2ARTs – Decision Support System for Exercise and Diet Prescriptions in Cardiac Recovery Patients

Master's Degree in Computer Engineering - Mobile Computing

Andreia Alexandra Sousa Pereira

Leiria, September of 2023

2ARTs – Decision Support System for Exercise and Diet Prescriptions in Cardiac Recovery Patients

Master's Degree in Computer Engineering - Mobile Computing

Andreia Alexandra Sousa Pereira

Project Report under the supervision of Professor Ricardo Martinho, Professor Rui Rijo and Professor Carlos Grilo, professors at the Escola Superior de Tecnologia e Gestão de Leiria.

Leiria, September of 2023

Originality and Copyrights

Acknowledgements

I want to express my sincere gratitude to the people and organizations that played a key role in the completion of this thesis. First and foremost, I owe an immeasurable debt of gratitude to my professors and the coordinators of this project, whose limitless patience and insightful feedback guided me through this academic journey.

To my classmates and groupmates, in particular, to my dear friends, I am grateful for their unwavering moral support. I must acknowledge the unfailing support of my family, with a special mention to my mother and boyfriend. Their unwavering belief in me has been the keystone of my perseverance and motivation throughout this process.

Finally, a note of gratitude to my feline companion, who is always fun and emotionally supportive, and who has provided moments of comfort amid academic challenges.

Abstract

The global health care system is faced with a variety of complicated challenges, ranging from limited access and increasing expenses to an aging population causing increased pressure on healthcare systems. Healthcare professionals are seeking alternative approaches to provide fair access and sustain high-quality care for everyone as a result of these challenges. Patients have historically been restricted from accessing essential healthcare services due to traditional barriers like geographic distance, financial and resource limitations. Innovative solutions to these problems are starting to take shape, thanks to the growth of eHealth platforms that use technology to improve patient care. Through a comprehensive study of existing solutions in the healthcare domain, particularly in cardiology, we identified the need for a Decision Support System (DSS) that would empower physicians with valuable insights and facilitate informed physical and diet prescribing practices into Cardiac Rehabilitation Programmes (CRPs).

The major goal of 2ARTs' project is to create and implement a cardiac rehabilitation platform into a hospital's infrastructure. A key aspect of this platform is the integration of a decision support system designed to provide physicians with valuable information when prescribing individualized treatment prescriptions for each patient, minimizing the potential of human error. The DSS uses algorithms and predictive models to classify patients into distinct groups based on their features and medical history. This classification provides critical insights and additional knowledge to doctors, allowing them to make informed judgments regarding the most effective treatment options for each patient's cardiac rehabilitation journey. By using the power of data-driven analytics and machine learning, the DSS enables doctors to better understand each patient's needs and personalize treatment actions accordingly.

In order to achieve the best possible results aligned with the goals of the project, a variety of approaches based on comprehensive studies were explored, specifically feature selection and feature reduction methods, where their performance metrics were evaluated, seeking the most effective solution. It was through this thorough analysis that Principal Component Analysis (PCA) emerged as the standout choice. PCA not only demonstrated superior outcomes in evaluation metrics, but also showcased excellent compatibility with the selected clustering algorithm along with the best results after an expert analysis. Moreover, with the analysis of the data types and features the dataset had, the K-Means algorithm produced the best results and was more adaptable to our dataset. We were able to identify useful insights and patterns within the data by employing both PCA and K-Means, opening the way for more accurate and informed decision-making in the 2ARTs project.

Keywords: Cardiac Rehabilitation Program, Clustering Algorithms, Decision Support Systems, eHealth, Machine Learning, Predictive Models

Resumo

O sistema global de saúde enfrenta diversos desafios complexos, que vão desde o seu acesso limitado ao envelhecimento da população, que provoca uma pressão crescente sobre estes sistemas. Em resultado destes problemas, os médicos procuram estratégias alternativas para proporcionar um acesso justo e manter cuidados de elevada qualidade para todos. Historicamente, os doentes têm sido impedidos de aceder a serviços de saúde essenciais devido a restrições tradicionais como a distância geográfica ou dificuldades financeiras. Soluções inovadoras para estes problemas começam a surgir, graças ao crescimento das plataformas eHealth que utilizam a tecnologia para melhorar os cuidados de saúde prestados aos doentes. Através de um estudo das soluções existentes no domínio dos cuidados de saúde, em particular na área da cardiologia, identificámos a necessidade de um sistema de apoio à decisão (DSS) que permitisse aos médicos obter informações valiosas e facilitar práticas de prescrição de exercício e nutrição informadas nos programas de reabilitação cardíaca (CRP).

O principal objetivo do projeto 2ARTs é criar e integrar uma plataforma de reabilitação cardíaca na infraestrutura de um hospital. Um aspeto fundamental desta plataforma é a integração de um DSS concebido para fornecer informações úteis aos médicos quando estes prescrevem tratamentos individualizados para cada doente, minimizando o potencial de erro humano. O DSS utiliza algoritmos complexos e modelos preditivos para classificar os doentes em grupos distintos com base nas suas características únicas e historial médico. Esta classificação fornece informações úteis e conhecimentos adicionais aos médicos, permitindo-lhes tomar decisões informadas relativamente às opções de tratamento mais eficazes para o percurso de reabilitação cardíaca de cada paciente. Ao utilizar o poder da análise de dados e do *Machine Learning*, o DSS permite aos médicos compreender melhor as necessidades de cada doente e personalizar as ações de tratamento em conformidade.

Para alcançar os melhores resultados possíveis, foram exploradas várias abordagens baseadas em estudos de referência, especificamente métodos de seleção e redução de características, em que foram avaliadas as suas métricas de desempenho, procurando a solução mais eficaz. Foi através desta análise que a Análise de Componentes Principais (PCA) emergiu como a escolha de destaque. O PCA não só demonstrou resultados superiores nas métricas de avaliação, como também revelou uma excelente compatibilidade com o algoritmo de *clustering* selecionado, juntamente com os melhores resultados após uma análise especializada. Além disso, com a análise das características do conjunto de dados, o algoritmo *K-Means* produziu bons resultados e foi o mais adequado. Conseguimos identificar informações e padrões úteis nos dados utilizando o PCA e o *K-Means*, abrindo caminho para uma tomada de decisões mais precisas e informadas no projeto.

Palavras-chave: Programa de Reabilitação Cardíaca, Algoritmos de *Clustering*, Sistemas de Apoio à Decisão, eHealth, *Machine Learning*, Modelos Preditivos

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

AACVPR	The American Association of Cardiovascular and Pulmonary Rehabilitation
AHA	American Heart Association
AHCPR	Agency for Health Care Policy and Research
AI	Artificial Intelligence
ANS	Autonomic Nervous System
API	Application Programming Interface
BMI	Body Mass Index
CAD	Coronary Artery Disease
CDS	Computerized Decision Support
CR	Cardiac Rehabilitation
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRP	Cardiac Rehabilitation Programmes
CRT	Cardiac Rehabilitation Therapy
CVD	Cardiovascular Disease
DSS	Decision Support System
ECG	Electrocardiogram
ECR	Exercise Cardiac Rehabilitation
EHR	Electronic Health Record
ESTG	School of Technology and Management
FAMD	Factorial Analysis of Mixed Data
FCT	Fundação para a Ciência e a Tecnologia
HIE	Health Information Exchange
HR	Heart Rate
HRQoL	Health-Related Quality of Life
HRR	Heart Rate Recovery
HRV	Heart Rate Variability
ICU	Intensive Care Unit
IoT	Internet of Things
MCA	Multiple Correspondence Analysis
MI	Myocardial Infarction
ML	Machine Learning

MSNA	Muscle Sympathetic Nerve Activity
PATHway	Physical Activity Towards Health
PCA	Principal Component Analysis
PNS	Parasympathetic Nervous System
PTCA	Percutaneous Transluminal Coronary Angioplasty
SNS	Sympathetic Nervous System
WCSS	Within-Cluster Sum of Square

1. Introduction

Healthcare systems around the world face numerous challenges, including accessibility, affordability, quality of care and an aging society increasing the burden on these systems, leading healthcare individuals looking for answers to both cost and quality challenges [1], therefore the importance of finding other ways to deliver healthcare for everyone. Many people struggle to access healthcare services due to various barriers such as geographical distance, financial constraints, and limited healthcare resources [1]. As a result, there is a growing need to find innovative solutions that can deliver healthcare to everyone, regardless of their location or socio-economic background, which is leading to several eHealth platforms that have grown into effective methods to enhance patient care [2]. Technology has emerged as a promising tool to address the challenges of healthcare delivery. By leveraging advancements in Information and Communications Technology (ICT), we can bridge the gaps in healthcare access and improve the overall delivery of care.

Here are some key solutions that ICT offers in delivering healthcare to everyone:

- **Telemedicine and Telehealth:** Telemedicine enables patients to connect with healthcare providers remotely, overcoming geographical barriers. Through video consultations, telemedicine allows individuals to receive medical advice, diagnosis, and even treatment from the comfort of their homes. Telehealth expands on this concept by incorporating a broader range of healthcare services, including remote monitoring of patients' health conditions and sharing of medical data [3].
- **Mobile Health (mHealth) Applications:** With the widespread availability of smartphones, mobile health applications have gained significant popularity. These apps provide patients with easy access to health information, appointment scheduling, medication reminders, and even personalized health tracking, empowering users to take control of their health and engage in self-care practices [3].
- **Electronic Health Records (EHR):** Traditional paper-based medical records can be inefficient and prone to errors, and that is where Electronic Health Records (EHR) can come in and digitize patient information, making it easily accessible to healthcare providers. EHR systems streamline the sharing of medical records among healthcare facilities, ensuring continuity of care and reducing redundant procedures [4].
- **Wearable Devices and Remote Monitoring:** Wearable devices, such as fitness trackers and smartwatches, enable individuals to monitor their health parameters in real-time, tracking multiple biomedical data such as heart rate, blood pressure, sleep patterns, and other vital signs. Remote monitoring systems allow healthcare providers to remotely monitor patients' health conditions and intervene when necessary, reducing the need for frequent in-person visits [3], [5].

- **Artificial Intelligence (AI) in Healthcare:** AI technology has the potential to revolutionize healthcare delivery, analysing vast amounts of medical data to identify patterns, detect diseases at early stages, and provide personalized treatment recommendations. AI-powered chatbots and virtual assistants can also offer 24/7 healthcare support, answering common medical questions and triaging patients based on their symptoms [6].
- **Health Information Exchange (HIE):** Interoperability and secure exchange of health information among different healthcare providers are essential for coordinated care. Health Information Exchange (HIE) platforms facilitate the seamless sharing of patient data, ensuring that healthcare professionals have access to comprehensive and up-to-date information when making treatment decisions [4].

By leveraging these technological solutions, healthcare systems can overcome the limitations of traditional care delivery models. They enable healthcare providers to reach underserved populations, improve efficiency, enhance patient engagement, and ultimately deliver better healthcare outcomes for everyone. However, it is important to ensure that these technologies are implemented in an inclusive and equitable manner, considering factors such as digital literacy, privacy and security concerns [3], [4], as well as addressing the digital divide to ensure that technology does not exacerbate existing healthcare disparities [5]. Collaborative efforts between healthcare providers, policymakers, and technology developers are crucial in harnessing the full potential of technology to deliver healthcare to everyone and transform the future of healthcare delivery [5].

It is essential to acknowledge the broader challenges faced by healthcare systems because even within specialized areas like cardiac rehabilitation, these challenges have a significant impact. Cardiac Rehabilitation (CR) is an integral part of the healthcare system, focusing on the recovery and improvement of heart health for individuals who have experienced cardiac events or undergone cardiac procedures [7]. The challenges that exist within healthcare systems, such as accessibility, affordability, and quality of care [1], also extend to cardiac rehabilitation services. These challenges can prevent individuals from receiving the necessary care and support to recover and prevent future cardiovascular issues. Moreover, the quality of cardiac rehabilitation programmes can be affected by systemic challenges within the healthcare system. We may more fully understand the environment in which cardiac rehabilitation occurs by comprehending the relationship of cardiac rehabilitation with the larger healthcare system concerns. It highlights the need for comprehensive solutions that address not only the specific challenges within cardiac rehabilitation but also the systemic issues impacting healthcare delivery.

1.1.Problem and Motivation

The existing CR programmes face significant challenges regarding patient participation and adherence, despite the well-documented evidence of their benefits in enhancing recovery and reducing mortality following a myocardial infarction. A considerable proportion of patients, approximately one third, do not participate in CR programmes, and the adherence rates are even lower, with only about one third of participants maintaining attendance after 6 months [7].

Numerous research studies have explored the factors associated with participation and adherence in CR programmes, shedding light on the multifaceted nature of this problem. Several consistent factors have emerged as barriers to participation, including:

- lack of referral by doctors (see Figure 1);

Figure 1 - Cardiac Rehabilitation Referral (source: [8]**)**

- associated illnesses;
- specific cardiac diagnoses;
- reimbursement issues;
- perceived benefits of CR;
- distance and transportation challenges;
- self-motivation;
- family composition;
- social support;
- self-esteem.

