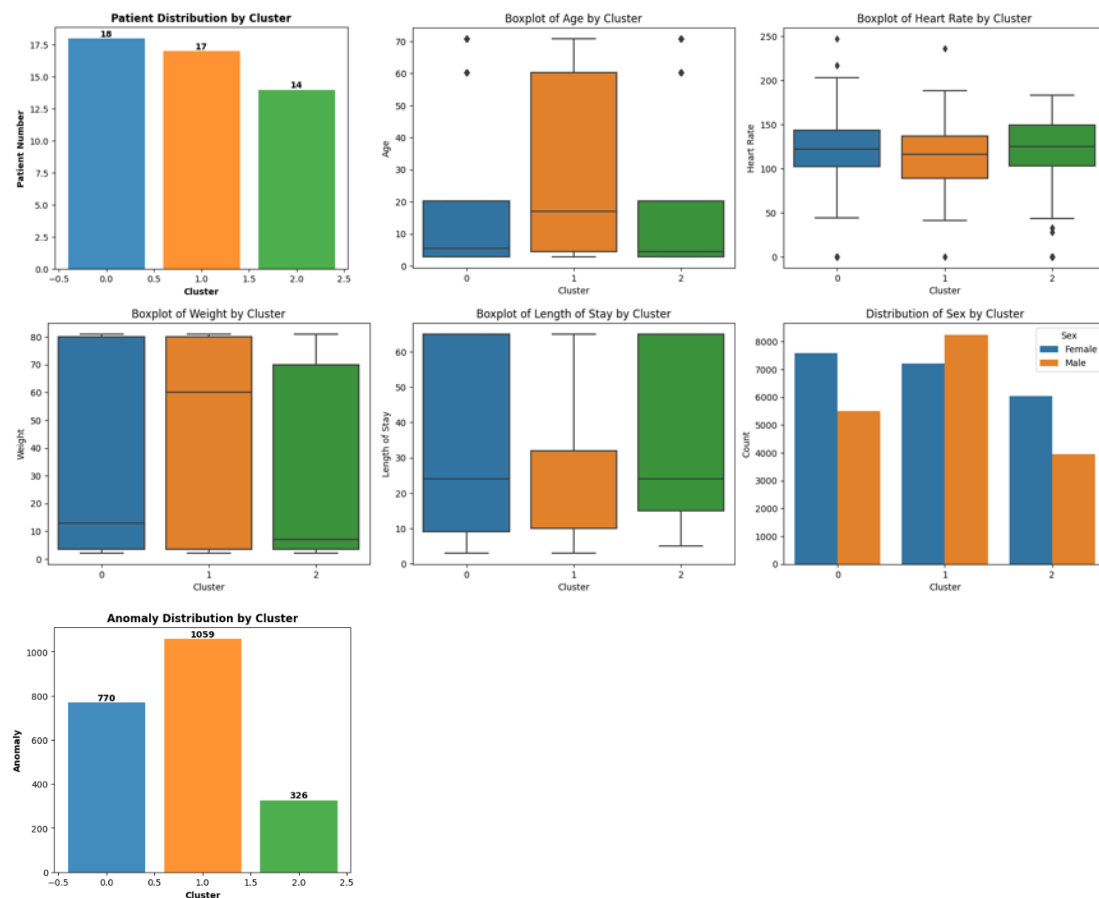


Figure 4.18. Myocarditis and Pneumonia Clusters' Variables Distribution

Evaluation

5.1. Introduction

This chapter includes an evaluation of both artifacts. As mentioned in the “Methodologies” section of this dissertation, the first artifact was evaluated according to hierarchical evaluation criteria for information systems artifacts proposed by Prat et al [87]. The evaluation of the second artifact, on the other hand, was conducted during a meeting with a physician.

5.2. Artifact 1

To perform the evaluation of the first artifact, the hierarchical evaluation criteria for information systems artifacts proposed by Prat et al. [87] was chosen, Figure 5.1. The evaluation of the second artifact was carried out according to the “Evaluation” phase of the CRISP-DM. The Data Collection, Monitoring and Health Alarm were selected as the capabilities to be evaluated, as presented in Table 5.1.

The assessment of the AIMHealth app was carried out by two physicians working at the Cardiology Service of Hospital de Santa Maria. These physicians were familiar with the app and had integrated it into their daily routine for data collection purposes. Over the period from November 2021 to July 2023, a total of 96 records were collected. These physicians were asked to evaluate using the ISO 15504 four-level NPLF scale [88], as illustrated in Table 5.2, in response to the provided questions outlined in Table 5.3.

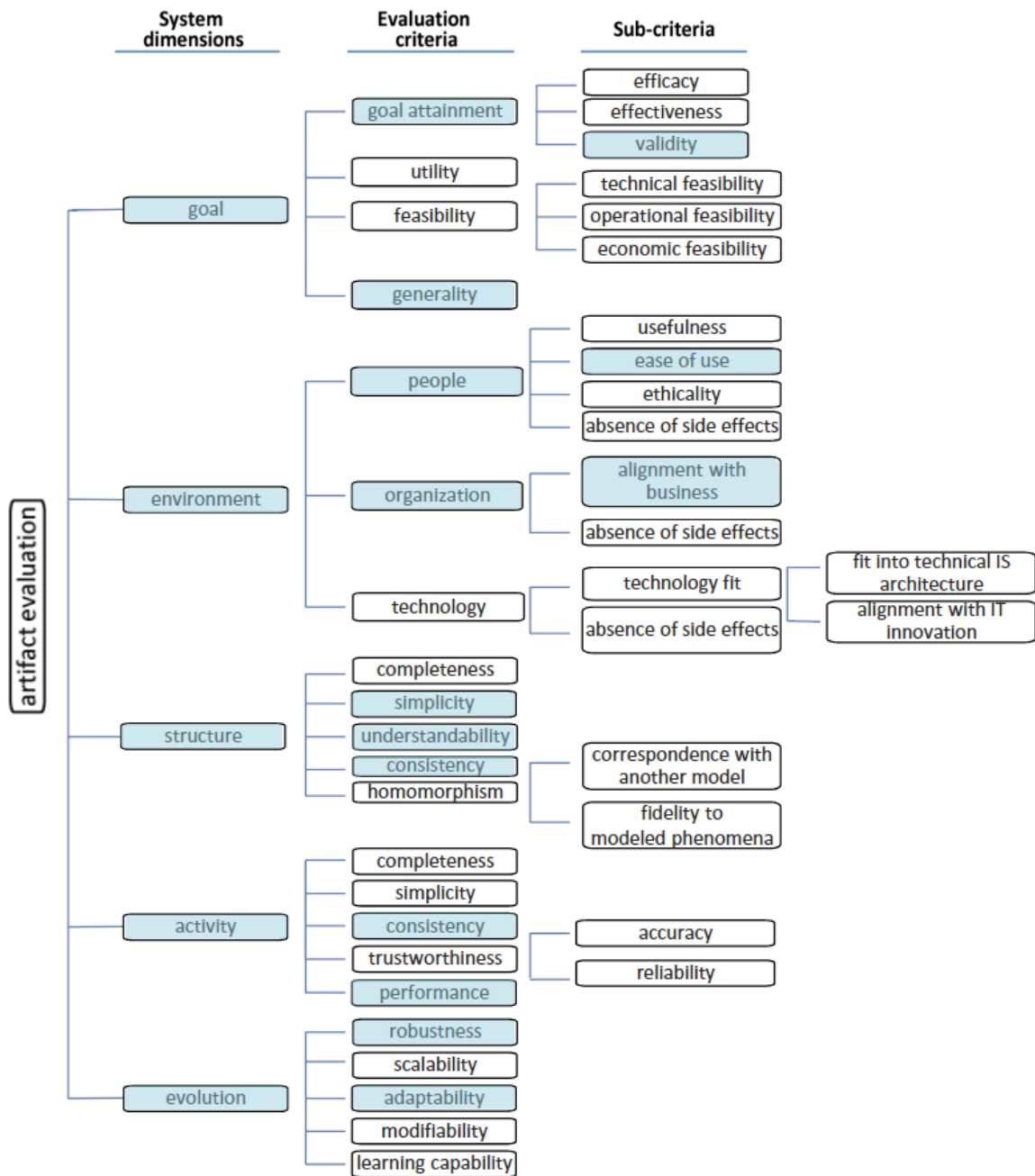


Figure 5.1. Information System Artifact Evaluation Criteria Hierarchy and Selected Criteria

Table 5.1. Capabilities, Dimensions, Criteria and Objective Statements used in the DSRM Evaluation

Capability	Dimension	Criteria	Objective Statement
Data Collection, Monitoring , Health Alarm	Goal	Goal Attainment - Validity	Data collection through the app is clinically valid
		Generality	The app is easily applied to other clinical departments
	Environment	People - Ease of Use	Easy to navigate through the app's features and functionalities
		Organization - Alignment with Business	Can successfully collect health data and monitor patient's health conditions
	Structure	Simplicity	The app's visual design is aesthetically appealing, and the information is well organized on the screens
		Understandability	The menu options, icons and buttons are clear indicators of the app's functionalities
		Consistency	The app maintains a consistent visual design
	Activity	Consistency	The app is able to recognize, diagnose and recover from errors or issues
		Performance	The app's features and functions have a good performance
	Evolution	Robustness	The app is robust in response to changes in the data collection environment
		Adaptability	The app quickly adapts to the user's needs

Table 5.2. ISO 15504 Four-Level NPLF Scale

Not Achieved (NA)	[0-15%]
Partially Achieved (PA)	]15-50%]
Largely Achieved (LA)	]50-85%]
Fully Achieved (FA)	]85-100%]

Table 5.3. App Evaluation Criteria, Questions, Objective Statements and Results

Criteria	Question	Objective Statement	Expert #1	Expert #2
Goal Attainment - Validity	On a scale of NA to FA, how clinically valid was the data collection through the app?	Data collection through the app is clinically valid	FA	LA
Generality	On a scale of NA to FA, how easy would it be to apply the app to other clinical departments?	The app is easily applied to other clinical departments	LA	PA
People - Ease of Use	On a scale of NA to FA, how easy was to navigate through the app's features and functionalities?	Easy to navigate through the app's features and functionalities	FA	FA
Organization – Alignment with Business	On a scale of NA to FA, did you find the app helpful in collecting health data and monitoring patient's health conditions?	Can successfully collect health data and monitor patient's health conditions	LA	FA
Simplicity	On a scale of NA to FA, did you find the app's visual design aesthetically appealing, and information well organized on the screens?	The app's visual design is aesthetically appealing, and the information is well organized on the screens	FA	FA
Understandability	On a scale of NA to FA, define how menu options, icons and buttons are clear indicators of the app's functionalities?	The menu options, icons and buttons are clear indicators of the app's functionalities	FA	FA
Consistency (Structure)	On a scale of NA to FA, did you find the overall app's visual design consistent throughout the app?	The app maintains a consistent visual design	FA	LA
Consistency (Activity)	On a scale of NA to FA, during the use of the app did you find the app able to recognize, diagnose and recover from errors or issues?	The app is able to recognize, diagnose and recover from errors or issues	LA	FA
Performance	On a scale of NA to FA, were all the app's features and functions working as expected?	The app's features and functions have a good performance	LA	FA

Robustness	On a scale of NA to FA, how robust is the app in response to changes in the data collection environment? Is the app robust when changing from online to offline mode?	The app is robust in response to changes in the data collection environment	FA	FA
Adaptability	On a scale of NA to FA, how fast has been the app to adapt to the user's needs?	The app quickly adapts to the user's needs	FA	FA

5.3. Artifact 2

In addition to the statistical assessment involving variable distribution within each cluster (sections 4.6.1, 4.6.2 and 4.6.3), meetings with a physician were conducted. These aimed to gain a deeper comprehension of the inherent cluster characteristics and to assess the potential insights offered by the models.

Between May 25th and July 17th, 2023, weekly meetings were conducted with a physician who is affiliated with the Cardiology Service of Hospital de Santa Maria. During these meetings, the data mining process was reviewed and fine-tuned based on the feedback provided by the physician. After employing various clustering algorithms and consulting with the physician, it was decided that the results obtained did not provide significant insights into health patterns. This outcome could be attributed to the lack of distinct characteristics and data points that differentiate patients from one another.

Despite the results obtained, it is visible that the models utilizing data from myocarditis and data from both myocarditis and pneumonia patients achieved higher silhouette scores.

Conclusions

6.1. Discussion and Conclusions

This dissertation focused on the development of the AIMHealth app, an AI-based mHealth application. The primary objective of this app is monitoring patients with various health conditions and providing predictive capabilities for forecasting disease worsening or decompensation. The research questions that guided this dissertation were:

- How can we optimize the user experience of the AIMHealth mHealth app to ensure ease of use and increased patient engagement during disease monitoring?
- Can we identify meaningful patterns in health data using data mining techniques and machine learning algorithms, within the context of the AIMHealth app's objectives?