Only 8% of patients discharged from the hospital following a myocardial infarction in Portugal are enrolled in CR programmes. The percentage of those admitted to CR programmes in Europe is 30%, while it is 20-30% in the United States [9]. Portugal has one of the lowest inclusion rates in these programmes when compared to other European countries. However, there has been a rise in the number of rehabilitation institutions in recent years, which has resulted in an increase in the number of referrals [10].

The integration of technology, such as monitoring devices or remote rehabilitation platforms, can improve the convenience and efficacy of non-hospital-based CR programmes. These advancements enable healthcare professionals to remotely monitor patients' progress, provide real-time feedback, and provide individual guidance while addressing the issue for physical relocation [11], as well as other above-mentioned issues like self-esteem or occupation since they do not need to meet with the doctors personally and go to the hospital at specific hours. This, in turn, can improve patient engagement, adherence, and, ultimately, the success of cardiac rehabilitation in improving outcomes for those suffering from cardiovascular disease. However, monitoring patients from afar involves many challenges that must be addressed for this approach to be reliable and efficient.

The flexibility to access patient records from anywhere, at any time, has improved the value of healthcare delivery [12]. These records can provide information about each patient that may be relevant, which leads to the importance of information. Information is crucial for new advancements and adjustments in organizations because it increases knowledge and enables the achievement of the best outcomes possible. However, doctors may lack the capacity to properly understand the vast amount of data collected, making it overwhelming to monitor as many patients as necessary [13]. Additionally, input errors from patients manually entering their data into these healthcare platforms can also be time-consuming for doctors, delay the delivery of crucial data, and result in incomplete or incorrect patient profiles. For that issue, computer-based solutions like Clinical Decision Support Systems (CDSS) are recommended as a solution for overcoming such barriers, as they have evolved into an important tool in the healthcare industry, supporting doctors in making informed choices [14].

CDSS are computer-based tools, that act as a digital assistant, giving doctors the background information that is needed to make informed decisions based on the patient's medical history, augmenting the capability of these doctors, rather than replacing them and supporting them in interpreting patient data, identifying patterns, and predicting future results [15]. By leveraging the power of advanced analytics and machine learning, these systems have shown that they can increase the effectiveness of healthcare delivery, improve patient outcomes, and lower medical errors [14], [16]. Besides, their predicting power is yet another beneficial feature. CDSS can spot trends, predict the progression of diseases, and calculate the likelihood of an outcome to happen by using historical patient data. This helps doctors anticipate risks properly, track patient progress and modify treatments as necessary.

1.2. Objectives

In this work, our focus is to integrate a CDSS with the 2ARTs health platform. By combining them, we aim to empower healthcare professionals by providing them with tools to closely monitor CR patient progress, analyse biomedical data, and make informed decisions.

Specifically, the objectives of this research are as follows:

- **Design and develop the 2ARTs CR health platform:** The first objective is to create a user-friendly health platform that makes data management easier, patient monitoring, and collaboration among healthcare professionals and their patients. The platform will be designed to integrate with an existing mobile app that patients use, ensuring that they are always communicating and have real-time data available between each. The aim is to enhance healthcare delivery and improve patient care outcomes through the development of a reliable and efficient health platform;

- **Design and develop a CDSS architecture:** The second objective is to design and develop an architecture for the CDSS. This involves identifying the key components, data requirements, and algorithms that will enable effective decision-making support within the health platform;
- **Integrate the CDSS with the health platform:** The third objective is to establish seamless integration between the CDSS and the existing health platform. This integration should allow for smooth data exchange and easy access to decision support functionalities involving leveraging patient biomedical data and other relevant information to generate predictive models and actionable insights;
- **Enhance decision-making capabilities:** The fourth objective is to empower healthcare professionals with the decision-making capabilities through the CDSS. This includes providing them with possible outcomes for each patient's health based on the analysis of their data, enabling doctors to make more informed decisions, improving patient care, and optimizing treatment outcomes;
- **Evaluate the performance and usability of the CDSS:** The final objective is to assess the performance and usability of the developed CDSS within the integrated health platform. This involves conducting thorough evaluations, analysis and gathering feedback from an expert, Professor and MD, Rui Pinto, from Instituto Politécnico de Leiria.

By accomplishing these objectives, this research aims to contribute to the advancement of healthcare technology by developing a CDSS that seamlessly integrates with a health platform. The goal is to empower healthcare professionals with valuable insights and tools that improve the quality of patient care and drive better health outcomes.

1.3. Document’s Structure

The rest of this document is structured as follows:

- **Chapter 2, Concepts and Related Work:** fundamental ideas used to explain the subject matter of the research and review of existing literature, studies, and advancements related to the research area, providing context and identifying gaps for the proposed study;
- **Chapter 3, Methodology:** detailed description of the procedures, techniques, and tools employed to conduct the platform development and the research study, enabling validation of the results;
- **Chapter 4, 2ARTs Backoffice:** provides a detailed account of the practical steps taken to develop and deploy the proposed health platform and system. It covers

aspects such as technologies used, the platform architecture and the connection between the model and the platform;

- **Chapter 5, Development of Classification Model:** describes the used dataset as well as the steps and approaches accounted to develop the decision support system model;
- **Chapter 6, Results:** presents the results and the analysis of the results of each approach used and the expert's analysis;
- **Chapter 7, Conclusions and Future Work:** a summary of the work accomplished, the goals achieved, the challenges faced, limitations and future work.

2. Concepts and Related Work

Cardiac patients often experience multiple cardiovascular events, making it crucial for them to receive continuous monitoring and care from healthcare professionals. The monitoring process enables early detection of potential complications and helps mitigate the risk of future cardiac events. By closely observing patients' health status and tracking their progress, healthcare professionals can tailor personalized interventions, provide timely interventions, and implement preventive measures to improve patient outcomes. This proactive approach is essential in minimizing the occurrence of adverse events and promoting long-term cardiovascular health and well-being.

In order to address this challenge, Cardiac Rehabilitation Programmes (CRP) play a vital role in the management of cardiovascular patients. However, certain patients may face barriers that prevent them from physically attending these programmes. To overcome this limitation, the integration of a virtual platform allows patients to participate in rehabilitation remotely, expanding access to a larger patient population. This is where a CDSS becomes invaluable. The CDSS enables healthcare professionals to collect and analyse data from multiple patients, providing a comprehensive and centralised view of their progress. By leveraging advanced analytics and minimizing the potential for human error, the CDSS empowers doctors to make informed decisions and deliver personalized care, ensuring optimal outcomes for a greater number of patients.

We will discuss key ideas and related work to provide a thorough grasp of our problem in the sections below. We aim to clarify the complexities of our subject and gather knowledge from previous research and solutions by outlining these concepts. We can comprehend the existing environment and identify the potential for improvement by analysing similar work.

2.1. Concepts

To give a thorough understanding, we go more into each of these concepts in the subsections that follow. We start by looking at cardiac rehabilitation programmes and go into details on the goals, factors, and advantages of them. The possible implications of untreated or improperly managed cardiac disorders are then looked at, with a focus on the significance of timely and efficient therapies. Finally, we explore the idea of a CDSS and how it relates to healthcare. We look at how a CDSS uses clinical knowledge and data analysis to help medical professionals make the best choices possible for patients.

2.1.1. Cardiovascular Diseases

Cardiovascular diseases (CVD) encompass a wide range of conditions that affect the heart and blood vessels. Some common cardiovascular diseases include coronary artery disease, myocardial infarction, heart failure, stroke, and peripheral artery disease. It develops gradually as fatty deposits, known as atheroma, accumulate in the arteries, leading to reduced blood supply to the heart and other parts of the body. This process is commonly referred to as "hardening of the arteries" [17]. When CVD affects the heart, it can result in episodes of angina, characterized by chest pain, or more severe events like a heart attack. If CVD affects the brain, it can lead to a stroke, either a major stroke or a mini-stroke (also called a transient ischemic attack).

The development of CVD is a gradual process that occurs over many years. As the blood supply to the heart is compromised by atheroma build-up, the risk of angina, heart attacks, strokes, and peripheral vascular issues increases [17]. Without proper medical care and support, individuals with cardiovascular diseases may face a range of complications and adverse outcomes. Additionally, untreated or poorly managed cardiovascular diseases can lead to long-term damage to the heart, impaired organ function, and even premature death. Regular monitoring of blood pressure, cholesterol levels, and other key indicators, along with appropriate adjustments to medications and lifestyle interventions, can significantly improve outcomes [18]. Cardiac rehabilitation programmes, as well as close collaboration with healthcare professionals, play a vital role in providing comprehensive care, education, and support to patients, helping them to minimize the risks associated with cardiovascular diseases and achieve better overall health and well-being.

2.1.2. Cardiac Rehabilitation Programmes

Cardiac rehabilitation programmes are designed to address the psychological and physiological stresses associated with cardiovascular disease (CVD) and to reduce the risk of mortality secondary to CVD. The main goal is to improve cardiovascular function and help patients achieve the highest possible quality of life, achieved through multiple interventions, including improving overall cardiac function and capacity, slowing or reversing the progression of atherosclerotic disease, and increasing patients' self-confidence through gradual conditioning [19], [20]. To ensure the effectiveness of cardiac rehabilitation programmes, several organizations, such as the American Heart Association (AHA), the American Association of Cardiovascular and Pulmonary Rehabilitation (AACVPR), and the Agency for Health Care Policy and Research (AHCPR), recommend specific core components, as seen in Figure 2. These components aim to optimize cardiovascular risk reduction, reduce disability, promote active and healthy lifestyle changes, and facilitate the maintenance of these healthy habits even after completing the rehabilitation program. The key components include patient assessment, nutritional counselling, risk factor control, patient education, psychosocial management, vocation advice, physical activity counselling and exercise training [21].

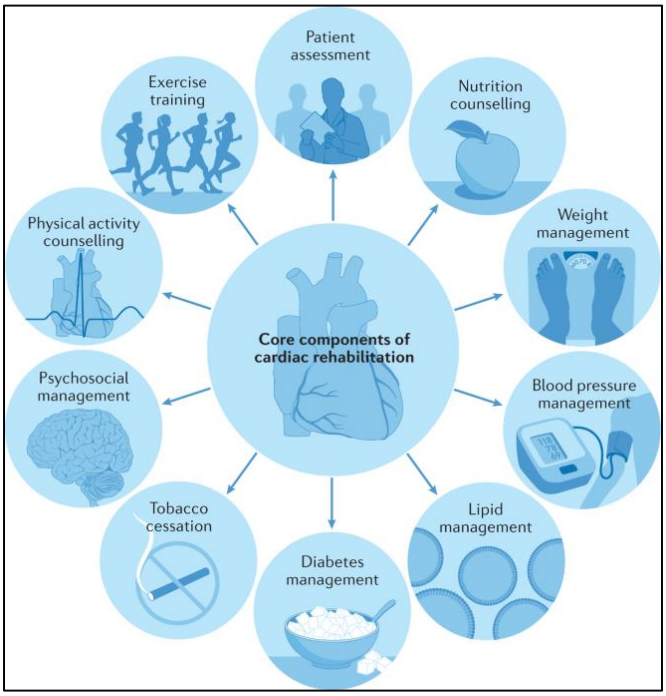

Figure 2 - Components of CRP (source: [20]**)**

Cardiovascular rehabilitation programmes are essential to comprehensive care for people with cardiovascular disease. They provide a multidisciplinary strategy that combines medical, lifestyle, and psychosocial therapies to improve patient long-term health and promote better results. These programmes enable patients to take an active role in their own cardiovascular health and make long-lasting lifestyle changes by offering suited care, information and support [20]. Cardiovascular rehabilitation programmes progressively help people with coronary cardiovascular disease live better overall.

2.1.3. Digital Health

Digital health refers to the usage of ICT in healthcare to manage illnesses, mitigate health risks, and promote overall well-being. It encompasses a wide range of applications, including wearable devices, mobile health, telehealth, health information technology, and telemedicine [22]. The field of digital health has gained significant traction due to its potential to bring multiple benefits, one of them being the improved access to healthcare services. Through telehealth and telemedicine, patients can receive remote consultations and access specialized care, overcoming geographical barriers and enhancing healthcare accessibility, especially for those in underserved areas. This not only increases convenience for patients but also alleviates the burden on healthcare systems by reducing unnecessary hospital visits [22].

In addition to improving access and quality of care, digital health can also help to reduce healthcare costs [23]. By enabling remote care, digital health can help to lower healthcare spending, improving efficiency by streamlining administrative tasks and reducing paperwork.

While digital health has the potential to revolutionize healthcare delivery, there are also several challenges that must be addressed to fully realize its benefits [23]. Some of these challenges include:

- **Data privacy and security:** relying on the collection and sharing of sensitive patient data, can be vulnerable to cyber-attacks and data breaches;
- **Interoperability:** these systems may use different standards and formats for data, making it difficult to share information between them;
- **Regulatory barriers:** they are subject to a complex web of regulations and standards, which can vary by country and region;
- **User adoption:** digital health technologies are only effective if they are adopted and used by patients and healthcare providers;
- **Cost:** while digital health has the potential to reduce healthcare costs in the long run, there may be significant upfront costs associated with implementing digital health technologies.