With these research questions guiding the dissertation, two distinct artifacts were created: "Artifact 1 - AIMHealth App for Data Collection" and "Artifact 2 - Data Mining for the AIMHealth App", each designed to address one of these research questions comprehensively.

Regarding the first artifact, a series of actions were executed to enhance the user experience of the AIMHealth app, using a user-centered design approach. Firstly, the UX research was conducted with 5 end-users as proposed by UX expert Jakob Nielsen [89], which involved collecting user insights and feedback to better understand the user's needs. With this information, a user persona was created, this persona represents the typical user which helped tailoring the app's design and features. After that, functional and non-functional requirements, and the user flow, were defined for the app, which led to the development of the low and high-fidelity prototypes. These helped to visualize the app's interface and functionalities. Lastly, based on the research and the prototypes, especially the high-fidelity prototype, the final improvements were made to the app's user interface, that was put into production in the Cardiology Service of the Hospital Santa Maria. Overall, the first artifact was well received by the 2 experts (medical doctor), following a user evaluation that adopted the ISO 15504 four-level NPLF scale, that are using the app in their clinical practice and routines.

The second artifact was based on the second research question, and through data visualization, discernible patterns were identified in the pneumonia dataset. We could observe what age groups were more commonly affected by pneumonia, the variation of patient influx in

terms of seasonal analysis, as well as the correlation between heart rate and age group, among others. As for the myocarditis dataset, due to the considerably lower patient number, not a lot of significant insights were extracted.

The results of the clustering model performed on the dataset that combined both myocarditis and pneumonia diagnoses demonstrated a better performance in terms of detecting distinct clusters and achieving a higher silhouette score. The models derived from the pneumonia and myocarditis datasets did not attain good clustering results, possibly due to insufficient distinguishing features within the data. In summary, the second artifact did not deliver favorable results, and the overall evaluation was inconclusive.

In conclusion, throughout this dissertation, a multidisciplinary approach was adopted, including user centered UX design, app development, data mining, and machine learning. Even though the second artifact did not gather significant insights from the data mining process, it lays a foundation for future research.

6.2. Research Limitations

The dissertation encountered several limitations, primarily associated with the availability and nature of the data. Firstly, the data collected from the app was not accessible at the time of the study, which led to the utilization of EHRs instead. The EHR data presented a high volume of variables and values, however, a substantial portion of them had missing values. The presence of these missing values and overall data quality issues limited the accuracy of the analysis.

6.3. Contributions

Through this dissertation, one contribution to the scientific community was made with a conference paper in Innovations in Bio-Inspired Computing and Applications regarding the AIMHealth app named 'AI-Based mHealth App for Covid-19 or Cardiac Diseases Diagnosis and Prognosis' [48].

6.4. Future Work

In the future, iterative improvements to the interface based on user feedback, attained by testing the high-fidelity prototype and/or the app in production with a sample of five end-users could be pursued. Expanding this work to include other diseases, as well as extending the time and incorporating a more comprehensive data from the available HSM dataset, could be a valuable direction to pursue. Additionally, exploring alternative machine learning models

beyond those used could be an opportunity to compare and assess their effectiveness in addressing the research objectives.

With further refinement, the models developed in this dissertation, not yet integrated into the AIMHealth app due to their unsatisfactory performance, could be considered for future integration into the app, from the UX perspective, following the insights gathered in the user requirements stage.

References

- [1] K. Peffers, T. Tuunanen, M. Rothenberger, and S. Chatterjee, 'A design science research methodology for information systems research', *Journal of Management Information Systems*, vol. 24, pp. 45–77, Jan. 2007.
- [2] 'Coronavirus disease (COVID-19)'. Accessed: Sep. 24, 2022. [Online]. Available: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19>
- [3] H. Rohmetra, N. Raghunath, P. Narang, V. Chamola, M. Guizani, and N. R. Lakkaniga, 'AI-enabled remote monitoring of vital signs for COVID-19: methods, prospects and challenges', *Computing*, 2021, doi: 10.1007/s00607-021-00937-7.
- [4] M. Polsinelli, L. Cinque, and G. Placidi, 'A light CNN for detecting COVID-19 from CT scans of the chest', *Pattern Recognition Letters*, vol. 140, pp. 95–100, 2020, doi: 10.1016/j.patrec.2020.10.001.
- [5] U. Sait *et al.*, 'A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images', *Applied Soft Computing*, vol. 109, 2021, doi: 10.1016/j.asoc.2021.107522.
- [6] 'European mhealth hub | Use case of disease monitoring and self-management – example of heart failure'. Accessed: Oct. 18, 2022. [Online]. Available: <https://mhealth-hub.org/use-case-of-disease-monitoring-and-self-management-example-of-heart-failure>
- [7] X. Jiang, J. Yao, and J. You, 'Cost-effectiveness of a Telemonitoring Program for Patients With Heart Failure During the COVID-19 Pandemic in Hong Kong: Model Development and Data Analysis', *JOURNAL OF MEDICAL INTERNET RESEARCH*, vol. 23, no. 3, Mar. 2021, doi: 10.2196/26516.
- [8] T. Jaarsma *et al.*, 'Heart failure management programmes in Europe', *Eur J Cardiovasc Nurs*, vol. 5, no. 3, pp. 197–205, Sep. 2006, doi: 10.1016/j.ejcnurse.2006.04.002.
- [9] J. G. F. Cleland, 'Patients with treatable malignant diseases--including heart failure--are entitled to specialist care', *CMAJ*, vol. 172, no. 2, pp. 207–209, Jan. 2005, doi: 10.1503/cmaj.045307.
- [10] K. Kalaiselvan, M. Sahithullah, G. Diron Balachandaran, V. Sakthi, and M. Srianth, 'Smart healthcare support for remote patient monitoring', presented at the 12th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2021, 2021, pp. 967–972. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117785111&partnerID=40&md5=b874b93519ad529d1b0efa7098bef0c3>
- [11] M. Peyroteo, I. A. Ferreira, L. B. Elvas, J. C. Ferreira, and L. V. Lapão, 'Remote Monitoring Systems for Patients With Chronic Diseases in Primary Health Care: Systematic Review', *JMIR Mhealth Uhealth*, vol. 9, no. 12, p. e28285, Dec. 2021, doi: 10.2196/28285.
- [12] C. Manta, S. Jain, A. Coravos, D. Mendelsohn, and E. Izmailova, 'An Evaluation of Biometric Monitoring Technologies for Vital Signs in the Era of COVID-19', *Clinical and Translational Science*, vol. 13, Aug. 2020, doi: 10.1111/cts.12874.
- [13] 'Heart Infection: Causes, Symptoms & Treatment', Cleveland Clinic. Accessed: Sep. 24, 2022. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/22054-heart-infection>
- [14] P. Ware *et al.*, 'Challenges of Telemonitoring Programs for Complex Chronic Conditions: Randomized Controlled Trial With an Embedded Qualitative Study', *JOURNAL OF MEDICAL INTERNET RESEARCH*, vol. 24, no. 1, Jan. 2022, doi: 10.2196/31754.
- [15] L. B. Elvas, A. G. Almeida, L. Rosario, M. S. Dias, and J. C. Ferreira, 'Calcium Identification and Scoring Based on Echocardiography. An Exploratory Study on Aortic Valve Stenosis', *Journal of Personalized Medicine*, vol. 11, no. 7, Art. no. 7, Jul. 2021, doi: 10.3390/jpm11070598.

- [16] M. Usama, B. Ahmad, W. Xiao, M. S. Hossain, and G. Muhammad, 'Self-attention based recurrent convolutional neural network for disease prediction using healthcare data', *Computer Methods and Programs in Biomedicine*, vol. 190, p. 105191, Jul. 2020, doi: 10.1016/j.cmpb.2019.105191.
- [17] 'European mhealth hub | Artificial Intelligence (AI)'. Accessed: Oct. 17, 2022. [Online]. Available: <https://mhealth-hub.org/artificial-intelligence-ai>
- [18] 'Sample Page', AIMhealth. Accessed: Oct. 01, 2023. [Online]. Available: <https://istar.iscte-iul.pt/aimhealth/>
- [19] K. Boikanyo, A. M. Zungeru, B. Sigweni, A. Yahya, and C. Lebekwe, 'Remote patient monitoring systems: Applications, architecture, and challenges', *Scientific African*, vol. 20, p. e01638, Jul. 2023, doi: 10.1016/j.sciaf.2023.e01638.
- [20] J. Espey and H. Dahmm, 'Innovations in Disease Surveillance and Monitoring', in *Handbook of Global Health*, R. Haring, I. Kickbusch, D. Ganten, and M. Moeti, Eds., Cham: Springer International Publishing, 2020, pp. 1–21. doi: 10.1007/978-3-030-05325-3_16-1.
- [21] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, 'Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement', *BMJ*, vol. 339, p. b2535, Jul. 2009, doi: 10.1136/bmj.b2535.
- [22] C. Abras, D. Maloney-Krichmar, and J. Preece, 'User-centered design', *User-Centered Design*, pp. 445–456, Jan. 2004.
- [23] 'IBM SPSS Modeler CRISP-DM Guide'.
- [24] N. Shimpi, S. McRoy, H. Zhao, M. Wu, and A. Acharya, 'Development of a periodontitis risk assessment model for primary care providers in an interdisciplinary setting', *Technology and Health Care*, vol. 28, no. 2, pp. 143–154, 2020, doi: 10.3233/THC-191642.
- [25] A. Shabbir, M. Shabbir, A. R. Javed, M. Rizwan, C. Iwendi, and C. Chakraborty, 'Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals', *Journal of Experimental and Theoretical Artificial Intelligence*, 2022, doi: 10.1080/0952813X.2021.1960634.
- [26] F. Babic, L. Majnaric, S. Bekic, and A. Holzinger, 'Machine Learning for Family Doctors: A Case of Cluster Analysis for Studying Aging Associated Comorbidities and Frailty', presented at the MACHINE LEARNING AND KNOWLEDGE EXTRACTION, CD-MAKE 2019, A. Holzinger, P. Kieseberg, A. Tjoa, and E. Weippl, Eds., 2019, pp. 178–194. doi: 10.1007/978-3-030-29726-8_12.
- [27] Z.-X. Li, Y.-M. Zha, G.-Y. Jiang, and Y.-X. Huang, 'AI Aided Analysis on Saliva Crystallization of Pregnant Women for Accurate Estimation of Delivery Date and Fetal Status', *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2320–2330, 2022, doi: 10.1109/JBHI.2021.3135534.
- [28] M. Bohanec *et al.*, 'HeartMan DSS: A decision support system for self-management of congestive heart failure', *Expert Systems with Applications*, vol. 186, 2021, doi: 10.1016/j.eswa.2021.115688.
- [29] C. Ke *et al.*, 'Prognostics of surgical site infections using dynamic health data', *Journal of Biomedical Informatics*, vol. 65, pp. 22–33, 2017, doi: 10.1016/j.jbi.2016.10.021.
- [30] M. Goni, S. B. Eickhoff, M. S. Far, K. R. Patil, and J. Dukart, 'Smartphone-Based Digital Biomarkers for Parkinson's Disease in a Remotely-Administered Setting', *IEEE Access*, vol. 10, pp. 28361–28384, 2022, doi: 10.1109/ACCESS.2022.3156659.
- [31] Y. Wang *et al.*, 'Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records', *Journal of Biomedical Informatics*, vol. 102, 2020, doi: 10.1016/j.jbi.2019.103364.
- [32] C. Lee and M. van der Schaar, 'Temporal Phenotyping using Deep Predictive Clustering of Disease Progression', presented at the INTERNATIONAL CONFERENCE ON MACHINE LEARNING, VOL 119, H. Daume and A. Singh, Eds., 2020.
- [33] S. C. Yu *et al.*, 'Sepsis Prediction for the General Ward Setting', *Frontiers in Digital Health*, vol. 4, 2022, doi: 10.3389/fdgth.2022.848599.