2.1.4. Decision Support Systems

A Decision Support System (DSS), as mentioned before, is an essential tool that utilizes the power of Artificial Intelligence (AI) and Machine Learning (ML) algorithms to assist decision-makers in complex and data-driven scenarios. It is designed to help managers and other decision-makers analyse complex problems and make informed decisions based on data and models. They can be used in a variety of industries, including healthcare, finance, marketing, and logistics. In healthcare, Clinical DSS (CDSS) can help healthcare professionals in making clinical decisions by providing them with relevant patient data, medical knowledge, and decision-making tools [15]. One of the key benefits of these systems, which is depicted in Figure 3, is that they can help decision-makers make more informed decisions by providing them with accurate and timely information [24]. They can also help decision-makers evaluate different options and scenarios as well as identify potential risks and opportunities.

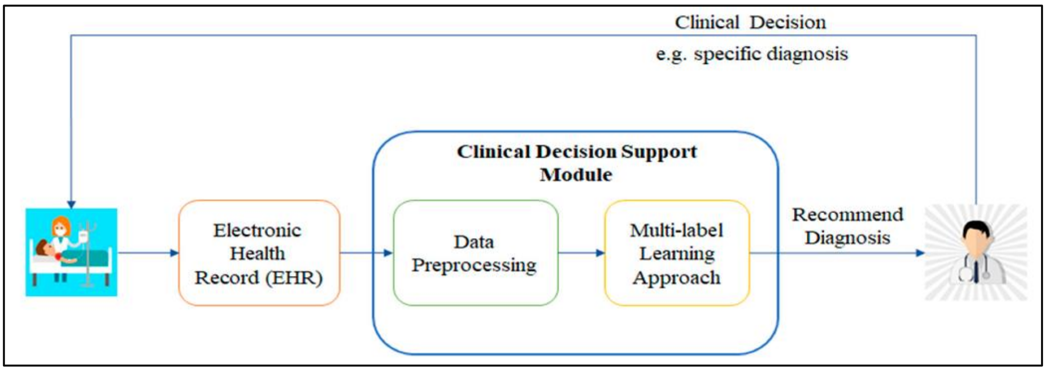

Figure 3 – Clinical Decision Support System for Medical Diagnosis (source: [24])

The following points highlight the key contributions of AI and ML in the context of CDSS:

- **Artificial Intelligence:** The data generated by the patients and the professionals who care for them can be analysed by AI technology, which can also identify patterns and insights that people would not have discovered on their own. AI algorithms can be used by healthcare organizations to assist them by making better operational and clinical choices and raise the standard of the services they offer [25]. There are many different forms and elements of healthcare data, and for that, organizations may integrate fragmented data to create a more comprehensive picture of the individuals behind the data by employing AI and Machine Learning technology [26];

- **Machine Learning:** As a subset of Artificial Intelligence, ML plays a fundamental role in many healthcare innovations, including the development of new medical solutions starting with the handling of patient data and records and the treatment of chronic diseases [27]. In this project, this method is mostly used to group patients at a single phenotype, i.e., from a data set, a clustering of several groups is built to detect the level where the patient is at and thus group all the patients at their respective levels. For each level, it is later possible to prescribe for everyone that belongs to that cluster, breaking down the traditional “one size fits all”.

2.2.Related Work

This section describes how the related works presented in this document were found, in which concept within the area of cardiac rehabilitation they are situated and how they relate to the problem tackled in this work.

Although there are tools, such as ResearchGate and ScienceDirect, we have used Google Scholar for convenience, since it enables us to find the articles existing in all the other platforms, as well as the number of citations. In terms of keywords, the articles noted in Table 1, were found from expressions such as "Artificial Intelligence Cardiac Disease Treatment", "Artificial Intelligence in Cardiology Decision Support Exercise", "Artificial Intelligence Cardiovascular Disease", “Decision Support for Cardiac Therapy Exercises” and "Cardiac Disease Rehab Decision Support". In addition to these, the references cited in the 2ARTs project application document for the Fundação para a Ciência e Tecnologia (FCT) were also reviewed.

Table 1 sums-up the referenced found for Artificial Intelligence in Cardiovascular Diseases, Decision Support Systems (DSS), and Exercise Cardiac Rehabilitation (ERC). The following sections detail the main conclusions of these works.

Table 1 - Articles related to Cardiac Rehabilitation

		CONCEPT – CARDIAC REHABILITATION		
		AI in Cardiovascular Diseases	DSS	ECR
ARTICLES RELATED	Haq et al. (2022). Artificial Intelligence in Cardiovascular Medicine: Current Insights and Future Prospects [28]	x		
	Long et al. (2019). Exercise-based cardiac rehabilitation for stable angina: systematic review and meta-analysis. [29]			x
	Bjarnason-Wehrens et al. (2020). Exercise-based cardiac rehabilitation in patients with reduced left ventricular ejection fraction: The Cardiac Rehabilitation Outcome Study in Heart Failure (CROS-HF): A systematic review and meta-analysis. [30]			x
	Ojha, S. (2022). Recent Advancements in Artificial Intelligence Assisted Monitoring of Heart Abnormalities and Cardiovascular Diseases: A Review. [31]	x		
	Vromen et al. (2020). A computerized decision support system did not improve personalization of exercise-based cardiac rehabilitation according to latest recommendations. [32]		x	x
	Triantafyllidis et al. (2018). Computerized decision support for beneficial home-based exercise rehabilitation in patients with cardiovascular disease. [33]		x	x
	Méa Plentz (2019). Relevance and Limitations of Decision Support Systems for Outpatient Cardiac Rehabilitation. [34]		x	
	Ishraque et al. (2018). Artificial intelligence-based cardiac rehabilitation therapy exercise recommendation system. [35]	x	x	x
	Rahman et al. (2019). Variables Influencing Machine Learning-Based Cardiac Decision Support System: A Systematic Literature Review. [36]	x	x	
	Pescatello et al. (2021). Development of a Novel Clinical Decision Support System for Exercise Prescription Among Patients With Multiple Cardiovascular Disease Risk Factors. [37]		x	x

	Lopez-Jimenez et al. (2020). Artificial Intelligence in Cardiology: Present and Future. [6] ; Diller et al. (2019). Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. [38] ; Lin et al. (2021). Artificial Intelligence in Cardiovascular Imaging for Risk Stratification in Coronary Artery Disease. [39]	x		
	Barrett et al. (2019). Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care. [40]	x		

2.2.1. Artificial Intelligence in Cardiovascular Diseases

As shown in Table 1, articles mentioning Artificial Intelligence (AI) in the field of cardiovascular diseases refer to advances and AI applications in the cardiovascular area. This section summarizes the key findings in order to proceed with the use of AI in Decision Support Systems.

To give a brief allusion to one of the problems, patients who have cardiac diseases are unaware of a variety of internal physiological changes and only seek medical attention when their symptoms become severely ill or when they start to become worse. However, disease-related symptoms are rare and only manifest once the illness has advanced to the point where irreversible alterations have occurred [31]. This is where the need to monitor the patient's ongoing physical condition comes in, and not just when the patient goes to a health centre for a check-up. With the help of ML algorithms, it is possible to work with the data obtained continuously by both the health professional and the patient and thus hope for better monitoring and better prescribed programmes suited to what the patient has been feeling.

Even though conventional statistical methods for risk stratification have been developed, the majority of the models exhibit shortcomings in terms of predicting a patient's risk. The use of Artificial Intelligence (AI) and other next-generation solutions will enable patients to engage in self-care through eLearning and decision-supporting processes. These approaches are intended to minimal expenses, raise standards of treatment, and guarantee that everyone has access to healthcare regardless of location, time, or personnel availability [40].

AI in cardiology began with the development of self-learning neural networks to trace the path of the atrioventricular accessory in patients with Wolff-Parkinson-White Syndrome, and is now able to precisely stratify the disease, integrate multimodality imaging, remotely monitor and diagnose continuously as well as suggest a therapy selection [28]. Lately, Artificial Intelligence has been used to interpret echocardiograms, automatically detect heart rhythms from an ECG, identify an individual using their ECG as a biometric signal and detect the presence of cardiac diseases from the surface ECG, such as left ventricular failure [6].

2.2.2. Decision Support System

Cardiac Rehabilitation Therapy (CRT) was created to aid in the recovery and risk reduction of cardiac patients. The programmes are all-inclusive, offering patients, who have had cardiac events or suffered a heart attack, medical assessments, nutrition and lifestyle counselling, education, advice on managing risk factors (such as blood pressure, weight, diabetes, and smoking), psychosocial interventions, as well as group-based physical activity and instructed exercise training. However, sometimes this counselling can be given in a generalized way, i.e. in a "one size fits all" approach and not always in a personalized way for each patient's risk level, which can lead to unadvisable prescriptions for certain risk levels.

In [33], the authors used a rule-based approach to process the data in their CDSS for the patient's guidance in an unsupervised exercise-based rehabilitation program, with the supplementation of smart devices such as mobile phones, smart sensors and smartwatches. This system has already been published in Physical Activity Towards Health (PATHway), a computer-assisted platform that employs a virtual coach for CR exercises at home, and so they could evaluate their system in a simulation and real-world study, taking into account two metrics, a) Recovery from low heart rate (HR) events, which measures the percentage of the number of exercises where the CDSS guides the patient to exercise within beneficial HR zones immediately after a low HR event and b) Recovery from low movement accuracy events, which measures the percentage of occasions where the CDSS guides the patient to exercise immediately after a low motion accuracy event. The results for a) reached 83% and 93% in 2-min and 3-min window, respectively, and for b) it reached 85% and 100% in a 1-min and 2-min time window, respectively.

The authors in [35], proposed a model that in addition to the CDSS, was also composed of a chatbot, responsible for user feedback (patients and professionals) that could understand text and even speech, in order to understand emotion or feelings and levels of trust or motivation, integrated with a data-driven dashboard and a cloud-based analytics platform. Although the platform has not yet been put into production, and so there are no reliable statistics yet, the authors compared their model with a randomized model and after 10 independent experiments, observed that their model can find the optimal therapy plan with a 43% higher success rate on average success rate compared to the other, making it more effective for therapy planning.

In [37], the authors developed an evidenced-based, guided, and time-efficient tool P3-EX, that can be used for professionals to prescribe exercise for patients who may also have other chronic illnesses and health issues or many cardiovascular disease risk factors. This tool is also not yet in production and for that reason, it cannot yet be considered a reliable tool to prescribe exercise, however, the authors foresee it as future work.

The authors in [32] however, after developing a CDSS and testing it with data available from 2258 patients, did not find major changes in Exercise Cardiac Rehabilitation (ECR) prescribing, saying that the implementation of a CDSS by itself is insufficient and it needs complementation in order to be used.

Finally, in [36], the authors investigated common variables used in ML-based CDSS and the most used techniques used in cardiac care unit, being the age, gender, admission type, duration of stay, heart rate, and respiration rate the most frequent variables used in ML methods, and for the techniques, logistic and multivariate regression.

2.2.3. Exercise Cardiac Rehabilitation

There are several articles conducted between the last five years that mention or review the importance and question the effectiveness to use exercise in cardiac rehabilitation. These articles vary from using IoT devices to measure important variables in the cardiovascular area, such as heart rate, or just prompting the patient to perform exercise, either at home or with professional monitoring so that later it can be verified if there is an improvement, for example, being measured by microneurography.

In [41], a study was conducted with the goal of determining the effect of ECR, compared to usual care and psychosocial and/or educational interventions on resting and post-exercise parasympathetic function in patients with coronary artery disease and to identify the possible moderating variables of the ECR-induced effect. The population used in the study were female/male adults who have ever had any relation to cardiovascular disease, where they had to report their HRV, HRR, or both. In total and after study selection, 25 studies, 1346 patients from 11 different countries, were counted. The result obtained was that ECR improves post-exercise parasympathetic function, responsible for regulating rest and recovery functions, such as slowing down the heart rate [42], in patients with coronary artery disease, with greater evidence in younger patients and patients treated with percutaneous transluminal coronary angioplasty (PTCA).

In [43], the authors also did a systematic review and meta-analysis with the same goal as the above. In a total of 20 studies, with most male patients within an average age between 49 and 70 years, the results indicate statistically significant improvements in all three examined parameters of autonomic function (HRR, HRV, Muscle Sympathetic Nerve Activity (MSNA)). Of these three parameters, microneurography is the only tool that provides a measure of activity.

In [29], the authors sought to find out the ECR effects compared to usual care on mortality, morbidity, hospital admissions, exercise capacity, health-related quality of life, adverse effects, and return to work for adults with stable angina. For a total of 7 studies and 581 patients, the analysis showed that there may be an improvement in exercise capacity after CR compared to usual control in the short term (up to 12 months follow-up, low quality evidence). However, there was insufficient evidence to be able to definitively assess the impact of CR on mortality, morbidity or Health-Related Quality of Life (HRQoL), leading to a conclusion that ECR may make little or no difference in all-cause mortality in studies with follow-up of less than 12 months.

3. Methodology

In this chapter, we delve into the methodology used for the development of both the 2ARTs digital platform and the CDSS. We describe the step-by-step approach and techniques employed to design and implement the platform, ensuring it meets the required functionality and integrates with the CDSS. Additionally, we outline the methodology used to build this decision support system, which is an academic adaptation of the Cross-Industry Standard Process for Data Mining (CRISP-DM).