- [34] N. Alexander, D. C. Alexander, F. Barkhof, and S. Denaxas, 'Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning', *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, 2021, doi: 10.1186/s12911-021-01693-6.
- [35] J. G. Diaz Ochoa and F. E. Mustafa, 'Graph neural network modelling as a potentially effective method for predicting and analyzing procedures based on patients' diagnoses', *Artificial Intelligence in Medicine*, vol. 131, 2022, doi: 10.1016/j.artmed.2022.102359.
- [36] M. Bernardini, M. Morettini, L. Romeo, E. Frontoni, and L. Burattini, 'Early temporal prediction of Type 2 Diabetes Risk Condition from a General Practitioner Electronic Health Record: A Multiple Instance Boosting Approach', *Artificial Intelligence in Medicine*, vol. 105, 2020, doi: 10.1016/j.artmed.2020.101847.
- [37] R. Ramaswamy *et al.*, 'CKD subpopulations defined by risk-factors: A longitudinal analysis of electronic health records', *CPT: Pharmacometrics and Systems Pharmacology*, vol. 10, no. 11, pp. 1343–1356, 2021, doi: 10.1002/psp4.12695.
- [38] R. J. Kate, N. Pearce, D. Mazumdar, and V. Nilakantan, 'A continual prediction model for inpatient acute kidney injury', *Computers in Biology and Medicine*, vol. 116, 2020, doi: 10.1016/j.combiomed.2019.103580.
- [39] I. Landi *et al.*, 'Deep representation learning of electronic health records to unlock patient stratification at scale', *npj Digital Medicine*, vol. 3, no. 1, 2020, doi: 10.1038/s41746-020-0301-z.
- [40] S. M. Lauritsen *et al.*, 'Early detection of sepsis utilizing deep learning on electronic health record event sequences', *Artificial Intelligence in Medicine*, vol. 104, 2020, doi: 10.1016/j.artmed.2020.101820.
- [41] E. R. Pfaff *et al.*, 'Identifying who has long COVID in the USA: a machine learning approach using N3C data', *The Lancet Digital Health*, vol. 4, no. 7, pp. e532–e541, 2022, doi: 10.1016/S2589-7500(22)00048-6.
- [42] S. D. Mohanty, D. Lekan, T. P. McCoy, M. Jenkins, and P. Manda, 'Machine learning for predicting readmission risk among the frail: Explainable AI for healthcare', *Patterns*, vol. 3, no. 1, 2022, doi: 10.1016/j.patter.2021.100395.
- [43] L. Rasmy *et al.*, 'Recurrent neural network models (CovRNN) for predicting outcomes of patients with COVID-19 on admission to hospital: model development and validation using electronic health record data', *The Lancet Digital Health*, vol. 4, no. 6, pp. e415–e425, 2022, doi: 10.1016/S2589-7500(22)00049-8.
- [44] Z. Nenova and J. Shang, 'Chronic Disease Progression Prediction: Leveraging Case-Based Reasoning and Big Data Analytics', *Production and Operations Management*, vol. 31, no. 1, pp. 259–280, 2022, doi: 10.1111/poms.13532.
- [45] H. Wang, Q. Zhang, F. Y. Chen, E. Y. M. Leung, E. L. Y. Wong, and E. K. Yeoh, 'Tensor Factorization-Based Prediction with an Application to Estimate the Risk of Chronic Diseases', *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 53–61, 2021, doi: 10.1109/MIS.2021.3071018.
- [46] F. Zhou, A. Gillespie, D. Gligorijevic, J. Gligorijevic, and Z. Obradovic, 'Use of disease embedding technique to predict the risk of progression to end-stage renal disease', *Journal of Biomedical Informatics*, vol. 105, 2020, doi: 10.1016/j.jbi.2020.103409.
- [47] M. D. Rinderknecht and Y. Klopfenstein, 'Predicting critical state after COVID-19 diagnosis: model development using a large US electronic health record dataset', *npj Digital Medicine*, vol. 4, no. 1, 2021, doi: 10.1038/s41746-021-00482-9.
- [48] A. Vieira *et al.*, 'AI-Based mHealth App for Covid-19 or Cardiac Diseases Diagnosis and Prognosis', in *Innovations in Bio-Inspired Computing and Applications*, A. Abraham, A. Bajaj, N. Gandhi, A. M. Madureira, and C. Kahraman, Eds., in *Lecture Notes in Networks and Systems*. Cham: Springer Nature Switzerland, 2023, pp. 765–777. doi: 10.1007/978-3-031-27499-2_71.

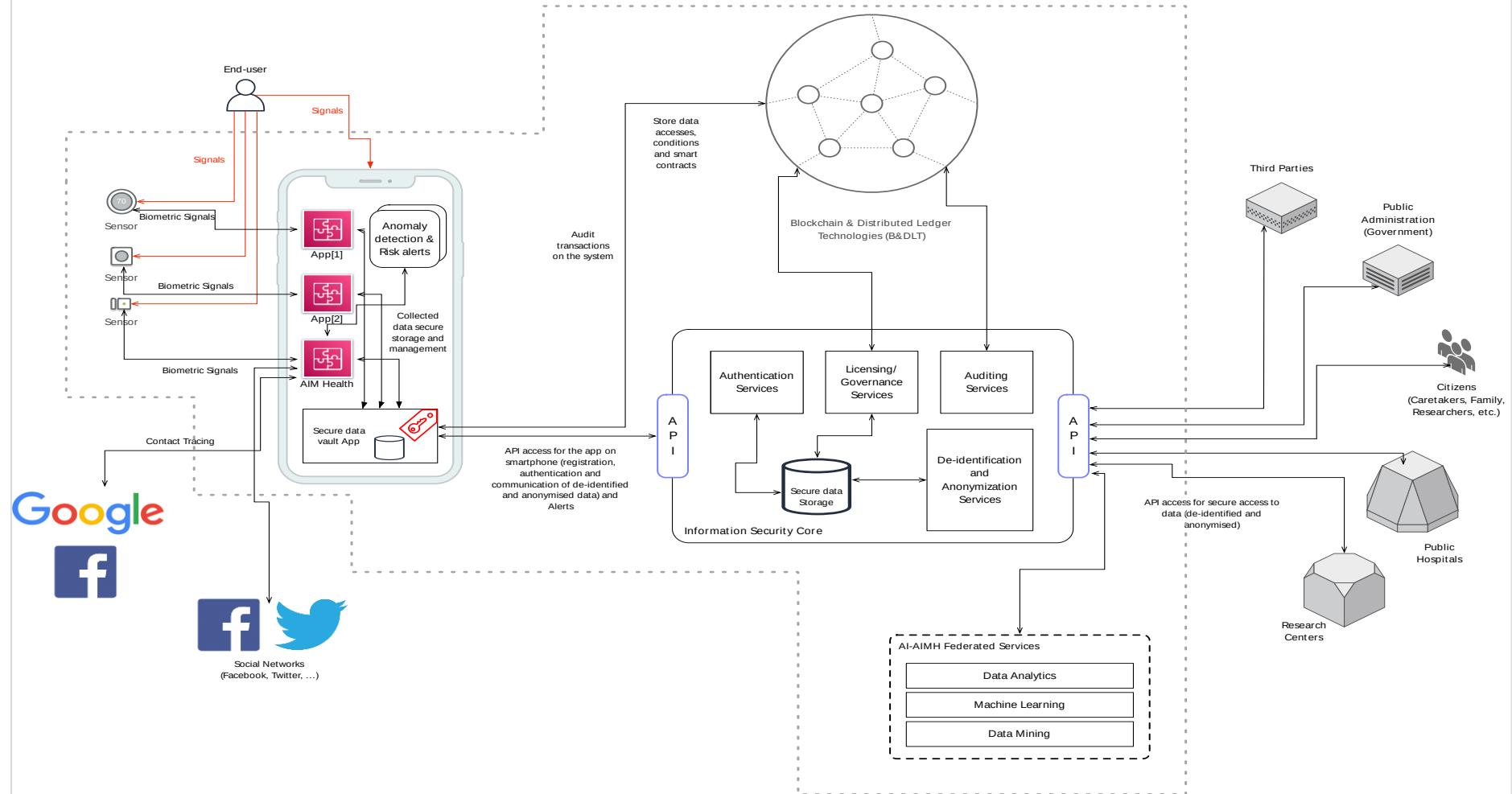
- [49] 'Flutter - Build apps for any screen'. Accessed: Oct. 10, 2022. [Online]. Available: [//flutter.dev/](https://flutter.dev/)
- [50] A. Raposo *et al.*, 'e-CoVig: A Novel mHealth System for Remote Monitoring of Symptoms in COVID-19', *Sensors*, vol. 21, no. 10, Art. no. 10, Jan. 2021, doi: 10.3390/s21103397.
- [51] A. S. for P. Affairs, 'System Usability Scale (SUS)'. Accessed: Oct. 11, 2022. [Online]. Available: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>
- [52] S. T. Acuña, J. W. Castro, and N. Juristo, 'A HCI technique for improving requirements elicitation', *Information and Software Technology*, vol. 54, no. 12, pp. 1357–1375, Dec. 2012, doi: 10.1016/j.infsof.2012.07.011.
- [53] B. Ferreira, W. Silva, E. Oliveira, and T. Conte, *Designing Personas with Empathy Map*. 2015. doi: 10.18293/SEKE2015-152.
- [54] Y.-K. Lim, E. Stolterman, and J. Tenenbergh, 'The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas', *ACM Trans. Comput.-Hum. Interact.*, vol. 15, no. 2, pp. 1–27, Jul. 2008, doi: 10.1145/1375761.1375762.
- [55] D. Wallach and S. C. Scholz, 'User-Centered Design: Why and How to Put Users First in Software Development', in *Software for People*, A. Maedche, A. Botzenhardt, and L. Neer, Eds., in Management for Professionals. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 11–38. doi: 10.1007/978-3-642-31371-4_2.
- [56] R. Sefelin, M. Tscheligi, and V. Giller, 'Paper prototyping - what is it good for?: a comparison of paper- and computer-based low-fidelity prototyping', in *CHI '03 extended abstracts on Human factors in computing systems - CHI '03*, Ft. Lauderdale, Florida, USA: ACM Press, 2003, p. 778. doi: 10.1145/765891.765986.
- [57] 'Figma', Figma. Accessed: Jan. 30, 2023. [Online]. Available: <https://www.figma.com/files/recent?fuid=1147474809979700666>
- [58] 'Appwrite - Open-Source End-to-End Backend Server', Appwrite. Accessed: Oct. 06, 2022. [Online]. Available: <https://appwrite.io/>
- [59] 'DataGrip: The Cross-Platform IDE for Databases & SQL by JetBrains', JetBrains. Accessed: Jun. 21, 2023. [Online]. Available: <https://www.jetbrains.com/datagrip/>
- [60] 'Welcome to Python.org', Python.org. Accessed: Jun. 21, 2023. [Online]. Available: <https://www.python.org/>
- [61] 'JupyterLab Documentation — JupyterLab 4.0.2 documentation'. Accessed: Jun. 21, 2023. [Online]. Available: <https://jupyterlab.readthedocs.io/en/latest/>
- [62] 'pandas - Python Data Analysis Library'. Accessed: Nov. 05, 2022. [Online]. Available: <https://pandas.pydata.org/>
- [63] 'NumPy'. Accessed: Jan. 30, 2023. [Online]. Available: <https://numpy.org/>
- [64] 'Matplotlib — Visualization with Python'. Accessed: Dec. 08, 2022. [Online]. Available: <https://matplotlib.org/>
- [65] M. Waskom, 'seaborn: statistical data visualization', *JOSS*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/joss.03021.
- [66] K. Thygesen *et al.*, 'Fourth universal definition of myocardial infarction (2018)', *European Heart Journal*, vol. 40, no. 3, pp. 237–269, Jan. 2019, doi: 10.1093/eurheartj/ehy462.
- [67] 'sklearn.cluster.KMeans', scikit-learn. Accessed: Jul. 17, 2023. [Online]. Available: <https://scikit-learn/stable/modules/generated/sklearn.cluster.KMeans.html>
- [68] 'sklearn.cluster.DBSCAN', scikit-learn. Accessed: Jul. 17, 2023. [Online]. Available: <https://scikit-learn/stable/modules/generated/sklearn.cluster.DBSCAN.html>
- [69] 'sklearn.cluster.HDBSCAN', scikit-learn. Accessed: Jul. 17, 2023. [Online]. Available: <https://scikit-learn/stable/modules/generated/sklearn.cluster.HDBSCAN.html>
- [70] 'sklearn.cluster.AgglomerativeClustering', scikit-learn. Accessed: Jul. 17, 2023. [Online]. Available: <https://scikit-learn/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>