3.1. 2ARTs Digital Platform

The development of the 2ARTs digital platform followed an agile methodology, which is a flexible and iterative approach to software development. The team used a combination of Scrum and Kanban principles to manage the development process effectively. Scrum was used to structure the development process into sprints, with each sprint lasting a fixed amount of time, typically a week. During each sprint, the team would focus on developing specific features and tasks, and at the end of the sprint, a potentially product increment would be delivered. Daily stand-up meetings were held to ensure clear communication and collaboration among team members, and regular sprint reviews and retrospectives were conducted to gather feedback and continuously improve the development process.

Kanban was used to manage the flow of tasks and ensure a smooth development process. The team used a Kanban board, specifically with the use of the visual tool Trello as shown in Figure 4, to visualize the progress of tasks, from backlog to in-progress, doing, testing to completed. This provided a clear overview of the development status and helped identify and resolve any bottlenecks.

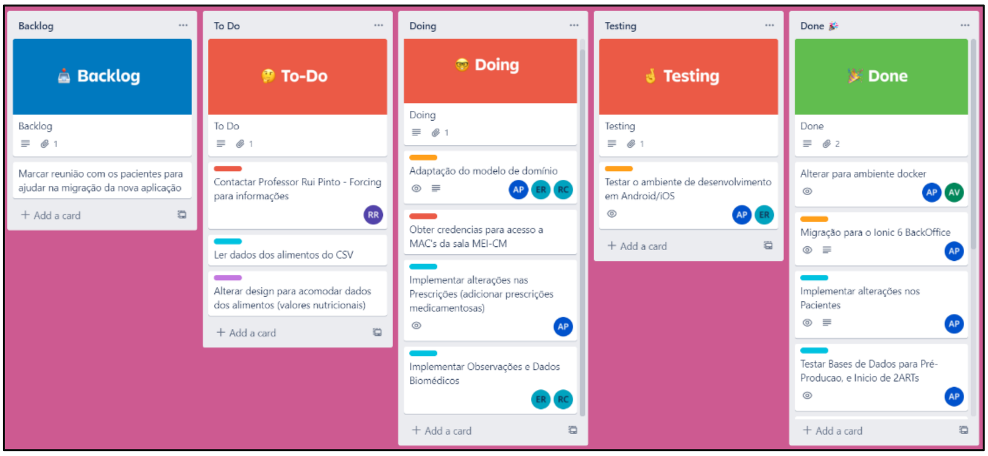

Figure 4 - Trello Board

The hybrid approach of Scrum and Kanban allowed the development team to be adaptable and respond quickly to changing requirements and priorities. It facilitated a collaborative and transparent development process, enabling the team to deliver a high-quality backoffice that met the project's objectives and requirements.

3.2. CDSS

The CRISP-DM framework is a widely recognized and well-established methodology for data mining projects. It provides a structured approach to guide the various stages involved in data analysis and knowledge discovery. In this chapter, we outline the academic adaptation of CRISP-DM specifically tailored to address the objectives and requirements of our problem. Our adaptation follows a similar structure to the original framework but incorporates modifications to suit the academic research context. It consists of six major stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment as shown in Figure 5. Each of them represents a distinct stage in the research process and contributes to the overall progression towards achieving the research objectives.

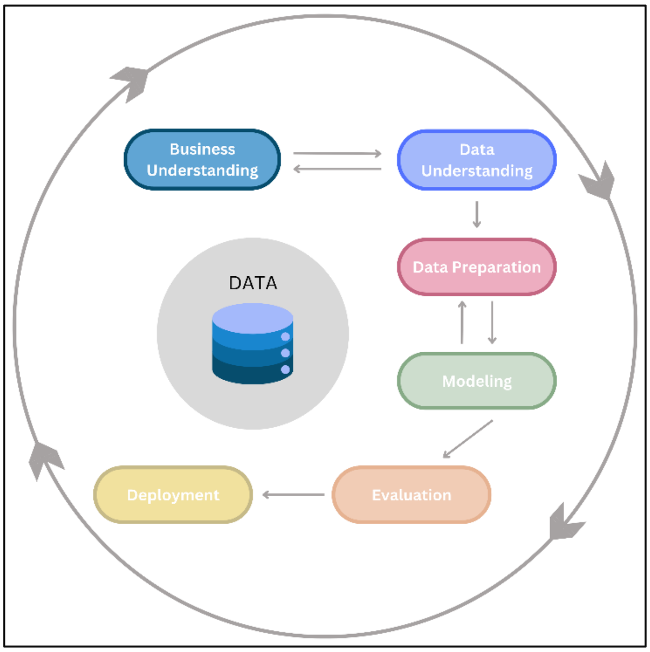

Figure 5 - CRISP-DM Stages (adapted from: [44]**)**

The process of moving between the Business Understanding and Data Understanding stages, as well as between the Data Preparation and Modeling stages, is fundamental to the success of data mining projects. Going back and forth between Business Understanding and Data Understanding ensures continuous alignment between the project goals and the characteristics of the data being analysed. This approach enables the integration of new insights, changes to project goals, and the refinement of data requirements. Moving between Data Preparation and Modeling, on the other hand, allows for iterative improvement of data cleansing, transformation, and feature engineering processes when new insights are discovered during the Modeling stage. This iterative approach guarantees that the analysis remains relevant, that the value obtained from the available data is optimized, and that the overall quality of the data mining project is improved.

In this section, we will examine the stages of the CRISP-DM adapted methodology and present them in the form of concise bullet points. These bullet points will explain each stage and provide a succinct summary of what was done.

- **Business Understanding:** Involves gaining a clear understanding of the research problem and defining the research objectives. We identified the key research questions, reviewed the existing literature, and conducted preliminary investigations to develop a comprehensive understanding of the problem domain. This helped to establish a solid foundation for subsequent stages by providing clarity on the research scope and objectives. We did a literature study and preliminary investigations to get a solid understanding of the area of cardiac rehabilitation and the challenges that doctors face when developing adapted rehabilitation prescriptions. The goal of the research was to address the need for better decision-making tools that use data-driven insights to improve patient outcomes in cardiac rehabilitation programmes, finding datasets that contained useful information about complications related to cardiac events, as this is an important factor in developing efficient rehabilitation strategies;

- **Data Understanding:** This stage focuses on acquiring and familiarizing oneself with the available data. It involves identifying relevant data sources, assessing data quality, and exploring the characteristics and structure of the dataset. During our research, we conducted a thorough search on Kaggle, a popular platform for datasets and data-related projects, with the intention of finding collaborators who had previously worked with similar datasets. However, despite our efforts, we did not find any collaborators or analyses specifically related to our dataset on Kaggle, even though the dataset was available there. While these absences limited our ability to leverage prior experiences and insights, the availability of the dataset itself still held significant value. It gave us an opportunity to independently explore and analyse the dataset to acquire a deeper understanding of its attributes and potential insights. By engaging in an extensive examination of the dataset, we aimed to uncover meaningful patterns and relationships that could contribute to our research objectives;

- **Data Preparation:** Process of preparing data for analysis. Data cleaning, feature engineering, and data transformation are all part of this step. This stage is a crucial step in ensuring the dataset is well-prepared for analysis. In order to try to find better results, we conducted thorough data cleaning procedures to enhance the quality and reliability of the data. These comprehensive data preparation steps were instrumental in ensuring the data was cleansed, transformed, and optimized for accurate and effective analysis in our thesis. This phase is iterative, as so, we went back and forth between data understanding, cleaning, and transformation to ensure the data was well-prepared for modeling. The basis for successful modeling and analysis in the succeeding phases of CRISP-DM is provided by a well-executed data preparation phase;

- **Modeling:** is the basis of the research process, during which we created and tested clustering models based on the prepared data. This stage consists of choosing appropriate modeling approaches, implementing the models, and optimizing their parameters. For that purpose, we conducted experiments with 4 different approaches explained in the Chapter 5 and three clustering algorithms: K-Means, K-Modes, and K-Prototypes. The main goal of this phase is to develop predictive models that can be used to make data-driven decisions and solve business problems effectively. It is essential to ensure that the models are accurate, reliable, and interpretable to gain insights and drive actionable results from the data. The modeling phase lays the foundation for the next phase, the evaluation phase, where the selected models are assessed in more detail before deployment;

- **Evaluation:** evaluates the produced models' performance and validity and can also reveal the model's strengths, flaws, and potential areas for improvement. To find these answers, we conducted a comprehensive evaluation of the K-Means clustering algorithm to assess the quality and effectiveness of its results. To evaluate the optimal number of clusters, we performed two widely used evaluation metrics: the Within-Cluster Sum of Squares (WCSS) and the Silhouette Coefficient. By analysing the WCSS values across different numbers of clusters, we could identify the number of clusters that resulted in the best overall compactness and separation. Following the selection of the optimal number of clusters, we proceeded to perform additional analyses to gain a deeper understanding of the cluster characteristics. Specifically, we conducted t-student and ANOVA tests to assess the statistical significance of differences between cluster groups in terms of the selected features. These tests allowed us to identify meaningful patterns and differences among the clusters, providing valuable insights for subsequent interpretation, helping to interpret and understand the cluster profiles;

- **Deployment:** validated models are applied to real-world scenarios or made available for additional analysis. During this phase, the models may be integrated into existing systems, user interfaces or APIs. In our case, after creating the clustering model, our next step was to implement the resulting decision support system on the 2ARTs digital platform. This deployment allowed healthcare professionals, specifically physicians, to gain valuable insights into the probable outcomes for each patient based on the clustering analysis. By accessing the system, doctors were empowered to make more informed decisions and modify their treatment prescriptions accordingly. The 2ART's platform provided a user-friendly interface that along with the DSS facilitated the possible outcomes of each patient, enhancing not only the efficiency and effectiveness of the CRP but also contributing to more personalized and targeted patient care.

It is important to note that the academic adaptation of CRISP-DM is an iterative process, where each phase informs and influences subsequent phases. We revisited earlier stages to refine the problem definition, explore additional data sources, and improve data pre-processing techniques based on the insights gained during later stages. This iterative nature allows for an evolving and refined research process.

4. 2ARTs Digital Platform

In this chapter, we discuss the creation and implementation of the 2ARTs digital platform, which serves as a monitoring tool for doctors participating in the cardiac rehab program. This system plays a crucial role in providing doctors with valuable insights into the progress and potential outcomes of their patients. To achieve this, a comprehensive set of steps was undertaken, building upon the methodology described in the previous chapter but delving into greater detail. This platform serves as the primary interface for doctors to access and analyse patient data. It offers a wide range of features that make the monitoring and assessment of patients' cardiac health easier throughout their rehabilitation journey.

The implementation of the platform involved consideration of the architecture and design. A user-friendly interface was developed, allowing doctors to navigate and access the desired information. The architecture of the platform was designed to ensure seamless integration with the CDSS, achieved through an API call, establishing a connection between the platform and the model developed for patient outcome prediction;

4.1. Architecture

In order to represent the architecture of our platform, we decided to use the C4 model[1] since it offers a clear and hierarchical visualization, allowing us to present the key components, relationships, and interactions of our system in a structured manner. This makes it easier for all stakeholders, including developers and project managers, to comprehend the platform's design and functionality, promoting effective communication by providing a common language for discussing architecture and design decisions. It helps us avoid ambiguities and misunderstandings, ensuring that everyone involved in the project is on the same page. This C4 model will serve as a valuable blueprint for the development, documentation, and communication of our platform.

In our project, we have chosen to focus on creating a level 2 C4 model, shown in Figure 6, because it provides the right level of detail and granularity for our specific needs. The level 2 of the C4 model allows us to represent the major containers and their interactions in our platform, offering an overview of the system's architecture.

[1] https://c4model.com

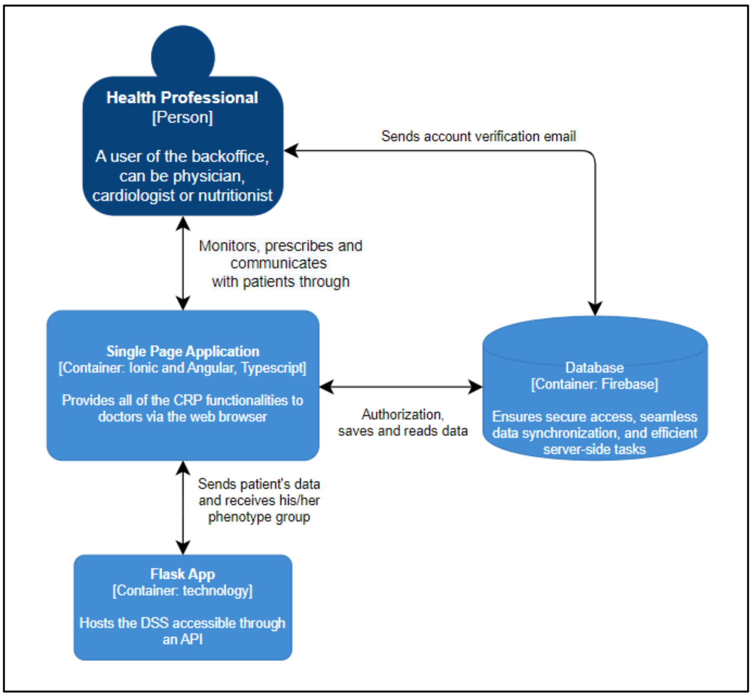

Figure 6 - 2Arts Backoffice Architecture

In our 2ARTs platform, the architecture is distributed as:

- **Major Containers:** Our platform consists of several major containers, including a single page application (SPA) for health professionals' use, a Firebase backend for user authorization, database and storage, and a Flask app for handling patient's data and providing an AP;
- **Container Relationships:** The SPA communicates with the Firebase backend to handle user authentication and store/retrieve user data. The Flask app acts as a data layer that interacts with the SPA for retrieving and sending data.
- **Key Technologies:** The SPA is built using modern front-end technologies like Ionic and Angular, while the Flask app uses Python and Flask for its implementation.

4.2. Technologies

The platform was developed using a combination of technologies including Ionic, Angular in Typescript and Firebase as we can see on Figure 7. Each of these technologies played a crucial role in different aspects of the platform's development and functionality.

An advantage of using Ionic was the ability to leverage Angular, a powerful TypeScript-based web application framework. Angular provided a solid foundation for building dynamic and responsive applications, and its seamless integration with Ionic made the development of complex features and functionalities easier. The combination of Ionic and Angular allowed us to create a scalable and maintainable codebase, ensuring that the platform could evolve and adapt to future requirements.

Additionally, using Firebase, a mobile and web application development platform, we had access to a range of services that enhanced the platform's functionality, security, and scalability, starting with the authentication service that was a pivotal factor in securing the platform as it allowed the implementation of user authentication and authorization mechanisms, ensuring that only authorized users could access the platform, helping in protect sensitive patient data and ensure privacy. The real-time database service was essential for the platform's dynamic nature, allowing the addition and retrieval of data in real-time. Doctors could store and access patient-related information, including biomedical data, prescriptions updates, and chat messages, ensuring that the data was always up to date and readily available. Finally, Firebase's cloud functions provided serverless computing capabilities, enabling the platform to scale easily and handle increased user demand. These functions allowed for the execution of server-side logic without the need to manage infrastructure. It enhanced the platform's performance and responsiveness, allowing us to perform actions based on certain triggers of the database.

The latest feature added to the platform focuses on achieving the primary goal of this work, which is to give doctors the ability to access predictions specific to each patient. This feature enhances the doctors' decision-making process by providing them with valuable insights that enable more accurate and individualized prescriptions that suit the unique needs and circumstances of each patient. To complete this feature, a predictive model was developed specifically to provide the desired predictions based on each patient's data. This model, explained in the following chapters, plays a crucial role in the decision support system by using advanced algorithms and machine learning techniques to analyse the patient's information and generate accurate predictions.

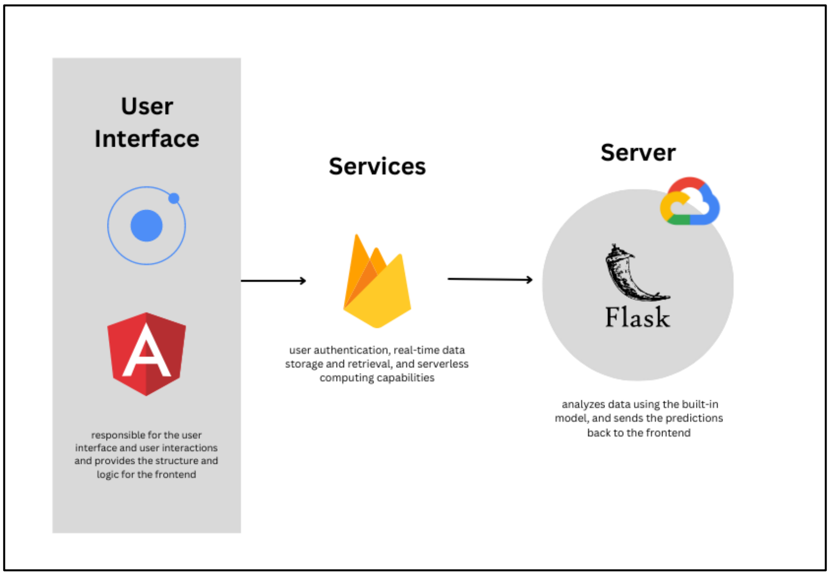

Figure 7 - Backoffice Technologies

4.3. User Interface Design

The platform’s UI was designed using Figma, a powerful web-based interface design tool as figured in Figure 8. Figma provides a collaborative environment that turns discussions and feedback sessions easier with doctors and other healthcare professionals. This collaborative approach ensured that the design of the platform effectively addressed the requirements and expectations of both patients and healthcare providers. Their valuable insights and expertise guided the design decisions, allowing for the creation of a platform that would cater to their specific workflows and enhance the rehabilitation program.

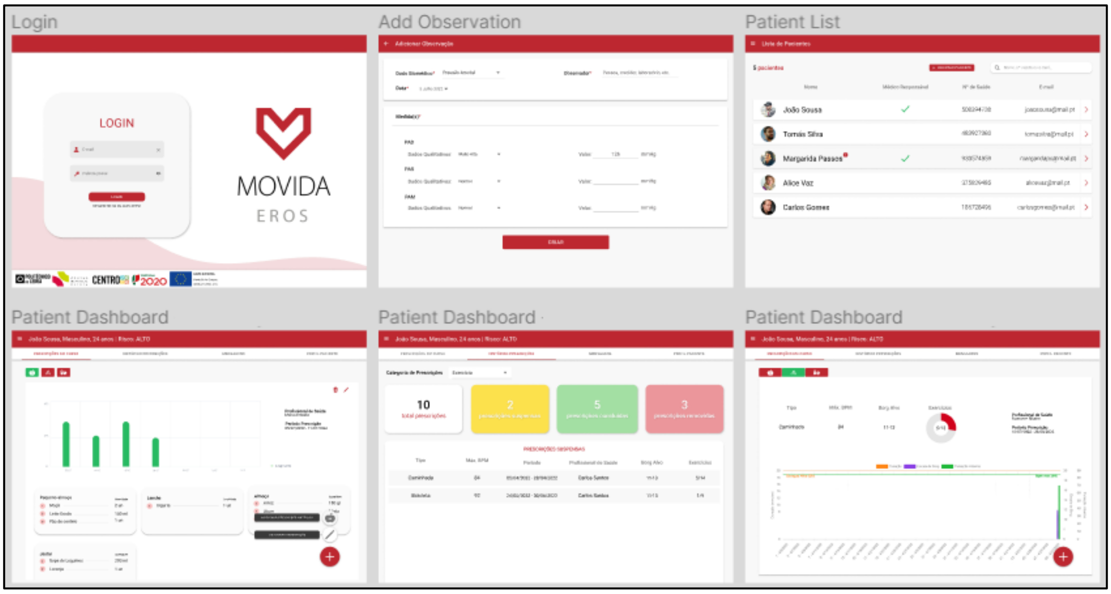

Figure 8 - Figma Prototype

Additionally, the design aims to strike a balance between functionality and aesthetics. While the platform needs to provide robust features and capabilities, it was also important to create a visually appealing interface that would engage users and create a positive user experience. The use of appropriate colour schemes, typography, and graphical elements helped in achieving a visually cohesive and pleasing design.

4.4. Software Features

This platform was created with a range of software features with the goal of allowing healthcare professionals to efficiently manage the cardiac rehabilitation of their patients. These features were developed to make communication easier, monitor patient progress, and support individualized care.

The multiple existent roles play crucial parts in the rehabilitation process. While the admin monitors the existent health professionals, patients and available biomedical data in the platform, the physician has the role to provide exercise prescriptions to their patients, ensuring that it is individualized and adapted to each patient’s needs. The nutritionist role focuses on providing personalized nutrition prescriptions, as well as seeing their history, as we can see in Figure 9, and monitoring patients' nutritional intake in collaboration with other doctors to ensure comprehensive care.

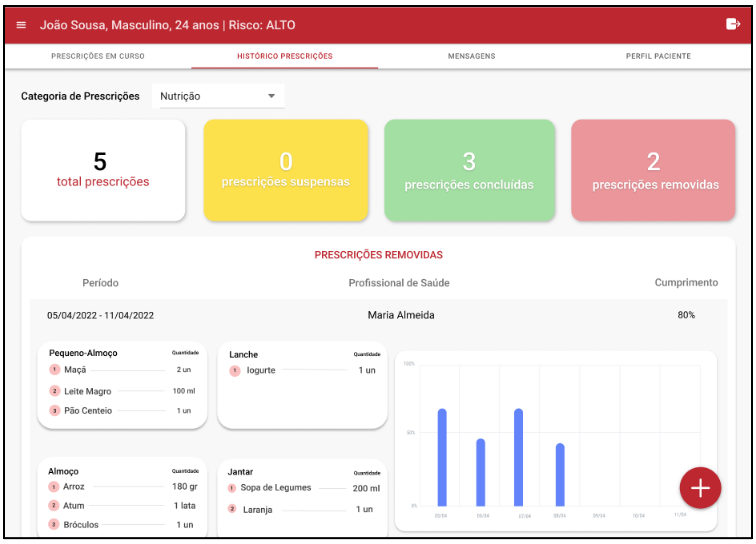

Figure 9 - History of Nutrition Prescriptions

The cardiologist role involves reviewing patients' medical data such as daily prescribed medication, presented in Figure 10 below, assessing cardiac health, and providing specialized recommendations or even insights through the group chat.

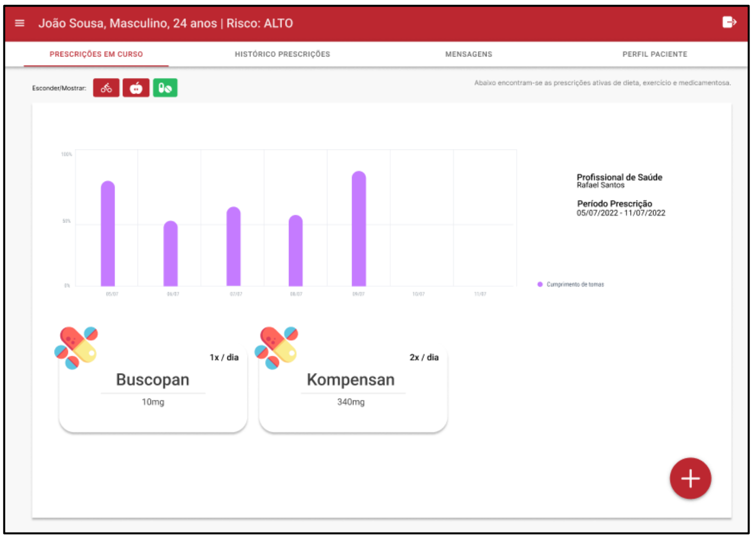

Figure 10 - Drug Prescription

All these health professional roles (physician, nutritionist and cardiologist) have the permission to add biomedical data to each of their patients, as well as see the history of every type of prescription already prescribed, suspended or completed. Besides working together on this platform, they can also communicate all together with the patient in a group chat, a feature that provides direct communication between patients and the entire healthcare team, ensuring that all doctors involved in a patient's rehabilitation program are updated with the latest information, fostering improved decision-making and collaborative care. Patients can interact with all doctors simultaneously, sharing their progress, asking questions, and receiving guidance from the entire healthcare team.

Besides this one, there is also a feature that aims to improve data management by dynamically adding biomedical data. This functionality enables doctors to collect and monitor a wide range of critical information related to patients' cardiac rehabilitation, enabling multiple input types of biomedical data into the platform's database as well as also receive it from the patient’s side. This includes medical test results such as electrocardiograms (ECGs) along with other diagnostic assessments and even body measurements such as weight, height, body mass index (BMI) or vital signs such as blood pressure, heart rate, and oxygen saturation levels which could also be logged in the system.

4.5. AI Model and Digital Platform Integration

It was essential to deploy the CDSS AI model in a cloud environment to make it accessible and useful in our digital platform, enabling it to be accessed via an API URL. We looked at several alternatives to accomplish this goal before deciding on a method that included moving the code tested in Google Colab to a Flask application and subsequently deploying this app on the Google Cloud platform.

Flask, a popular Python web framework, provided a suitable foundation for hosting the model as an API [45]. By encapsulating the model code within a Flask app, we created a server-side application that could handle incoming requests and provide predictions based on the trained model. The process started with familiarizing with Flask and transferring the Python model code from Google Colab to the main.py file in the Flask app. This code includes loading the trained machine learning model, performing the necessary data pre-processing explained before, and making predictions so we could expose the model as an API endpoint. To handle incoming requests, we defined Flask routes in the main.py file. Routes define the URLs that the app will respond to and the corresponding functions to execute, for example, for our study we defined a route that accepts an HTTP POST request with data to be classified using the model created before. In addition to the main.py file, we also included other necessary files in the Flask app's file structure like app.yaml, which provides configuration settings for deploying the app on Google Cloud Run, the Dockerfile that defines the instructions for building the Docker image of the app, the CSV file which is the dataset we need to read to create the model initially, and a requirements.txt file that lists the required Python packages for the app.

We tested the application on our local environment before deploying it to the cloud to assure its performance and functioning. During this testing process, we were able to find and fix any problems or bugs to make sure the application was running as planned and to finally deploy it to Google Cloud Run (Figure 11), once we were sure of its reliability and usability. Cloud Run is a serverless compute platform that allows to run containers in a fully managed environment [46]. By deploying the Flask app to Cloud Run, chosen for its ease of use, scalability and cost-effectiveness, we can easily access the API endpoint through a generated URL.