- [71] R. Bateja, S. K. Dubey, and A. Bhatt, 'Evaluation and Application of Clustering Algorithms in Healthcare Domain Using Cloud Services', in *Second International Conference on Sustainable Technologies for Computational Intelligence*, A. K. Luhach, R. C. Poonia, X.-Z. Gao, and D. Singh Jat, Eds., in *Advances in Intelligent Systems and Computing*. Singapore: Springer, 2022, pp. 249–261. doi: 10.1007/978-981-16-4641-6_21.
- [72] P. Wang and J. Li, 'Implementation of Real-Time Medical and Health Data Mining System Based on Machine Learning', *Journal of Healthcare Engineering*, vol. 2021, pp. 1–5, Nov. 2021, doi: 10.1155/2021/7011205.
- [73] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, 'Hybrid Prediction Model for Type 2 Diabetes and Hypertension Using DBSCAN-Based Outlier Detection, Synthetic Minority Over Sampling Technique (SMOTE), and Random Forest', *Applied Sciences*, vol. 8, no. 8, Art. no. 8, Aug. 2018, doi: 10.3390/app8081325.
- [74] J. Zhang, W. Xu, J. Guo, and S. Gao, 'A temporal model in Electronic Health Record search', *Knowledge-Based Systems*, vol. 126, pp. 56–67, Jun. 2017, doi: 10.1016/j.knosys.2017.03.029.
- [75] S. O. Adams, H. U. Yahaya, and T. Mohammed, 'Cluster Analysis of HIV/AIDs Incidence in Sub-Saharan Africa (1990 – 2018)', *IJEHS*, vol. 4, no. 2, May 2023, doi: 10.51757/IJEHS.4.2023.701311.
- [76] A. Rueda and S. Krishnan, 'Clustering Parkinson's and Age-Related Voice Impairment Signal Features for Unsupervised Learning', *Adv. Data Sci. Adapt. Data Anal.*, vol. 10, no. 02, p. 1840007, Apr. 2018, doi: 10.1142/S2424922X18400077.
- [77] D. H. Solomon *et al.*, 'Assessing clusters of comorbidities in rheumatoid arthritis: a machine learning approach', *Arthritis Res Ther*, vol. 25, no. 1, p. 224, Nov. 2023, doi: 10.1186/s13075-023-03191-8.
- [78] Y. Nia, H. Alavi Majd, M. Zamanian Azodi, and N. Khayer, 'Using partitioning and non-partitioning clustering algorithms for included proteins sequences in esophagus, stomach and colon cancer', *Journal of Paramedical Sciences*, vol. 2, Jan. 2011.
- [79] K. I. Neijenhuijs, C. F. W. Peeters, H. van Weert, P. Cuijpers, and I. V. Leeuw, 'Symptom clusters among cancer survivors: what can machine learning techniques tell us?', *BMC Medical Research Methodology*, vol. 21, no. 1, p. 166, Aug. 2021, doi: 10.1186/s12874-021-01352-4.
- [80] C.-H. Lin, K.-C. Hsu, K. R. Johnson, M. Luby, and Y. C. Fann, 'Applying density-based outlier identifications using multiple datasets for validation of stroke clinical outcomes', *International Journal of Medical Informatics*, vol. 132, p. 103988, Dec. 2019, doi: 10.1016/j.ijmedinf.2019.103988.
- [81] D. B. Davidson, *Chapter 6 Cluster Analysis | Data Analytics Living Textbook*. Accessed: Jun. 29, 2023. [Online]. Available: https://bookdown.org/brittany_davidson1993/bookdown-demo/cluster-analysis.html#density-based-clustering
- [82] 'Background — clusteval clusteval documentation'. Accessed: Jun. 30, 2023. [Online]. Available: <https://erdogant.github.io/clusteval/pages/html/Abstract.html#aim>
- [83] M. Chaudhary, 'Silhouette Analysis in K-means Clustering', Medium. Accessed: Jul. 17, 2023. [Online]. Available: <https://medium.com/@cmukesh8688/silhouette-analysis-in-k-means-clustering-cefa9a7ad111>
- [84] 'pandas.get_dummies — pandas 2.0.3 documentation'. Accessed: Jul. 09, 2023. [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.get_dummies.html
- [85] 'sklearn.preprocessing.MinMaxScaler', scikit-learn. Accessed: Jul. 10, 2023. [Online]. Available: <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
- [86] R. Khandelwal, 'T-distributed Stochastic Neighbor Embedding(t-SNE)', Medium. Accessed: Jul. 10, 2023. [Online]. Available: <https://towardsdatascience.com/t-distributed-stochastic-neighbor-embedding-t-sne-bb60ff109561>

- [87] 'A Taxonomy of Evaluation Methods for Information Systems Artifacts'. Accessed: Sep. 29, 2023. [Online]. Available: <https://www.tandfonline.com/doi/epdf/10.1080/07421222.2015.1099390?needAccess=true>
- [88] A. Sampaio, E. Gray, H. Moreira, and M. Martins, 'Selection of Projects in Software Process Assessment: New Perspectives', in *New Perspectives on Information Systems Development: Theory, Methods, and Practice*, G. Harindranath, W. G. Wojtkowski, J. Zupančič, D. Rosenberg, W. Wojtkowski, S. Wrycza, and J. A. A. Sillince, Eds., Boston, MA: Springer US, 2002, pp. 135–146. doi: 10.1007/978-1-4615-0595-2_11.
- [89] J. Nielsen and T. K. Landauer, 'A mathematical model of the finding of usability problems', in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*, Amsterdam, The Netherlands: ACM Press, 1993, pp. 206–213. doi: 10.1145/169059.169166.

Annex A

AIM Health Secure Platform



Annex B



FORMULÁRIO DE SUBMISSÃO PARA APROVAÇÃO ÉTICA

AIMHealth

Ana Vieira, ISTAR-Iscte, mestranda
(Prof. Miguel Sales Dias e Luís Elvas, ISTAR-Iscte, equipa de orientação)

Índice

IDENTIFICAÇÃO DO ESTUDO E INVESTIGADOR(ES).....	3
DESCRIÇÃO DO ESTUDO	5
PROBLEMA DE INVESTIGAÇÃO E RELEVÂNCIA DO ESTUDO	5
OBJETIVOS/ PERGUNTAS DE INVESTIGAÇÃO	5
MÉTODO.....	5
PARTICIPANTES	7
NÚMERO, IDADE E ORIGEM DOS/AS PARTICIPANTES	7
MÉTODO DE RECRUTAMENTO	8
CONSENTIMENTO INFORMADO E <i>DEBRIEFING</i>	8
OBTENÇÃO DO CONSENTIMENTO INFORMADO	8
ELEMENTOS DO CONSENTIMENTO INFORMADO	9
PRESTAÇÃO DO <i>DEBRIEFING</i>	11
ELEMENTOS DO <i>DEBRIEFING</i>	12
PROTEÇÃO E SEGURANÇA DOS PARTICIPANTES.....	13
AMOSTRA PROVENIENTE DE POPULAÇÕES VULNERÁVEIS	13
RISCOS ASSOCIADOS À PARTICIPAÇÃO.....	13
DECLARAÇÃO DE RESPONSABILIDADE E DE CONDUTA ÉTICA.....	14
ANEXOS	15
Anexo 1 - QUESTIONÁRIO SOBRE TRATAMENTO DE DADOS PESSOAIS	16
Anexo 2 - EXPLICAÇÃO DA EXPERIÊNCIA	23
Anexo 3- CONSENTIMENTO INFORMADO	25
Anexo 4 – QUESTIONÁRIO PRELIMINAR	27
Anexo 5 – GUIÃO DE TAREFAS.....	28
Anexo 6 – QUESTIONÁRIO SEQ.....	29
Anexo 7 – QUESTIONÁRIO SUS.....	31
Anexo 8- QUESTIONÁRIO GUI.....	32
Anexo 9 - <i>DEBRIEFING</i> /EXPLICAÇÃO DA INVESTIGAÇÃO	33
Anexo 10 - GRELHA DE REGISTO DE REALIZAÇÃO DE TAREFAS.....	34
Anexo 11 – PROTÓTIPO EM PAPEL	37

IDENTIFICAÇÃO DO ESTUDO E INVESTIGADOR(ES)

Título do projeto:	FCT DSAIPA/AI/0122/2020 AIMHealth – Aplicações Móveis Baseadas em Inteligência Artificial para Resposta de Saúde Pública
Investigador/a proponente:	Ana Vieira
Investigador/a responsável:	Ana Vieira (Prof. Miguel Sales Dias e Luís Elvas, equipa de orientação)
Contactos (e-mail):	ana_beatriz_vieira@iscte-iul.pt
Escola:	Escola de Tecnologias e Arquitetura
Departamento ou centro de investigação	ISTAR-Iscte
Equipa de investigação:	Ana Vieira, Prof. Miguel Sales Dias, Luís Elvas
Financiamento (se aplicável):	FCT- Fundação para a Ciência e a Tecnologia, através da bolsa DSAIPA/AI/0122/2020
Submissão:	Primeira submissão ☒ Re-submissão ☐ Alteração ☐

CHECKLIST PARA QUESTÕES DE ÉTICA

Indique se o estudo envolve algum dos seguintes elementos (assinale todos os que se aplicam):

Amostra proveniente de populações vulneráveis

- Crianças e jovens com menos de 18 anos. ☐
- Pessoas com dificuldades físicas ou psicológicas. ☐
- Pessoas com relação de dependência em relação aos/ às responsáveis pela investigação (e.g., superiores hierárquicos; assimetria de poder/ estatuto) ou no contexto onde decorre a investigação (e.g., universidade; empresas). ☐

Pessoas pertencentes a grupos minoritários em situação de vulnerabilidade e/ou em situação ilegal. ☐

Riscos significativos para os/as participantes

Recolha de informação sobre assuntos sensíveis para os/as participantes (e.g., experiências traumáticas; limitações físicas; sofrimento psicológico). ☐

Indução de estados de desconforto físico (e.g., tarefas físicas prolongadas ou muito repetitivas) ou psicológico (e.g., ansiedade; humilhação). ☐

Atribuição de rótulos ou categorias com consequências potencialmente negativas para o autoconceito (e.g., manipulação de competências percebidas; manipulação de situações de exclusão). ☐

Atividades invasivas (e.g., administração de substâncias; ingestão de alimentos). ☐

Recolha de tecidos humanos, sangue ou outros materiais biológicos. ☐

Indique se o estudo envolve tratamento de dados pessoais:¹

Sim ☒
Não² ☐

SE INDICOU QUE O ESTUDO ENVOLVE O TRATAMENTO DE DADOS PESSOAIS, INCLUA EM ANEXO O QUESTIONÁRIO SOBRE TRATAMENTO DE DADOS PESSOAIS

¹ Por dados pessoais entende-se qualquer informação, de qualquer natureza e em qualquer suporte (e.g. registo de voz ou imagem), relativa a uma pessoa singular identificada ou identificável. É considerada identificável uma pessoa singular que possa ser identificada direta ou indiretamente, por exemplo, através de um nome, um número de identificação, dados de localização, identificador eletrónico (e.g. IP) ou de outros elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular.

² Se pretende que o seu estudo não envolva o tratamento de dados pessoais, isto é, recolha e trate exclusivamente dados anónimos, o momento do processo de anonimização é fundamental.

O facto de um estudo não reportar respostas individuais de participantes, não é por si só indicador de não haver tratamento de dados pessoais. Pode considerar-se que um estudo nunca trata dados pessoais apenas na condição de que o investigador não tenha acesso a nenhum suporte com registos de dados pessoais durante a recolha e posteriores tratamentos.

Se a anonimização dos dados ocorrer numa fase posterior à recolha de dados, por exemplo, quando o investigador remove informações de identificação pessoal de uma transcrição áudio de uma entrevista, ou quando transfere os dados pessoais recolhidos para outra base de dados, os dados brutos ainda são pessoais e deve indicar “sim” nesta questão, anexando ainda o questionário sobre tratamento de dados pessoais.

DESCRIÇÃO DO ESTUDO

PROBLEMA DE INVESTIGAÇÃO E RELEVÂNCIA DO ESTUDO

Indique o problema de investigação e a relevância do estudo, clarificando qual o contributo original que apresenta para o avanço do conhecimento e/ ou outros benefícios esperados para indivíduos ou comunidades. [até 200 palavras]

A 11 de março de 2020, a Organização Mundial de Saúde declarou a Covid-19 como pandemia. Durante este surto, em que o distanciamento social e a quarentena se tornaram a norma, um rápido diagnóstico e prognóstico e a existência de telemonitorização tornaram-se de elevada importância.

Neste contexto, as aplicações de saúde foram reconhecidas como uma solução para o diagnóstico e vigilância de pacientes com Covid-19. Estas aplicações visam promover a saúde dos utilizadores, através da monitorização do seu bem-estar e prevenção ou gestão de doenças.

A aplicação que estamos a desenvolver, AIMHealth, permite registar diversos parâmetros de saúde, com o objetivo de acompanhar doentes de grupos de risco, bem como estabelecer diagnósticos, prognósticos e/ou acompanhar pacientes COVID-19 em ambulatório.

Este estudo irá contribuir para o aumento do nosso conhecimento necessário ao desenvolvimento da nossa aplicação de saúde adaptada aos utilizadores e às suas necessidades. O presente estudo tem como objetivo avaliar a forma como os potenciais utilizadores interagem com o protótipo desta aplicação, cumprindo com o princípio da iteração (*feedback* contínuo dos utilizadores com base em prototipagem) no desenvolvimento centrado no utilizador. Serão feitos testes formativos para identificar problemas e potenciais melhoramentos, assim como uma avaliação sumativa.

OBJETIVOS/ PERGUNTAS DE INVESTIGAÇÃO

Indique os objetivos gerais e específicos do estudo e/ ou a(s) pergunta(s) de investigação [até 150 palavras]

O presente estudo visa avaliar uma aplicação de *smartphone*. Numa primeira fase, essa avaliação será feita através de um protótipo de baixa fidelidade, em papel, tendo como referência as tarefas já definidas para o utilizador, de forma a identificar os problemas e as oportunidades de melhoramento da aplicação. Numa segunda fase, depois de identificados os problemas e oportunidades de melhoria e após a melhoria da aplicação, será feita a avaliação sumativa, de modo a ser possível avaliar a usabilidade da aplicação e a satisfação dos utilizadores.