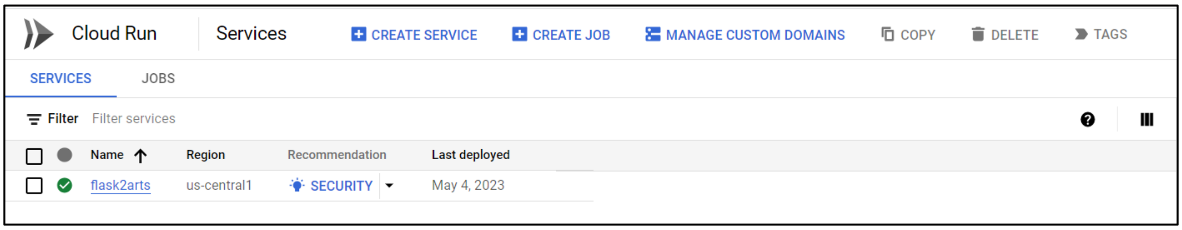

Figure 11 - Cloud Run

After deploying and testing the application, we receive data from the health platform, which is then inserted into our clustering model. This model helps classifying the new patient into a specific cluster or group based on her/his characteristics. The assigned cluster phenotype provides valuable information to the doctor, enabling the making of informed decisions and tailoring the patient's treatment plan accordingly. By understanding the patient's cluster phenotype, the doctor can proactively modify certain factors or interventions to mitigate the risk of future cardiac events and improve the patient's overall cardiovascular health. Ultimately, this feedback loop between the clustering model and the doctor helps in the personalized prevention and management of cardiac events, enhancing patient care and outcomes.

As shown in Figure 12, the doctor can easily navigate to a patient's profile and click on the statistics button, which triggers an automatic call to the route defined in the Flask app, starting the clustering model on the patient's data. Once the clustering analysis is complete, the doctor receives an alert that includes the name of a clinical phenotype assigned by the expert who collaborated on the project with us, allowing the doctor to make informed decisions and adjust the patient's prescriptions based on the assigned group, ensuring personalized and targeted care.

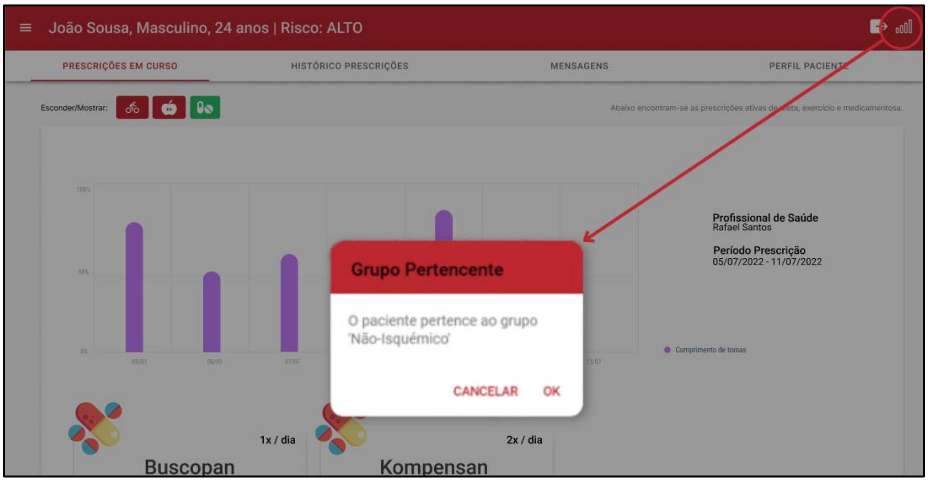

Figure 12 - Patient's predicted outcome in the platform

5. Development of the Classification Model

The CDSS, implemented as a Flask application, forms the core of the predictive analytics component. It leverages the clustering model developed to provide doctors with insights into the probable outcomes for each patient. Through the 2ARTs digital platforms interface, doctors can make API calls to retrieve the necessary predictions and recommendations. This integration empowers doctors to make informed decisions and adapt their treatment plans based on the insights provided by the system.

5.1. Dataset

During the research process, both the expert and the team diligently searched for the most appropriate dataset that would effectively address the objectives of our study. In this quest, the expert suggested the dataset, known as the Myocardial Infarction Complications Database. This recommendation played a crucial role in influencing our choice since his extensive knowledge and experience in the field allowed him to identify the dataset that best aligned with the requirements of our research.

The database under discussion includes extensive information from a group of 1700 patients. There are 123 features in all, 111 of which are related to the clinical phenotypes observed in the patients, and the remaining 12 aspects are probable complications of the relevant disease, namely myocardial infarction (MI). The detailed character of actual clinical data is shown by this rich dataset, which also brings special analytical challenges.

It should be noted that 7.6% of the data in this database is missing, requiring additional estimating and imputation methods to address these missing values properly [47]. In the field of cardiology and clinical research, the inclusion of such a thorough dataset is crucial as researchers can better understand each of the clinical phenotypes associated with heart failure and investigate potential consequences by using this data. These discoveries aid in the creation of more effective treatment plans as well as the classification of patient subgroups.

Table 2 provides an overview of the information contained in the dataset:

Table 2 - Dataset Information

Category	Number of Features	Description
Complication Predictions	111	Features describing all the clinical characteristics
Possible Complications	12	Features representing potential heart failure-related complications

5.2. Classification Model Testing in Google Colab

In the process of developing the CDSS for the cardiac rehabilitation program platform, testing of different algorithms and data cleaning functions was conducted using Google Colab, a cloud based Jupyter notebook environment, which provided a convenient and efficient platform for testing and experimenting with multiple machine learning algorithms as well as data cleaning techniques that were applied to ensure optimal model performance. Using these, we proceeded to determine the most suitable methods that would result in accurate and relevant results for the decision support system.

Unlike homogeneous datasets where all features belong to the same type, this one exhibits heterogeneity, comprising a combination of categorical, binary, and numerical features. So, we needed to be careful with the methods we used since many of them might be specific to the usual numeric feature.

5.2.1. Data Cleaning and Pre-processing

The data cleaning process played a crucial role in ensuring the quality and reliability of the dataset used for model training in this study. Our chosen dataset underwent multiple data cleaning techniques to address common data quality issues and improve the integrity of the data. In order to achieve this, before we went under testing the best methods, we met with the expert in the area of studying, to check the whole dataset and choose the features that weren't important for our problem, which lead to removing three features that were related to emergency events (e.g. inside an ambulance) and 12 that were features related to the outputs, which means post-event and thus not important for our main goal. After the deletion, we were still left with 108 features (excluding the ID column), so we proceeded to get the insights of the remaining columns, deciding what would be the best option for each one of them.

- **Handling Null Values:** The dataset was examined to identify columns with a significant number of missing or null values. Columns with more than 60% null values were deemed insufficient in providing meaningful insights and, thus, four features (IBS_NASL, KFK_BLOOD, S_AD_KBRIG, D_AD_KBRIG) were consequently removed from the dataset, leaving 104 features to work with. For the remaining columns with null values, appropriate imputation techniques were employed based on the data distribution. If the numerical feature presented a normal distribution, like shown in the Figure 13 (feature AGE), the mean was chosen to replace the null values, else if the distribution was skewed, also shown in Figure 13 (feature L_BLOOD), we used the median. In case there was null values in our categorical features, we opted to replace these values with their mode, ensuring that the imputed values aligned with the underlying data distribution [48].

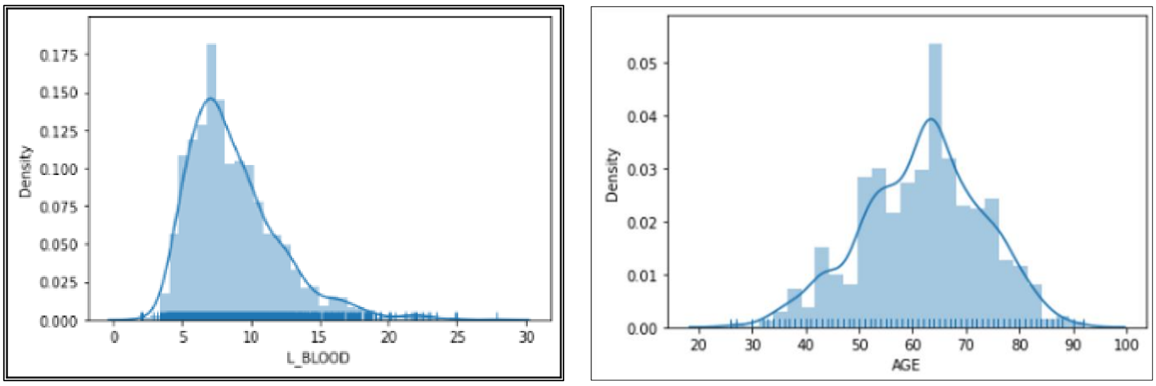

Figure 13 - Distribution of 'AGE' and 'L_BLOOD'

- **Correlation Matrix:** A correlation matrix was initially calculated to identify highly correlated features within the dataset, as we can see in Figure 14. This step aimed to detect and mitigate multicollinearity issues, where redundant or strongly correlated features (>0.7 or <-0.7) could potentially introduce bias or noise to the model. During the multiple tests conducted to identify the best methods for our end goal, we found that the use of Principal Component Analysis (PCA) gave beneficial outcomes. Instead of removing features with strong correlations, PCA treated these correlations and transformed the features into a new set of uncorrelated variables called principal components [49]. This approach allowed us to retain all features in the dataset while effectively reducing dimensionality and addressing multicollinearity issues. Therefore, PCA was chosen as a suitable method for our data cleaning process.

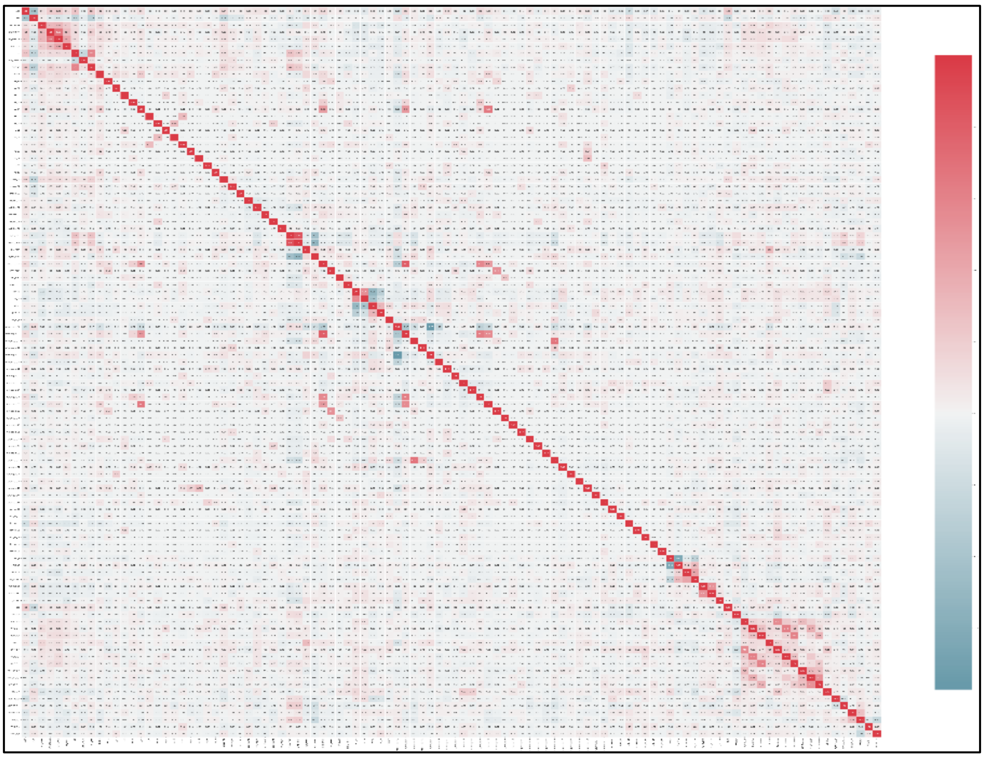

Figure 14 - Correlation Matrix

- **Scaling:** To ensure consistency and comparability among the numerical features in our mixed dataset, a scaling method that standardized the data by subtracting the mean and dividing by the standard deviation was employed. Scaling involves transforming the numerical values to a standardized scale, removing differences in magnitude, and allowing for fair comparisons across features. Doing this step eliminates the biases that may arise due to differences in measurement scales, ensuring that all feature types contribute equally to the modeling process. By bringing the numerical features to a common scale, we promote a more reliable and meaningful interpretation of their relationships and effects on the desired outcomes as well as facilitate the effective integration of different feature types in subsequent modeling steps. It ensures that the numerical features are comparable to categorical and binary features, enabling a comprehensive analysis that considers the diverse nature of the dataset.

- **One-Hot Encoding:** Between the 104 features we had to work on, besides numerical features, categorical features were also present. These features were encoded using one-hot encoding. This technique transforms categorical variables into binary variables, enabling the model (namely the chosen PCA method, which only accepts numerical features) to effectively interpret and use the categorical data. Each category within a categorical variable is represented as a separate binary variable, making proper feature representation easier and inclusion in the model.

The presence of mixed data leads to tailored analytical approaches, so, we must account for the distinct characteristics of each feature type during the analysis and modeling phases. Categorical features require techniques such as one-hot encoding, where numerical features can be subjected to traditional statistical analyses, including measures of central tendency or correlation.