Para avaliar a usabilidade serão utilizadas as seguintes métricas:

- **Eficácia:** número de tarefas realizadas pelo utilizador com sucesso;
- **Eficiência:** tempo que os utilizadores demoram a realizar as tarefas definidas;
- **Satisfação:** comportamento e reação dos utilizadores durante e depois da realização das tarefas propostas.

MÉTODO

Explique a escolha dos métodos de investigação e descreva todos os procedimentos para a recolha e registo de dados, participação e tarefas solicitadas aos/às participantes, intervenções realizadas, duração da participação e frequência da recolha de dados.

Se são tratados dados pessoais, inclua informação sobre:

- i) Os dados pessoais recolhidos, quem são os titulares dos dados e os tratamentos previstos;
- ii) O fundamento jurídico para o tratamento, caso não seja o consentimento;
- iii) Quem é o responsável pelo tratamento, nos casos em que não é o Iscte ou há responsabilidade conjunta;
- iv) Os procedimentos e os momentos ou os prazos previstos de pseudonimização e anonimização ou destruição, consoante os casos aplicáveis. Se existe anonimização, indique as medidas adotadas para reduzir o risco de re-identificação.

[até 700 palavras]

No contexto deste estudo, a recolha de informação será realizada através da experimentação de um protótipo por parte dos participantes, realizando tarefas predefinidas e respondendo a questionários para a avaliação do mesmo. Adicionalmente, o investigador anota comportamentos e comentários dos participantes, e regista a eficácia, a satisfação e a eficiência da realização das tarefas.

O tempo previsto de duração do estudo é de cerca de 30 minutos, e a recolha de dados será realizada duas vezes: testes formativos (fase preliminar) e avaliação sumativa.

Este estudo irá ter em consideração a seguinte metodologia:

- Definição do protótipo em papel (protótipo de baixa fidelidade);
- Definição das tarefas a realizar;
- Definição das medidas a avaliar;
- Definição dos questionários (questionário de caracterização preliminar, questionário de perguntas fechadas SEQ, SUS e GUI, e questionário de perguntas abertas);
- Contacto com os potenciais utilizadores;
- Realização dos testes;
- Análise e sistematização dos resultados e definição de conclusões.

O procedimento para a realização das entrevistas é o seguinte:

- O investigador agradece a disponibilidade e introduz brevemente a experiência (Anexo 2 - EXPLICAÇÃO DA EXPERIÊNCIA)
- O participante lê e assina o termo de consentimento informado. Caso o participante não aceite participar não se avança com a experiência (Anexo 3- CONSENTIMENTO INFORMADO)
- O participante preenche o questionário preliminar (Anexo 4 – QUESTIONÁRIO PRELIMINAR)
- O investigador explica como vai decorrer a experiência (Anexo 2 - EXPLICAÇÃO DA EXPERIÊNCIA)
- O investigador fornece o guião de tarefas ao participante (Anexo 5 – GUIÃO DE TAREFAS)
- O participante realiza as tarefas. Durante a realização das tarefas, o computador humano simula a resposta da interface do protótipo em papel às interações do participante, e o investigador tira notas sobre o que está a acontecer e regista/anota o sucesso/insucesso da realização das tarefas, as suas quebras, as dúvidas e hesitações do participante e mede o tempo de execução de cada tarefa. O participante não pode ser ajudado na realização da tarefa. Caso não consiga realizá-la é registada como não cumprida e o investigador pode intervir, apenas para se poder passar à próxima tarefa.
- Após a realização de cada tarefa o participante responde a um questionário SEQ (Single Ease Question) (Anexo 6 – QUESTIONÁRIO SEQ)
- No final, é fornecido um questionário SUS (System Usability Scale) ao participante (

- Anexo 7 – QUESTIONÁRIO SUS)
- Depois do questionário SUS, é fornecido um questionário da usabilidade da interface gráfica GUI (Graphical User Interface) ao participante (Anexo 8- QUESTIONÁRIO GUI)
- O investigador faz duas perguntas (de resposta aberta) ao participante (Anexo 2 - EXPLICAÇÃO DA EXPERIÊNCIA)
- Debriefing: O investigador explica o que deveria ter sido feito no caso das tarefas que não tenham sido concluídas. Agradece novamente a participação e reforça os dados de contato para caso o participante deseje colocar uma dúvida ou partilhar algum comentário, e manifestar a intenção de informar sobre os resultados e conclusões do estudo (Anexo 9 - DEBRIEFING/EXPLICAÇÃO DA INVESTIGAÇÃO)

Para além do referido anteriormente, irá ser feito um registo fotográfico da experiência, sem foco nos participantes, mas sim na atividade (exemplo: registo das mãos dos participantes a interagir com o protótipo em papel). Ainda assim, esse registo será autorizado pelo participante no consentimento informado. Caso o participante mostre desconforto em ser fotografado, não será feito o registo fotográfico.

Sobre dados pessoais:

Os dados pessoais recolhidos são o nome do entrevistado, contacto e registo fotográfico.

O fundamento jurídico para o tratamento é o consentimento informado do titular dos dados, de acordo com o art.6, nº1, alínea a) do Regulamento Geral de Proteção de Dados.

De acordo com o documento da Comissão de ética de Orientações aos investigadores sobre proteção de dados em atividades de investigação científica no Iscte (D1), além do Iscte, o responsável pelo tratamento dos dados é a estudante Ana Vieira e o doutorando Luís Elvas, sob orientação do Professor Miguel Sales Dias.

Os dados recolhidos são pseudonimizados, sendo os questionários respondidos por cada participante e dados anotados pelo investigador relativamente a cada participante identificados por código (ex: H1, H2, H3 (...)). Estes dados, alvo de tratamento estatístico, serão guardados em Excel.

O nome e contato dos participantes são armazenados à parte, sendo que o documento com os dados que serão alvo de tratamento estatístico não tem qualquer informação pessoal dos participantes.

O registo fotográfico da experiência irá focar na utilização do protótipo e não nos participantes, pelo que as fotografias serão tiradas de forma que os participantes não sejam identificáveis.

INCLUA EM ANEXO OS MATERIAIS A UTILIZAR NO ÂMBITO DA RECOLHA DE DADOS

(Ao enviar a submissão, por favor anexe os questionários, guiões de entrevista ou de atividade, grelhas de registo/observação, etc., devidamente identificados)

PARTICIPANTES

NÚMERO, IDADE E ORIGEM DOS/AS PARTICIPANTES

Caracterize os / as participantes no estudo, no que respeita ao número esperado, critérios de seleção, intervalo de idades e origem (e.g., contexto de recrutamento). [até 100 palavras]

No total, estima-se que aproximadamente 20 pessoas participem no estudo: para o protótipo de baixa fidelidade serão realizados testes com 5 participantes; na avaliação sumativa prevê-se a realização de testes com pelo menos 20 participantes.
O estudo tem como população-alvo a população geral. Relativamente aos critérios de inclusão para participação no estudo, os participantes têm de ter idade igual ou superior a 18 anos e um domínio básico da língua portuguesa escrita.

MÉTODO DE RECRUTAMENTO

Descreva o método de recrutamento dos/as participantes. [até 100 palavras]

Os participantes serão recrutados por contato telefónico ou pessoalmente.
O agendamento da entrevista será feito de acordo com a preferência e disponibilidade do participante. No caso da ocorrência de imprevistos, a entrevista será remarcada de acordo com a disponibilidade de ambas as partes.

CONSENTIMENTO INFORMADO E DEBRIEFING

OBTENÇÃO DO CONSENTIMENTO INFORMADO

Indique o momento e o local de obtenção do consentimento informado, bem como medidas para superar barreiras linguísticas (caso existam). [até 100 palavras]

O consentimento informado é obtido presencialmente no início do estudo, após uma pequena introdução e antes de se dar início à atividade.
Não existem barreiras linguísticas, uma vez que o estudo será realizado em português, com participantes portugueses. A linguagem utilizada será clara e sem utilização de termos complexos, de forma que qualquer pessoa consiga entender.

Indique o meio de obtenção do consentimento informado:

Documento no qual o/ a participante assina o seu consentimento (e.g., estudo com participação presencial) ☒

Documento/texto que o/a participante lê antes de transmitir a sua intenção de participar (e.g., estudo online, via videoconferência, etc.) ☐

Se o consentimento não é obtido presencialmente e há recolha de dados pessoais, explique por favor como regista a manifestação positiva e explícita do participante de forma a poder comprová-la.

[Click here to enter text](#)

Explicação oral dada ao/à participante antes de transmitir a sua intenção de participar (e.g., quando a identificação pessoal possa implicar riscos para o/a participante) ☐

Consentimento obtido através de terceiros que assegurem os direitos dos/as participantes, tais como os/as cuidadores/as principais ou representantes legais ☐

Se através de terceiros, por favor descreva quem irá consentir e como o consentimento será obtido [até 50 palavras]

[Click here to enter text](#)

Outro meio ou Não Aplicável ☐

Se através de outro meio ou Não Aplicável, por favor descreva/justifique [até 50 palavras]

ELEMENTOS DO CONSENTIMENTO INFORMADO

Assinale os elementos que incluiu no consentimento informado:

Identificação do estudo e do/s investigador/es responsável/eis ☒

Descrição dos objetivos gerais do estudo, número sessões, tempo estimado e características gerais da participação ☒

- Natureza voluntária da colaboração que inclui a possibilidade de interromper a participação em qualquer momento sem necessidade de justificação ☒
- Informação sobre eventuais riscos, desconfortos ou outros efeitos adversos associados à participação ☐
- Informação sobre eventuais benefícios associados ao estudo e/ou à participação ☐
- Informação sobre eventuais limites à confidencialidade, quando aplicável ☐
- Informação sobre incentivos à participação, quando aplicável ☐
- Dados de contacto caso o/a participante deseje fazer perguntas ou comentários sobre o estudo ☒
- Medidas previstas para lidar com eventuais consequências negativas para os/as participantes, quando aplicável ☐

Caso o estudo envolva tratamento de dados pessoais, assinale ainda os elementos que incluiu no consentimento informado:

- Identificação do Iscte-Instituto Universitário de Lisboa como Responsável pelo Tratamento, e/ou de outros Responsáveis pelo Tratamento se aplicável. ☒
- O fundamento legal para o tratamento (art. 6º, nº1, a) ou art 9º, nº2, a) do RGPD)³ ☒
- Os direitos que o participante titular de dados pode exercer junto do Responsável pelo Tratamento e a forma e contactos a quem se dirigir (direitos de acesso, retificação, de retirar o consentimento, e de apagamento) ☒
- O direito de apresentar reclamação à CNPD ☒
- O prazo de conservação dos dados pessoais (após o qual são destruídos ou anonimizados) ☒
- Caso haja tratamento de dados pessoais por terceiros (e.g. subcontratantes) ou uso posterior dos dados pessoais por outras equipas de investigação, informação sobre os fins do tratamento e a identificação dos terceiros ☐

³ O art 9º, nº2, a) aplica-se ao dados pessoais que revelem a origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas, ou a filiação sindical, bem como o tratamento de dados genéticos, dados biométricos para identificar uma pessoa de forma inequívoca, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa; nos outros casos aplica-se o art. 6º, nº1, a).

Se for caso disso, o facto de haver transferências para um país terceiro ou organização internacional fora do Espaço Económico Europeu e a existência ou não de uma “decisão de adequação”. Caso não exista “decisão de adequação”, informações sobre as transferências, dos riscos que daí podem decorrer para os participantes e de medidas adotadas para mitigar o risco. ☐

Os contactos do Encarregado de Proteção de Dados ☒

Caso o tratamento de dados envolva potenciais riscos para os direitos e liberdades dos participantes, a importância e as consequências previstas de tal tratamento para os participantes ☐

Caso existam dados pessoais não recolhidos junto dos participantes, a origem dos dados pessoais e, eventualmente, se provêm de fontes acessíveis ao público ☒

Caso existam decisões automatizadas⁴, incluindo a definição de perfis referida no art. 22º, nº 1 e nº4 do RGPD, informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento para o participante ☐

Caso inclua no consentimento informado outros elementos, assinale-os:

Outros elementos ☒

Se incluiu outros elementos, por favor descreva [até 50 palavras]
Informação sobre registo fotográfico. Caso o participante não autorize o registo fotográfico, mas aceite participar no estudo, prossegue-se sem tirar fotografias.

PRESTAÇÃO DO DEBRIEFING

Indique o meio de prestação do *debriefing*:

Documento/ texto apresentado ao/à participante no final da participação ☐

Explicação oral dada ao/à participante no final da participação ☒

Outro meio ou Não Aplicável ☐

Se através de outro meio ou Não Aplicável, por favor descreva/justifique [até 50 palavras]:
Click here to enter text

⁴ Decisões individuais automatizadas ocorrem quando são tomadas decisões sobre uma pessoa singular por meios tecnológicos e sem envolvimento humano. Podem ser efetuadas mesmo sem definição de perfis. Por exemplo, se a decisão de um banco em conceder um empréstimo bancário a uma pessoa singular for exclusivamente tomada por um algoritmo, sem intervenção humana. Se uma pessoa controlar a decisão final fornecida pelo algoritmo, com efetiva competência ou possibilidade de influenciar o resultado final, a decisão pode ser considerada não “exclusivamente” automatizada.