To provide a clear overview of the feature distribution, Table 3 presents the breakdown of the dataset with the remaining 104 features by feature type:

Table 3 - Distribution of Features by Type

Feature Type	Number of Features
Categorical	17
Binary	78
Numerical	8

5.2.2. Modeling

In order to optimize the prescription of physical exercise and tailor it to individual patients, we must assign them to specific phenotype classes. However, our dataset lacks predefined classes for these phenotypes. To address this, we employed an unsupervised learning model, using clustering techniques to group patients into distinct clusters which are then evaluated by the expert to determine if they align with the above-mentioned phenotypes, enabling us to gain insights into the effectiveness of the exercise prescription process. By identifying patterns and similarities among patients, we can offer more informed and personalized exercise recommendations, ultimately enhancing patient outcomes and overall cardiac rehabilitation.

Given the heterogeneity of the dataset, it was imperative to employ multiple modelling approaches to identify the optimal solution and achieve the desired results. For that, we explored four distinct approaches to feature selection: Feature Selection Methods, Expert's Feature Selection, and feature reduction like PCA and Factorial Analysis of Mixed Data (FAMD) techniques. These methods allowed us to identify the most relevant features and dimensions for creating meaningful clusters within the dataset.

In the four approaches, we used common techniques to determine the optimal number of clusters and to later evaluate and compare all of the approaches together. The Elbow Method, which involves plotting the variance explained as a function of the number of clusters and where at the point at which adding more clusters does not significantly decrease the variance explained represents the optimal number of clusters and Silhouette Scores, which measures the compactness and separation of clusters, with higher scores indicating better-defined clusters, helped us identify the suitable number of clusters for each approach. Additionally, we assessed the quality of the clusters using evaluation metrics such as Within-Cluster Sum of Squares (WCSS) and Silhouette Coefficient. The WCSS metric measures the compactness and tightness of data points within each cluster, quantifying the sum of squared distances between each data point and its assigned cluster centroid. A lower WCSS value indicates more cohesive and well-defined clusters, helping us assess the effectiveness of the feature reduction and selection techniques in producing clusters with minimal within-cluster variability. The Silhouette Coefficient, on the other hand, assesses both the cohesion and separation of clusters. It measures the extent to which each data point is similar to its own cluster compared to other clusters, ranging from -1 to 1, where higher values indicate well-separated and distinct clusters, while lower values suggest overlapping or poorly separated clusters. To ensure the accuracy and validity of the clustering results, we also collaborated with the expert who reviewed the clustering outcomes and assessed the accuracy and interpretability of each cluster. His expertise and domain knowledge were instrumental in validating the clustering results and providing valuable insights into the relevance and meaningfulness of the obtained clusters.

The following sections discuss the various approaches tried during this modeling step.

5.2.3. Feature Reduction

Two feature reduction techniques, namely Principal Component Analysis (PCA) and Factor Analysis of Mixed Data (FAMD), were employed. These techniques aim to reduce the dimensionality of the dataset while preserving the essential information contained within the features. This reduction in dimensionality aids in simplifying subsequent modelling processes and can improve the efficiency and interpretability of the results.

i. PCA

We proceeded to explore the use of Principal Component Analysis (PCA) as a dimensionality reduction technique after completing all the necessary pre-processing processes, which included the removal of correlated variables, one-hot encoding, and scaling. By converting the original features into a collection of orthogonal components, PCA enables us to extract the dataset's most important information.

Before proceeding, we calculated the Cumulative Explained Variance to decide how many principal components to keep. To determine the point at which adding more components would result in diminishing returns in terms of capturing the dataset's variability, we evaluated the cumulative explained variance results. Based on our analysis, the first principal component accounted for 28,3% of the variance, the second accounted for 54.7%, and the third accounted for 79.9%. To capture at least 60% of the variance [50], we chose to retain only two principal components. This decision allowed us to gather a substantial amount of the dataset's variability, surpassing the minimum threshold we set and to successfully perform PCA, where the resulting output allowed us to visualize the dataset in a reduced-dimensional space.

To provide a preview of the results, the plot below in Figure 15 displays the two principal components obtained from the PCA analysis:

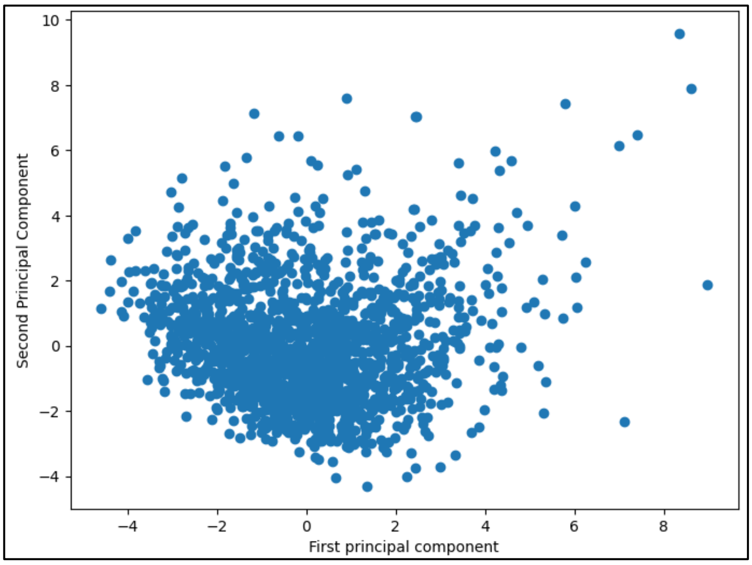

Figure 15 - Scatter Plot of the PCA

Before proceeding with the application of the clustering algorithms, we needed to determine the optimal number of clusters to consider. To achieve this, we employed the two commonly used methods mentioned before: the Elbow method and Silhouette scores, both based on the results obtained from the PCA. The first method helped identify the point that the reduction in within-cluster sum of squares (WCSS) significantly diminished, where the silhouette scores assessed cluster cohesion and separation. By considering both methods, we determined that the ideal number of clusters would be three based on the plot in Figure 16 and Table 4.

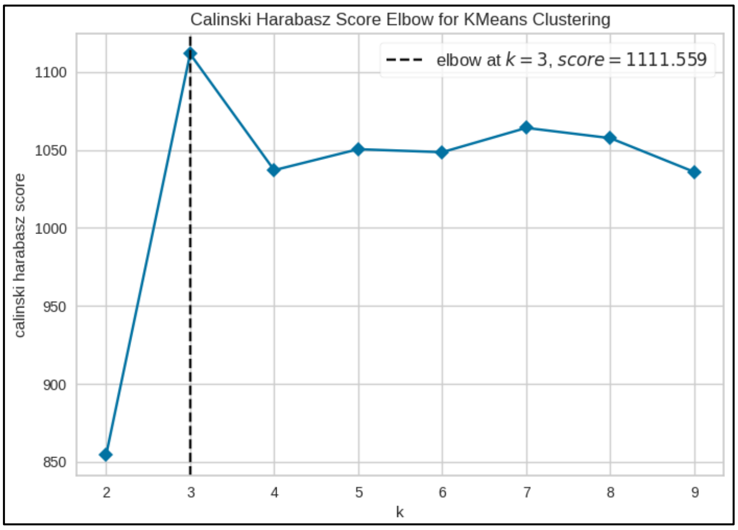

Figure 16 - Elbow Method with PCA

Table 4 - Silhouette Scores with PCA

Number of Clusters	Silhouette Score
2	0.337529
3	0.377201
4	0.341437
5	0.347945
6	0.350231
7	0.340203

In order to select the most suitable clustering algorithm for our analysis, we compared two prominent algorithms: K-Means and K-Prototypes, leaving K-Modes behind since it is only suitable for categorical data. While K-Prototypes is specifically designed for mixed data, we also explored K-means as it is widely used and has proven effective in various scenarios.

Given the nature of our dataset, which consists of a combination of categorical and numerical features, K-Prototypes appeared to be the most suitable choice. K-prototypes combines the strengths of K-Means for numerical data and K-Modes for categorical data, allowing it to handle mixed data appropriately. This algorithm considers the dissimilarity between data points based on their numerical and categorical attributes, offering more flexibility in capturing the underlying patterns within the dataset. However, during the data pre-processing phase, we encountered a challenge when applying principal component analysis (PCA) to reduce dimensionality, as explained before. To use PCA, we had to transform our categorical features into numerical ones through one-hot encoding, resulting in the loss of the original categorical features, as they were converted into binary numerical representations.

As a result, we realized that the K-Prototypes algorithm would not be the best choice for our current situation, as it relies on the presence of categorical features. Without the original categorical features, the effectiveness and interpretability of the K-Prototypes algorithm would be compromised. Given this limitation, we had to reconsider our options and decided to proceed with the K-Means algorithm. With the optimal number of clusters identified before, we applied the K-Means algorithm and visualized the results as in the Figure 17, where each data point is colored and labeled according to the cluster it belongs to as determined by the clustering and the axis is represented by the two principal components.

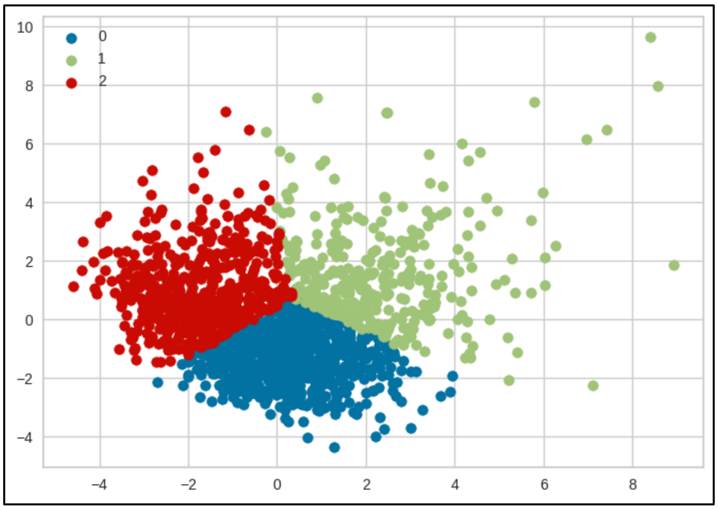

Figure 17 - Clustering Analysis with PCA

To measure how well this approach behaves when comparing with the others, we performed the evaluation metrics WCSS and Silhouette Coefficient which gave the scores of 4754.24 and 0.3772, respectively.

ii. FAMD

After testing PCA, we decided to explore Factor Analysis of Mixed Data (FAMD) one of the methods for handling mixed data, which includes both categorical and numerical variables. FAMD is specifically designed to handle mixed data and offers advantages over traditional PCA when dealing with such datasets. This method already scales the numerical data, so, in this approach we skipped the scaling and the one hot encoding steps because FAMD already performs one-hot encoding to represent the categorical data numerically [51]. This is necessary because this method is based on factor analysis, which requires numerical inputs. In our case, we have binary features that can be considered categorical since they don't have any distance or meaningful order between their options. However, when these binary features are encoded using one-hot encoding, it can create duplicate entries in the dataset which are redundant and perfectly correlated negatively with each other due to the one-hot encoding process. To avoid this redundancy and correlation, we can treat these binary features as numerical instead of categorical.

To determine the optimal number of components to retain in FAMD, we first examined the percentage of (cumulative) variance explained by each component. We conducted FAMD with a maximum of 10 components and calculated the cumulative percentage of variance explained, as seen in Table 5. To strike a balance between dimensionality reduction and retaining meaningful information, we decided to select only those components that accounted for at least 80% of the data variance, and after careful consideration, we ended up retaining seven components.

Table 5 - FAMD % of Variance

Component	% of variance (cumulative)
0	4.78%
1	8.78%
2	11.62%
3	13.92%
4	16.10%
5	18.23%
6	20.29%
7	22.29%
8	24.21%
9	26.11%

To determine the optimal number of clusters within the FAMD-transformed data, we also employed the two techniques used before: the Elbow Method as shown in Figure 18 and the Silhouette Score shown in Table 6. By applying these methods to the FAMD-transformed data, we were able to identify that the optimal number of clusters was two, proceeding to apply the K-Means clustering algorithm.

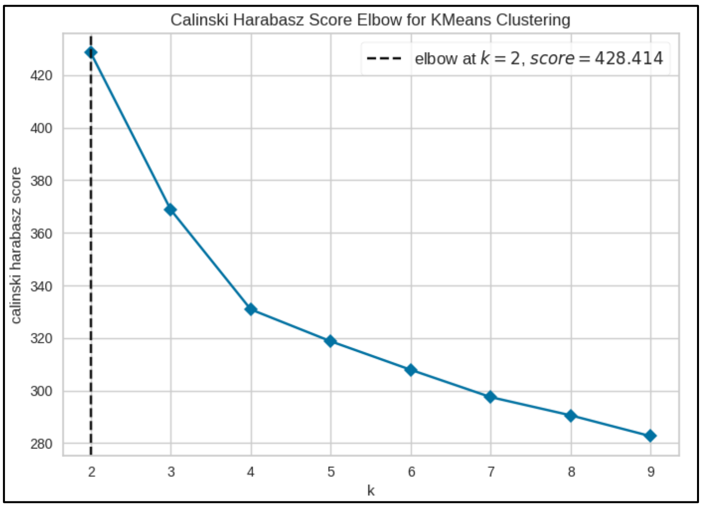

Figure 18 - Elbow Method FAMD

Table 6 - Silhouette Score FAMD

Number of Clusters	Silhouette Score
2	0.225
3	0.194
4	0.208
5	0.211
6	0.214

In this approach we do not have a scatter plot of the clustering as shown in the PCA approach since in this case we have more than two features (seven) and it would not be possible to scatter the seven features together. We evaluated the quality of the clustering results using both evaluation metrics also mentioned before: the WCSS and the Silhouette Coefficient, whereas the results were 45060.39 and 0.2255, respectively.