ELEMENTOS DO DEBRIEFING

Assinale os elementos que incluiu no *debriefing*:

- Agradecimento pela participação ☒
- Informação mais específica sobre objetivos, hipóteses, procedimentos e/ou contributos esperados da investigação do estudo, quando aplicável ☐
- Clarificação sobre *deception* na investigação, quando aplicável ☐
- Dados de contacto caso o/a participante deseje fazer perguntas ou comentários sobre o estudo ☒
- Meios de obter informação posterior sobre os resultados e conclusões do estudo ☒
- Meios de obter informação sobre o tema de investigação, quando aplicável ☐
- Medidas previstas para lidar com eventuais consequências negativas para os/as participantes, quando aplicável ☐
- Outros elementos ☒

Se incluiu *outros elementos*, por favor descreva [até 50 palavras]:
 Explicação do que era necessário fazer para realizar a(s) tarefa(s) que o participante não tenha conseguido concluir com sucesso, quando aplicável.
 Reforço de que a aplicação é que estava a ser avaliada e não a capacidade do participante em realizar as tarefas.

Se desejar clarificar ou justificar algum aspeto relacionado com os elementos do consentimento informado e/ou do *debriefing*, por favor descreva. [até 100 palavras]

INCLUA EM ANEXO OS DOCUMENTOS DE CONSENTIMENTO INFORMADO E DEBRIEFING

(Ao enviar a submissão, por favor anexe os documentos/textos de consentimento informado e *debriefing* ou, no caso da explicação oral, a transcrição do discurso direto)

PROTEÇÃO E SEGURANÇA DOS PARTICIPANTES

AMOSTRA PROVENIENTE DE POPULAÇÕES VULNERÁVEIS

Se a amostra é constituída por:

- Crianças e jovens com menos de 18 anos;
- Pessoas com dificuldades físicas ou psicológicas;
- Pessoas com relação de desigualdade ou dependência em relação aos/às responsáveis da investigação, ou no contexto onde decorre a investigação;
- Ou outras populações que possam ser consideradas vulneráveis (e.g., grupos minoritários em situação de vulnerabilidade e/ou em situação ilegal).

Indique as medidas previstas para assegurar que a participação é estritamente voluntária (e.g., no caso de estudantes universitários/as em que a participação integre um componente curricular, devem ser dadas alternativas à participação para obtenção de créditos). [até 100 palavras]

A amostra não é proveniente de populações vulneráveis.

RISCOS ASSOCIADOS À PARTICIPAÇÃO

Se existirem riscos potencialmente significativos para os/as participantes, tais como:

- Recolha de informação sobre assuntos sensíveis (e.g., experiências traumáticas; limitações físicas; sofrimento psicológico);
- Indução de estados de desconforto físico (e.g., tarefas físicas prolongadas ou muito repetitivas) ou psicológico (e.g., ansiedade; humilhação);
- Atribuição de rótulos ou categorias no contexto experimental com consequências potencialmente negativas para o autoconceito (e.g., manipulação de competências percebidas; manipulação de situações de exclusão);
- Atividades invasivas (e.g., administração de substâncias);
- Recolha de tecidos humanos, sangue ou outros materiais biológicos;
- Ou outras atividades que se antecipe que possam implicar riscos significativos para os/as participantes.

Indique os procedimentos previstos para minimizar os riscos e/ou monitorizar a segurança dos participantes. [até 100 palavras]

Não existem riscos associados à participação.

Indique as medidas previstas para lidar com eventuais consequências negativas para os/as participantes. [até 100 palavras]

Não existem consequências negativas para os participantes.

DECLARAÇÃO DE RESPONSABILIDADE E DE CONDUTA ÉTICA

Enquanto investigador/a responsável pelo estudo declaro que:

- Toda a informação prestada nesta submissão é verdadeira; ☒
- Procurei antecipar todos os riscos que possam surgir associados à participação no estudo, delinear estratégias para minimizar os riscos, e definir medidas para lidar com eventuais consequências negativas para os participantes; ☒
- Detenho (individualmente ou em equipa) as competências e os recursos necessários para concretizar o projeto tal como foi apresentado nesta submissão; ☒
- A minha conduta e as minhas decisões em todas as matérias relacionadas com o presente projeto terão em consideração as disposições do Código de Conduta Ética na Investigação – ISCTE-IUL. ☒

Nome: Ana Beatriz da Cruz Baptista Galamba Vieira

Data: 10 de janeiro de 2023

Assinatura:

Assinado por: **Ana Beatriz da Cruz Baptista Galamba Vieira**
Num. de identificação: 14704024
Data: 2023.01.10 15:44:18+00'00'



ANEXOS

Anexo 1 - QUESTIONÁRIO SOBRE TRATAMENTO DE DADOS PESSOAIS

QUADRO 1 - TRATAMENTO DE DADOS PESSOAIS		
A.	O estudo envolve tratamento de dados pessoais? ⁱ	
A.1	Sim	☒
A.2	Não	☐
Se o estudo <u>não</u> envolve o tratamento de dados pessoais, deixe em branco as demais respostas do quadro 1 e do quadro 2		
B.	Além do Iscte, existem outros responsáveis pelo tratamento dos dados pessoais? ⁱⁱ	
B.1	Sim Identifique as instituições ou as pessoas singulares responsáveis: Ana Vieira, estudante de Mestrado do Iscte, Luís Elvas, estudante de doutoramento do Iscte, ISTAR-Iscte, Prof. Miguel Sales Dias, Professor Associado com Agregação do Iscte, ISTAR-Iscte (Deve consultar o GAI para a realização de um acordo deresponsabilidade conjunta, de acordo com o art. 26º do RGPD).	☒
B.2	Não	☐
C.	Qual a base legal para o tratamento de dados pessoais?	
C.1	Consentimento dos titulares de dados – art. 6º, nº1, a) do RGPD	☒
C.2	Consentimento dos titulares de dados (para o tratamento de categorias especiais de dados - ‘dados sensíveis’) ⁱⁱⁱ – art. 9º, nº2, a) do RGPD	☐
C.3	Outro: Click here to enter text.	☐

ⁱ Por dados pessoais entende-se qualquer informação, de qualquer natureza e em qualquer suporte (e.g. registo de voz ou imagem), relativa a uma pessoa singular identificada ou identificável. É considerada identificável uma pessoa singular que possa ser identificada direta ou indiretamente, por exemplo, através de um nome, um número de identificação, dados de localização, identificador eletrónicos (e.g. IP) ou de outros elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular.

ⁱⁱ O Responsável pelo Tratamento é a pessoa singular ou coletiva, a autoridade pública, a agência ou outro organismo que, individualmente ou em conjunto com outras, determina as finalidades e os meios de tratamento de dados pessoais. Quando dois ou mais responsáveis pelo tratamento determinem conjuntamente as finalidades e os meios desse tratamento, ambos são responsáveis conjuntos pelo tratamento. Os responsáveis conjuntos pelo tratamento determinam, por acordo entre si e de modo transparente, as respetivas responsabilidades pelo cumprimento do RGPD (art. 4º, nº7 e arts 24º e 26º do RGPD).

ⁱⁱⁱ Aplica-se ao tratamento de categorias especiais de dados, i.e. dados pessoais que revelem a origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas, ou a filiação sindical, bem como o tratamento de dados genéticos, dados biométricos para identificar uma pessoa de forma inequívoca, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa.

D.	Como são recolhidos os dados pessoais?
----	--

D.1	Junto do titular de dados	☒
D.2	Dados pessoais de outros estudos, que têm como responsável pelo tratamento o Iscte	☐
D.3	Dados pessoais de outros estudos, que têm como responsável pelo tratamento outras instituições	☐
D.4	Dados disponíveis publicamente	☐
D.5	Outros (especifique)	☐
E. Qual a natureza de dados pessoais tratados?		
E.1	Categorias especiais de dados (dados 'sensíveis')^{iv} Indique as categorias: Click here to enter text.	☐
E.2	Dados de natureza altamente pessoal^v Indique quais os dados tratados: Click here to enter text.	☐
E.3	Dados pessoais relacionados com condenações penais e infrações Indique quais os dados tratados: Click here to enter text.	☐
E.4	Registos de voz, imagem ou vídeo	☒
E.5	Outros Indique quais: ^{vi} Nome e contato	☒

^{iv} Categorias especiais de dados, ou "dados sensíveis", são dados pessoais que revelem a origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas, ou a filiação sindical, os dados genéticos, dados biométricos que identifiquem uma pessoa de forma inequívoca, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa.

^v Tratam-se de dados pessoais ligados a atividades privadas ou familiares (tais como comunicações eletrónicas cuja confidencialidade deve ser protegida) ou porque afetam o exercício de um direito fundamental (tais como dados de localização cuja recolha põe em causa a liberdade de circulação) ou porque a sua violação implica claramente que a vida quotidiana do titular dos dados será gravemente afetada (tais como dados financeiros que possam ser utilizados para a prática de um crime de fraude).

^{vi} Exemplos: nome, números de identificação, contactos, morada, dados de localização, estado civil, dados financeiros, registo de imagens, voz ou vídeo, dados sociodemográficos, etc

F.	Quem são os titulares dos dados pessoais?	
F.1	Crianças ou jovens com idade inferior a 18 anos Número previsto de titulares de dados: Click here to enter text.	☐
F.2	Grupos vulneráveis, na aceção em que há acentuado desequilíbrio com o responsável pelo tratamento dos dados pessoais, significando isto que os indivíduos podem não ser capazes de consentir, ou opor-se, facilmente ao tratamento dos seus dados ou de exercer os seus direitos. ^{vii} Número previsto de titulares de dados: Click here to enter text.	☐
F.3	Estudantes do Iscte Número previsto de titulares de dados: Click here to enter text.	☐
F.4	Colaboradores do Iscte (e.g. docentes, funcionários, etc.) Número previsto de titulares de dados: Click here to enter text.	☐
F.5	Outros Indique número previsto de titulares e quem são: Está prevista a realização de entrevistas a cerca de 5 participantes.	☒
G.	Quantas pessoas na equipa do estudo se prevê terem acesso aos dados pessoais?	3
H.	Medidas técnicas e organizativas para a proteção dos dados pessoais e prazos de conservação:	
H.1	Anonimização Indique o prazo de conservação dos dados pessoais até à anonimização: Click here to enter text.	☐
H.2	Pseudonimização ^{viii}	☒
H.3	Destruição Indique o prazo de conservação:	☐

^{vii} É o caso, por exemplo, de crianças, empregados, segmentos vulneráveis da população que necessitem de proteção especial, e.g. pessoas com doenças mentais, requerentes de asilo, idosos, doentes, etc

^{viii} Tratamento de dados pessoais de forma que deixem de poder ser atribuídos a um titular de dados específico sem recorrer a informações suplementares, desde que essas informações suplementares sejam mantidas separadamente e sujeitas a medidas técnicas e organizativas para assegurar que os dados pessoais não possam ser atribuídos a uma pessoa singular identificada ou identificável. Por exemplo, criando uma cópia do conjunto de dados, mas onde as informações de identificação pessoal (e.g. o nome de um indivíduo) foram substituídas por identificadores codificados, e prosseguindo o tratamento no novo conjunto de dados que, por si só e sem a chave de descodificação, não permite a identificação dos titulares de dados

I.	Assinale a opção aplicável para o software utilizado para o tratamento de dados pessoais:	
I.1	Software licenciado pelo Iscte (e.g. Excel do Office 365)	☒
I.2	Software não licenciado pelo Iscte, e cujo funcionamento está em conformidade com o RGPD Especifique qual o software utilizado: Click here to enter text.	☐
J.	Indique a forma e local de armazenamento dos dados pessoais:	
J.1	Em servidores do Iscte ou através de serviços de nuvem fornecidos pelo Iscte (e.g. Sharepoint do Office 365)	☐
J.2	Em computadores institucionais do Iscte	☐
J.3	Em computadores do próprio	☒
J.4	Outro (especifique): Click here to enter text.	☐
K.	Outras medidas técnicas e organizativas – assinale quais as medidas adotadas:	
K.1	As sessões de acesso aos dados pessoais são protegidas e autenticadas com as credenciais pessoais.	☒
K.2	Encriptação dos dados pessoais nos dispositivos de armazenamento	☐
K.3	Membros da equipa de estudo que têm acesso aos dados e não têm contrato com o Iscte (e.g. alunos) assinam termo de responsabilidade e confidencialidade	☐
K.4	Descreva medidas adicionais de mitigação de risco que considere aplicar, em especial se assinalou positivamente algum dos itens E.1, E.2 ou E.3: Click here to enter text.	
L.	Existem subcontratantes, i.e. algum organismo ou terceiro que não faça parte da equipa do estudo e que tratará os dados pessoais ou parte dos dados pessoais por conta do responsável pelo tratamento?	
L.1	Sim Indique os dados tratados por conta de outrem e nome do organismo: Click here to enter text. (Deve consultar o GAI para a realização de um contrato que garanta os requisitos do art.28º do RGPD.)	☐
L.2	Não	☒

M.	Há transferências de dados pessoais para organizações em países fora do Espaço Económico Europeu e sem ‘decisão de adequação’? ^{ix}	
M.1	Sim Indique os países. Se a base legal para o tratamento é o consentimento do titular de dados, indique os riscos que podem decorrer das transferências para os participantes e medidas de mitigação. Caso a base legal seja outra que não o consentimento, descreva as garantias adequadas de acordo com o art.46º do RGPD. Click here to enter text. Pode juntar documentação relevante.	☐
M.2	Não	☒
N.	Se assinalou uma ou mais das opções D.2, D.3, D.4 ou D.5 e a base legal para o tratamento não é o consentimento dos titulares dos dados, os participantes foram ou serão informados sobre o novo tratamento de dados, nos termos do art. 14º do RGPD?	
N.1	Sim	☐
N.2	Não Justifique detalhadamente, i) porque não se mostra possível facultar essa informação, ou ii) porque implicaria um esforço desproporcional ou comprometeria os objetivos do tratamento, e iii) quais as medidas para proteger os direitos, liberdades e interesses dos titulares de dados. Click here to enter text.	☐
O.	Enquanto os dados pessoais não forem anonimizados ou destruídos, prevê-se que o estudo restrinja de alguma forma os direitos de acesso, retificação, limitação do tratamento ou de oposição previstos nos artigos 15.º, 16.º, 18.º e 21.º do RGPD?	
O.1	Sim Indique quais os direitos que ficam prejudicados e como e em que medida ficam prejudicados. Explique ainda i) porque são suscetíveis de tornar impossível ou prejudicar gravemente a realização dos objetivos do estudo; ii) como avalia o impacto que pode vir a ter nos titulares de dados. Click here to enter text.	☐
O.2	Não	☒

^{ix} Em certos casos, pode declarar-se que um país terceiro oferece um nível adequado de proteção através de uma decisão da Comissão Europeia («decisão de adequação»), o que significa que é possível transferir dados para uma instituição situada no país terceiro sem que o exportador dos dados tenha de apresentar garantias suplementares e sem que esteja sujeito a condições adicionais. Por outras palavras, as transferências para um país terceiro «adequado» serão semelhantes a uma transmissão de dados no interior da UE. Consultar https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-rules-apply-if-my-organisation-transfers-data-outside-eu_pt. A lista de decisões de adequação pode ser consultada em: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_pt

QUADRO 2 - TRATAMENTOS ESPECÍFICOS DE DADOS PESSOAIS SUSCETÍVEIS DE ELEVAR O RISCO

P.	Indique se o tratamento de dados pessoais envolve alguma das seguintes operações de risco: ^x	
P.1	Avaliação ou classificação, incluindo definição de perfis e previsão ^{xi}	☐
P.2	Decisões automatizadas que produzam efeitos jurídicos ou afetem significativamente o titular de dados de modo similar ^{xii}	☐
P.3	Controlo sistemático ^{xiii}	☐
P.4	Tratamento de dados em grande escala ^{xiv}	☐
P.5	Estabelecer correspondências ou combinar conjuntos de dados ^{xv}	☐
P.6	Utilização de soluções inovadoras ou aplicação de novas soluções tecnológicas ou organizacionais ^{vi}	☐
P.7	Tratamento que impede ou prejudica os titulares dos dados de exercer um direito ou de utilizar um serviço ou um contrato ^{xvii}	☐
P.8	Caso tenha assinalado alguma das operações P.1-P.7, descreva-as no contexto do trabalho de investigação e porque são necessárias. Não existem tratamentos específicos de dados pessoais suscetíveis de elevar o risco.	

Se assinalou positivamente dois ou mais pontos relativos ao seguinte conjunto de critérios, pode ponderar e juntar – ou pode vir a ser-lhe sugerido realizar pela Comissão de Ética ou o Encarregado de Proteção de Dados – uma Avaliação de Impacto sobre a Proteção de Dados:^{xviii}

Um dos itens E.1, E.2 – 1 ponto
E.3 – 1 ponto
Um dos itens F.1, F.2 – 1 ponto
P.1, P.2, P.3, P.4, P.5, P.6, P.7 – Cada um destes critérios vale um ponto

^x Exemplos sobre o significado de cada critério estão disponíveis no documento do GT29 para a proteção de dados: 'Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679', em particular os critérios aí apresentados e os exemplos das páginas 10-14 (tomando como referência a versão portuguesa), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

^{xi} Em especial de «aspectos relacionados com o desempenho profissional, a situação económica, saúde, preferências ou interesses pessoais, fiabilidade ou comportamento, localização ou deslocações do titular dos dados» (considerandos 71 e 91 do RGPD). Os exemplos deste critério podem incluir: Uma plataforma de ensino à distância onde se recolhem, analisam e classificam dados das atividades e comportamentos dos alunos para o docente aplicar pedagogias diferenciadas e melhorar os objetivos de aprendizagem; Uma instituição financeira que faça um controlo seletivo dos seus clientes a partir de uma base de dados de referências de crédito bancário ou a partir de uma base de dados de combate ao branqueamento de capitais

e ao financiamento do terrorismo ou de combate à fraude; Uma empresa debioteclologia que ofereça testes genéticos diretamente aos seus clientes por forma a avaliar e prever riscos de doença ou para a saúde; ou Uma investigação que desenvolva perfis comportamentais ou de comercialização baseados na utilização ou navegação num sítio web.

^{xii} Tratamento destinado à tomada de decisões sobre os titulares dos dados e que produza «efeitos jurídicos relativamente à pessoa singular» ou que «a afetem significativamente de forma similar» (art. 35.º, n.º 3, alínea a) do RGPD). Por exemplo, o tratamento pode implicar a exclusão ou a discriminação de indivíduos. O tratamento que produza poucos ou nenhuns efeitos relativamente aos indivíduos não satisfaz estes critérios específicos

^{xiii} Tratamento utilizado para observar, monitorizar ou controlar os titulares dos dados, incluindo dados recolhidos através de redes, ou um «controlo sistemático de zonas acessíveis ao público» (art. 35.º, n.º 3, alínea c) do RGPD). Por exemplo, captação de vídeo para o tratamento e investigação dos trajetos usados pelas pessoas em circulação num edifício acessível ao público, e.g numa universidade.

^{xiv} Não existe uma definição do que constitui grande escala no RGPD. Os seguintes fatores podem ser considerados na apreciação:

- O número de titulares de dados afetados como número concreto ou em percentagem da população em causa, por exemplo uma percentagem alta dos alunos do Iscte;
- O volume de dados e/ou o alcance dos diferentes elementos de dados objeto de tratamento;
- A duração, ou permanência, da atividade de tratamento de dados;
- O âmbito geográfico da atividade de tratamento.

Contam-se como exemplos de tratamento de grande escala: i) O tratamento de dados de uma tecnologia para uso de indivíduos de uma população e que faz o rastreio de contactos, como o StayawayCovid; ii) O tratamento de dados de doentes no exercício normal das atividades de um hospital; iii) O tratamento de dados de viagem das pessoas que utilizam o sistema de transportes públicos de uma cidade; iv) O tratamento de dados de clientes no exercício normal das atividades de uma companhia de seguros ou de um banco.

Exemplos que **não** constituem tratamento de grande escala, incluem: i) O tratamento de dados de doentes pacientes por um médico; ii) o tratamento de dados pessoais relacionados com condenações penais e infrações por um advogado.

Ver a secção 3 do seguinte do documento do GT29: https://www.cnpd.pt/media/meplvdie/wp243rev01_pt.pdf

^{xv} Por exemplo, com origem em duas ou mais operações de tratamento de dados realizadas com diferentes finalidades e/ou por diferentes responsáveis pelo tratamento de dados de tal forma que excedam as expectativas razoáveis do titular dos dados. Por exemplo, o tratamento de dados pessoais do percurso curricular e de desempenho de alunos de uma instituição universitária e que, para esse efeito, recorre também a dados pessoais relativos aos mesmos alunos e disponíveis publicamente numa rede social.

^{xvi} Quando a utilização da tecnologia pode envolver novas formas de recolha e utilização de dados, possivelmente com elevado risco para os direitos e as liberdades dos indivíduos. Por exemplo, combinar a utilização da impressão digital e do reconhecimento facial para melhorar o controlo de acesso físico a um edifício. O uso de big data, técnicas de Inteligência Artificial ou aplicações de Internet das Coisas são candidatos a este critério.

^{xvii} Por exemplo, operações de tratamento destinadas a autorizar, alterar ou recusar o acesso dos titulares dos dados a um serviço ou que estes celebrem um contrato. Por exemplo, quando um banco faz um controlo seletivo dos seus clientes a partir de uma base de dados de referências de crédito bancário com vista a decidir se lhes concede ou não um empréstimo.

^{xviii} Consultar o documento do GT20: 'Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «susceptível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679', https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

Anexo 2 - EXPLICAÇÃO DA EXPERIÊNCIA

Este guião deve ser verbalizado pelo investigador aos participantes e poderá ser lido ou memorizado e dito, desde que a mesma informação chegue às pessoas que irão efetuar o teste.

“Antes de mais quero agradecer a sua participação neste teste, a sua opinião é uma parte importante para o estudo que está a ser realizado. Irá participar na avaliação de uma aplicação de saúde que permite registar vários parâmetros de saúde, com o objetivo de acompanhar doentes de grupos de risco, assim como estabelecer diagnósticos e/ou acompanhar pacientes com COVID-19. Esta aplicação surgiu no contexto da pandemia com o distanciamento social e a necessidade de telemonitorização.

O objetivo deste teste é então perceber a sua opinião relativamente à aplicação que lhe vai ser apresentada, nomeadamente em relação à facilidade de uso.

Numa primeira fase ser-lhe-á entregue um termo de consentimento, que deve ser lido e assinado se concordar com o conteúdo, para garantir que o teste é feito voluntariamente. Caso esteja de acordo, e depois de responder a um pequeno questionário preliminar, prosseguimos para a próxima fase onde lhe será pedido que realize algumas tarefas para que possa experimentar o protótipo e dar a sua opinião.

Solicito-lhe agora que leia o Termo de Consentimento, onde tem acesso a informação mais detalhada sobre objetivos e condições do estudo, e, caso concorde com tudo, assine e preencha o questionário preliminar.”

[Entrega e assinatura do Termo de consentimento]

[Entrega do Questionário Preliminar]

Antes de serem fornecidas tarefas ao participante, o investigador deverá dar as seguintes instruções:

“Vou agora passar à explicação de como a experiência vai ocorrer:

Este é o protótipo que irá ser avaliado. Vai ter a oportunidade de experimentá-lo através da realização de seis tarefas simples, que estão descritas num guião que lhe vou disponibilizar no início da experiência. São sete ecrãs, que correspondem às diferentes funcionalidades da aplicação com as quais irá interagir: medições de dados fisiológicos, visualização dos dados registados, visualização e adição de lembretes de medicamentos, exames e consultas.

Enquanto estiver a realizá-las, vai ser acompanhado por uma simulação da resposta do sistema às suas ações. Peço-lhe que preste atenção a todos os elementos do “ecrã”, e às indicações que vão surgindo, porque eu não o posso ajudar na realização das tarefas, uma vez que o que está a ser avaliado é a capacidade que o sistema tem de transmitir como se deve proceder. Reforço que em nenhum momento está a ser avaliado(a), o único alvo de avaliação neste teste é a aplicação. Peço-lhe também que durante a realização das tarefas diga em voz alta o que está a pensar e quais as suas dificuldades e opiniões que surgem no momento.

Após a realização de cada tarefa deve responder a um pequeno questionário que lhe será igualmente disponibilizado. Deve ser respondido logo após a realização de cada tarefa, porque é quando a informação ainda está fresca na memória. Se não conseguir completar uma tarefa, não se preocupe, passaremos à próxima.

Quando terminar ser-lhe-á pedido que responda a um pequeno questionário final de satisfação em relação à aplicação em geral.

Caso tenha alguma dúvida, esteja à vontade para colocá-la. Quando estiver pronto, podemos iniciar a experiência”.

[Entrega do Guião de tarefas]

[Entrega do Questionário SEQ]

[Realização das tarefas e resposta ao Questionário SEQ por parte do participante]

[Entrega do Questionário SUS]

[Entrega do Questionário GUI]

“Diga por favor, os três aspetos que mais gostou”

[Anotar]

1-

2-

3-

“Diga por favor os três aspetos que menos gostou”

[Anotar]

1-

2-

3-

Anexo 3- CONSENTIMENTO INFORMADO

O presente estudo surge no âmbito do projeto AIM Health – Aplicações Móveis Baseadas em Inteligência Artificial para Resposta de Saúde Pública, a decorrer no Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura (ISTAR-Iscte), financiado pela FCT- Fundação para a Ciência e a Tecnologia, através da bolsa DSAIPA/AI/0122/2020.

Este estudo tem por objetivo avaliar a aplicação de saúde AIMHealth, permitindo que os potenciais utilizadores a experimentem em primeira pessoa. Pretende-se fazer uma avaliação numa fase preliminar através do uso de um protótipo em papel. Esta avaliação serve para identificar problemas e oportunidades de melhoramento da aplicação. Após esta iteração e o melhoramento da aplicação, será realizada a avaliação sumativa para avaliar a sua usabilidade e a satisfação dos potenciais utilizadores.

A sua participação no estudo, que será muito valorizada, irá contribuir para o avanço do conhecimento neste domínio da ciência. Consiste em realizar algumas tarefas que lhe serão indicadas para experimentar a usar a interface e responder a pequenos questionários que lhe serão disponibilizados. Esta experiência poderá durar cerca de 30 minutos.

Os seus dados pessoais recolhidos são o seu nome, contato e recolha de imagem através de fotografias para registo da experiência (caso assim o autorize).

Além do Iscte, a estudante Ana Vieira, estudante de Mestrado do Iscte e ISTAR-Iscte e o estudante de doutoramento do Iscte Luis Elvas (com supervisão do Prof. Miguel Sales Dias, Professor Associado com Agregação do Iscte, ISTAR-Iscte), é responsável pelo tratamento dos seus dados pessoais, recolhidos e tratados exclusivamente para as finalidades do estudo, tendo como base legal o seu consentimento de acordo com o art.6, nº1, alínea a) do Regulamento Geral de Proteção de Dados.

O estudo é realizado por Ana Vieira (ana_beatriz_vieira@iscte-iul.pt), que poderá contactar caso pretenda esclarecer uma dúvida, partilhar algum comentário ou exercer os seus direitos relativos ao tratamento dos seus dados pessoais. Poderá utilizar o contacto indicado para solicitar o acesso, a retificação, o apagamento ou a limitação do tratamento dos seus dados pessoais.

A participação neste estudo é confidencial. Os seus dados pessoais serão sempre tratados por pessoal autorizado vinculado ao dever de sigilo e confidencialidade. O Iscte garante a utilização das técnicas, medidas organizativas e de segurança adequadas para proteger as informações pessoais. É exigido a todos os investigadores que mantenham os dados pessoais confidenciais.

Além de confidencial, a participação no estudo é estritamente voluntária: pode escolher livremente participar ou não participar. Se tiver escolhido participar, pode interromper a participação e retirar o consentimento para o tratamento dos seus dados pessoais em qualquer momento, sem ter de prestar qualquer justificação. A retirada de consentimento

não afeta a legalidade dos tratamentos anteriormente efetuados com base no consentimento prestado.

Não existem riscos significativos expectáveis associados à participação no estudo.

O Iscte não divulga ou partilha com terceiros a informação relativa aos seus dados pessoais.

O Iscte tem um Encarregado de Proteção de Dados, contactável através do email dpo@iscte-iul.pt. Caso considere necessário tem ainda o direito de apresentar reclamação à autoridade de controlo competente – Comissão Nacional de Proteção de Dados.

Declaro ter compreendido os objetivos de quanto me foi proposto e explicado pelo/a investigador/a, ter-me sido dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter obtido resposta esclarecedora. **Aceito** participar no estudo e consinto que os meus dados pessoais sejam utilizados de acordo com a informações que me foram disponibilizadas.

Sim ☐ Não ☐

Nome: _____

Assinatura: _____

Anexo 4 – QUESTIONÁRIO PRELIMINAR

Número de participante:

1. Idade

- ☐ 25 – 30
☐ 31 – 35
☐ 36 – 40
☐ 41 – 45

2. Género

- ☐ Masculino
☐ Feminino
☐ Outro (Por favor, especifique) _____

3. Como avalia a sua relação com as tecnologias digitais?

- ☐ Muito má
☐ Insuficiente
☐ Razoável
☐ Boa
☐ Muito Boa

4. Alguma vez teve contato com aplicações de saúde?

- ☐ Sim
☐ Não

5. Neste momento sinto-me ... (assinalar as que se aplicam)

- ☐ Motivado(a)
☐ Apreensivo(a)
☐ Outro (Por favor especifique) _____

Obrigado pela sua participação!

Anexo 5 – GUIÃO DE TAREFAS

Tarefas a serem realizadas pelo participante:

- **Tarefa 1:** Imagine que entrou na aplicação e deseja saber a sua frequência cardíaca. Meça o seu valor de frequência cardíaca.
- **Tarefa 2:** Imagine agora que conseguiu medir com sucesso a sua frequência cardíaca. Prossiga agora para a visualização dos valores que acabou de registar para a frequência cardíaca.
- **Tarefa 3:** Imagine agora que vai tomar o medicamento “ibuprofeno” e antes de o fazer quer confirmar que o está a tomar na dose e no horário corretos. Veja nos lembretes os medicamentos que tem de tomar e assinala o medicamento que tomou.
- **Tarefa 4:** Imagine que tem uma consulta de otorrinolaringologia marcada e quer confirmar os detalhes sobre a consulta (tipo de consulta, horário, local, etc.). Veja nos lembretes que essa consulta se encontra marcada.
- **Tarefa 5:** Imagine que acabou de sair de uma consulta e agendou um ecocardiograma. Adicione esse exame aos seus lembretes.
- **Tarefa 6:** Imagine que deseja aceder às informações associadas à sua conta. Veja o seu perfil.

Anexo 6 – QUESTIONÁRIO SEQ

Número de participante:

Por favor, responda a cada pergunta após a realização de cada tarefa.

Tarefa 1

De uma forma geral, quão fácil ou difícil foi realizar esta tarefa?

Muito difícil

Muito fácil

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Tarefa 2

De uma forma geral, quão fácil ou difícil foi realizar esta tarefa?

Muito difícil

Muito fácil

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Tarefa 3

De uma forma geral, quão fácil ou difícil foi realizar esta tarefa?

Muito difícil

Muito fácil

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Tarefa 4

De uma forma geral, quão fácil ou difícil foi realizar esta tarefa?

Muito difícil

Muito fácil

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Tarefa 5

De uma forma geral, quão fácil ou difícil foi realizar esta tarefa?

Muito difícil

Muito fácil

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Tarefa 6

De uma forma geral, quão fácil ou difícil foi realizar esta tarefa?

Muito difícil

Muito fácil

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Obrigado pela sua participação!

Anexo 7 – QUESTIONÁRIO SUS

Número de participante:

Para cada uma das afirmações abaixo, selecione a caixa (apenas uma por linha) que melhor descreve a sua opinião relativamente à aplicação que acabou de experimentar:

	Discordo totalmente						Concordo totalmente
1- Acho que gostaria de utilizar esta aplicação com frequência.							
	1	2	3	4	5		
2- Considerei esta aplicação mais complexa do que necessária.							
	1	2	3	4	5		
3- Achei a aplicação fácil de usar.							
	1	2	3	4	5		
4- Acho que precisaria de ajuda de um técnico para conseguir utilizar a aplicação.							
	1	2	3	4	5		
5- Considerei que as várias funcionalidades da aplicação estavam bem integradas.							
	1	2	3	4	5		
6- Achei que a aplicação tinha muitas inconsistências.							
	1	2	3	4	5		
7- Imagino que a maioria das pessoas aprenderia a usar esta aplicação rapidamente.							
	1	2	3	4	5		
8- Considerei que a aplicação era muito complicada de usar.							
	1	2	3	4	5		
9- Senti-me muito confiante a utilizar a aplicação.							
	1	2	3	4	5		
10- Precisei de aprender muito antes de conseguir utilizar a aplicação.							
	1	2	3	4	5		

Obrigado pela sua participação!

Anexo 8- QUESTIONÁRIO GUI

Número de participante:

Para cada uma das afirmações abaixo, selecione a caixa (apenas uma por linha) que melhor descreve a sua opinião relativamente à aplicação que acabou de experimentar:

	Discordo totalmente					Concordo totalmente
1- Entendi os nomes dos ecrãs da aplicação.						
	1	2	3	4	5	
2- Não é fácil navegar pelos diferentes ecrãs da aplicação.						
	1	2	3	4	5	
3- Entendi as diferentes opções e instruções dos ecrãs da aplicação.						
	1	2	3	4	5	
4- Os ícones e botões dos ecrãs não indicam bem a sua funcionalidade.						
	1	2	3	4	5	
5- Os diferentes botões e opções são fáceis de ler e entender.						
	1	2	3	4	5	
6- O design gráfico dos ecrãs não está bem desenhado.						
	1	2	3	4	5	
7- As informações estão bem organizadas nos ecrãs.						
	1	2	3	4	5	
8- Não é fácil aceder às informações de que preciso.						
	1	2	3	4	5	
9- O tamanho dos botões e das palavras ou frases é adequado.						
	1	2	3	4	5	

Obrigado pela sua participação!

Anexo 9 - DEBRIEFING/EXPLICAÇÃO DA INVESTIGAÇÃO

Muito obrigado por ter participado neste estudo. Conforme adiantado no início da sua participação, o estudo incide sobre a avaliação de uma aplicação de saúde, permitindo que os potenciais utilizadores a experimentem em primeira pessoa. Esta avaliação serve para identificar problemas e oportunidades de melhoramento da aplicação. Após esta iteração e o melhoramento da aplicação, será realizada a avaliação sumativa para avaliar a sua usabilidade e a satisfação dos potenciais utilizadores.

No âmbito da sua participação (explicação do procedimento que era expectável, para a realização das tarefas que não foram concluídas com sucesso, quando aplicável). Reforço que em nenhum momento esteve a ser avaliado. A aplicação de saúde é que estava a ser avaliada e não sua capacidade em realizar as tarefas.

Reforçamos os dados de contacto que pode utilizar caso deseje colocar uma dúvida, partilhar algum comentário, ou assinalar a sua intenção de receber informação sobre os principais resultados e conclusões do estudo:

Ana Vieira (ana_beatriz_vieira@iscte-iul.pt) – Investigadora responsável

Luís Brito Elvas (Luis.Elvas@iscte-iul.pt) – Investigador

Equipa de orientação:

Prof. Miguel Sales Dias (miguel.dias@iscte-iul.pt) - Orientador

Mais uma vez, obrigado pela sua participação.

Anexo 10 - GRELHA DE REGISTO DE REALIZAÇÃO DE TAREFAS

Número de participante:

Tarefa 1

Passos	Realizado	Não realizado	Tentativas*	Tempo
Clicar no botão "Frequência Cardíaca"				
Inserir o dedo na "câmara" traseira				
Notas: (comentários do participante durante a realização da tarefa, comportamentos, etc)				

*Assinalar o número de ações realizadas até realizar a ação correta. Se o participante não conseguir realizar a tarefa após três tentativas, registar como não concluída, e passar à seguinte

Tarefa 2

Passos	Realizado	Não realizado	Tentativas*	Tempo
Clicar no botão do ecrã dos registos				
Clicar no botão "Frequência Cardíaca"				
Notas: (comentários do participante durante a realização da tarefa, comportamentos, etc)				

*Assinalar o número de ações realizadas até realizar a ação correta. Se o participante não conseguir realizar a tarefa após três tentativas, registar como não concluída, e passar à seguinte

Tarefa 3

Passos	Realizado	Não realizado	Tentativas*	Tempo
Clicar no botão do ecrã dos lembretes				
Marcar uma cruz no medicamento “Ibuprofeno”				
Notas: (comentários do participante durante a realização da tarefa, comportamentos, etc)				

*Assinalar o número de ações realizadas até realizar a ação correta. Se o participante não conseguir realizar a tarefa após três tentativas, registar como não concluída, e passar à seguinte

Tarefa 4

Passos	Realizado	Não realizado	Tentativas*	Tempo
Clicar no botão “Exames e Consultas”				
Notas: (comentários do participante durante a realização da tarefa, comportamentos, etc)				

*Assinalar o número de ações realizadas até realizar a ação correta. Se o participante não conseguir realizar a tarefa após três tentativas, registar como não concluída, e passar à seguinte

Tarefa 5

Passos	Realizado	Não realizado	Tentativas*	Tempo
Clicar no botão “+”				
Notas: (comentários do participante durante a realização da tarefa, comportamentos, etc)				

*Assinalar o número de ações realizadas até realizar a ação correta. Se o participante não conseguir realizar a tarefa após três tentativas, registar como não concluída, e passar à seguinte

Tarefa 6

Passos	Realizado	Não realizado	Tentativas*	Tempo
Clicar no botão do ecrã do perfil				
Notas: (comentários do participante durante a realização da tarefa, comportamentos, etc)				

*Assinalar o número de ações realizadas até realizar a ação correta. Se o participante não conseguir realizar a tarefa após três tentativas, registar como não concluída, e passar à seguinte

Anexo 11 – PROTÓTIPO EM PAPEL