5.2.4. Feature Selection

To mitigate the potential impact of irrelevant or redundant features, two feature selection approaches were also tested. The first approach involved using methods like Variance Threshold, whereas the second involved the input of the domain expert. The expert was consulted to provide insights and knowledge in order to select a subset of features that were deemed highly relevant for the analysis. The expert's input allowed a more focused examination of features that held significant meaning within the context of the study.

i. Feature Selection Methods

Initially, several feature selection techniques were explored, including Random Forest, SelectKBest, Recursive Feature Elimination, Mutual Information and SelectFromModel. However, these techniques typically require the presence of an output variable to guide the feature selection process. In our dataset, we did have features related to patient outcomes, specifically the lethal outcome of each patient. However, using this output feature for feature selection would not align with the primary research objective of our study. Our aim was not to predict or analyze the lethal outcomes of patients, but rather to investigate and understand the resulting clusters in order to give them a phenotype to label the patients for doctors to be able to give a more individual prescription. Given this deviation from the conventional feature selection approaches, an alternative method like Variance Threshold was explored to address the unique requirements of our study.

A VarianceThreshold method was instantiated with a threshold of 0.25, indicating that any feature with a variance below this value would be considered for removal. By applying this threshold, we ensured that the final dataset contained only the selected features, present in Table 7, that exhibited sufficient variability.

Table 7 - Variance Threshold Feature Selection

Feature	Description
AGE	Age of the patient
K_BLOOD	Serum potassium content
NA_BLOOD	Serum sodium content
ALT_BLOOD	Serum Alanine aminotransferase content
AST_BLOOD	Serum Aspartate aminotransferase content
L_BLOOD	White blood cell count
ROE	Erythrocyte sedimentation rate
D_AD_ORIT	Diastolic blood pressure
INF_ANAM	Quantity of myocardial infarctions in the anamnesi
STENOK_AN	Exertional angina pectoris in the anamnesis
GB	Presence of essential hypertension

DLIT_AG	Duration of arterial hypertension
ZSN_A	Presence of CHF in the anamnesis
TIME_B_S	Time between CHD attack and hospital admission
R_AB_1_n	Relapse of pain in the first hours of hospital
NA_R_1_n	Use of opioid drugs in the first hours of hospital
NOT_NA_1_n	Use of NSAIDs in the first hours of hospital

Following the feature selection process, we sought to determine the optimal number of clusters for our analysis with the techniques Elbow Method, Silhouette Score, and Gap Statistics, which measures the statistical significance of clustering results compared to random data, to make sure we chose the optimal number of clusters. These metrics helped assess the quality and distinctiveness of clustering solutions.

In our case, the optimal number found with the Elbow Method was two clusters, whereas with the Silhouette Score the number was 2 and with the Gap Statistics was 4, shown in Figure 19. However, due to the nature of our problem and the need to accurately identify the phenotypes of patient clusters, we, along with the expert, prioritized having a number of clusters between 3 and 6. Considering these criteria, we selected 4 as the number of clusters for our analysis.

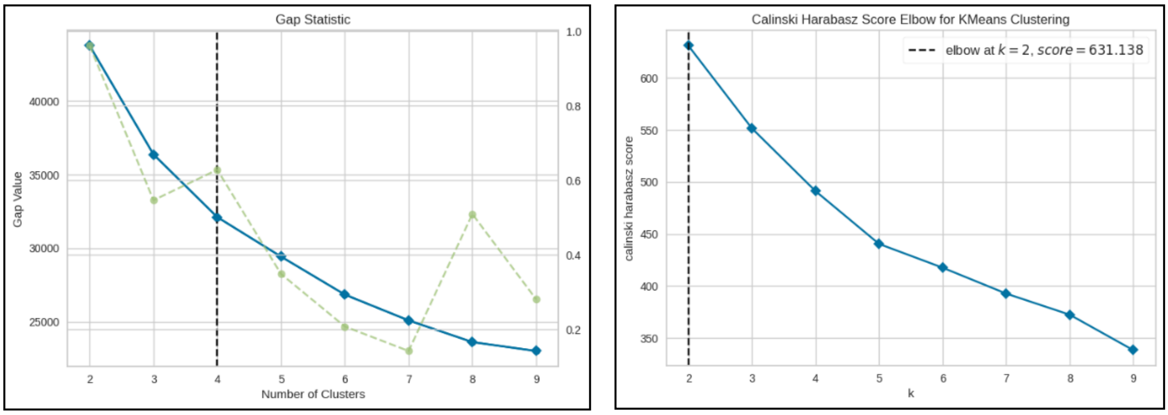

Figure 19 - Gap Statistic and Elbow Method (Variance Threshold)

Following the clustering analysis using the K-Means algorithm with four clusters, we assessed the quality of the clustering results using the WCSS and the Silhouette Coefficient, giving the scores of 538953.114 and 0.2280, respectively.

ii. Expert's Feature Selection

To explore alternative approaches in search of the best results for our problem, we collaborated with the expert to select a subset of features that were deemed important for achieving our goal. Out of the initial 108 features, we narrowed down the selection to 16 based on their relevance and potential impact. However, upon closer examination, we discovered that three of the selected features contained over 60% null values, leading us to the decision of excluding these features (S_AD_KBRID, D_AD_KBRIG, KFK_BLOOD) from further analysis because of the potential bias and limitations associated with missing data, being left with a reduced set of 13 features to work with, as presented in Figure 20 , where six were categorical (ordinal and binary) in nature.

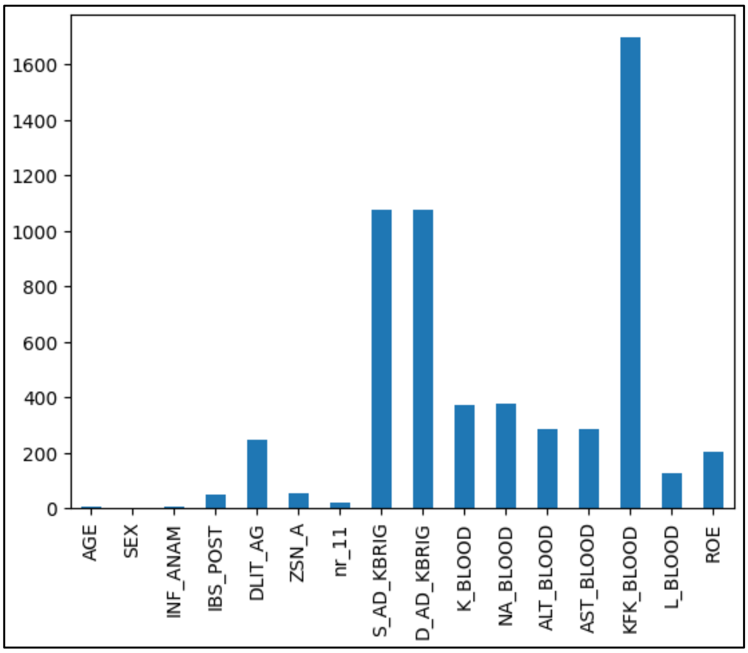

Figure 20 - Null Count on the 16 features

To prepare these features for subsequent analysis, we applied one-hot encoding, as explained earlier, to the categorical ones that exhibited an ordinal nature and lacked any inherent distance or numerical relationship between the available options, which in this approach is equivalent to only doing it on the IBS_POST feature, as shown in Figure 21, by converting it into binary form, ensuring that each distinct category was represented by its own binary column. On the other hand, for the features in our dataset that were already represented in binary form, containing values of either 0 or 1, or for the ordinal features where the values exhibited a meaningful distance or numerical relationship between them, we made the decision to retain their original format. This choice was based on the understanding that these particular features were already appropriately encoded, and their values could be treated as either binary or numerical with inherent distances between the available options.

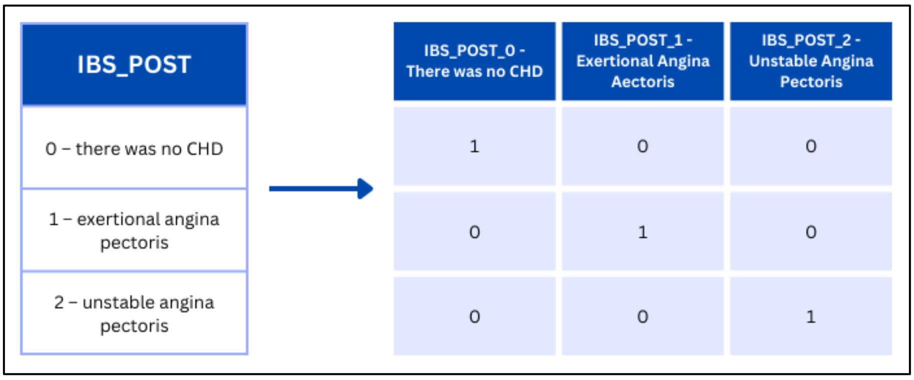

Figure 21 - One Hot Encoding on Categorical Feature

In preparation for applying the K-Means algorithm, we recognized the importance of feature reduction to enhance the performance and efficiency of the clustering process. Considering various dimensionality reduction techniques such as PCA (Principal Component Analysis), MCA (Multiple Correspondence Analysis), and FAMD (Factor Analysis of Mixed Data), we carefully evaluated their suitability for our dataset.

While MCA appeared to be a viable option, we observed that it was specifically designed for categorical features. As our dataset contained a combination of categorical and numerical features, we concluded that MCA would not be the most appropriate choice for our feature reduction needs. We had previously used FAMD, as described above, but found that compared with the results when using PCA, the latter would be our and the expertise best choice.

While the use of one-hot encoding in conjunction with Principal Component Analysis (PCA) may be a subject of debate, we considered various perspectives. Some experts argue against combining one-hot encoding with PCA due to the potential loss of interpretability and sparsity. However, others claim that it can be a viable approach depending on the specific context and objectives of the analysis. In our case, after consideration and weighing the available options, we made the decision to proceed with one-hot encoding of the categorical features while employing PCA. This allowed us to capture the underlying variability and structure of the data while accommodating the categorical nature of the selected features.

After examination of the components of PCA, we determined that selecting three components would account for approximately 60% of the total variance, as explained in approaches done before. To finally determine the ideal number of clusters for the K-Means algorithm with the resulting components, we conducted an in-depth analysis, employing the already explained techniques such as the Elbow method and Silhouette score and getting three as the optimal number of clusters, as plotted in Figure 22.

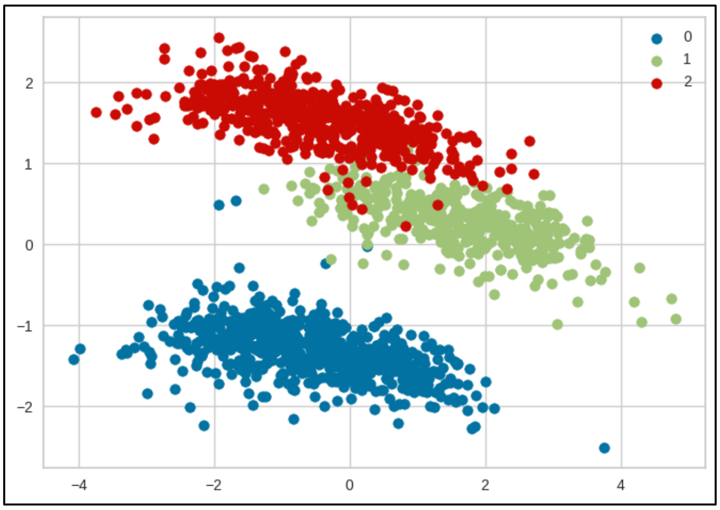

Figure 22 - Clustering Analysis on Expert's Feature Selection

Lastly and like we did in the approaches before, we performed the evaluation metrics WCSS and Silhouette Coefficient which gave the scores of 6230.614 and 0.1764, respectively.

6. Results

In this Chapter, we discuss the results and outcomes obtained from the CDSS. We focus on the analysis of the resulting clusters and their phenotype identification and providing a thorough analysis of the results gathered from the different modelling techniques used, including feature selection, feature reduction, clustering, and expert analysis. The goal is to evaluate the model's performance and efficacy in achieving the study's stated research goals. This includes detailed discussions on the performance metrics employed, such as the within-cluster sum of squares (WCSS), silhouette coefficients, and expert evaluations. The results are discussed considering the research objectives and their implications for the field of study, emphasizing the significance of the findings and their potential implications for real-life applications within the specific domain. These contribute to the existing body of knowledge in the field, providing the way for further advancements in CDSSs, treatment strategies, and subtyping approaches.

6.1. WCSS and Silhouette Coefficient Performances

As previously discussed in the preceding chapter, we employed the WCSS and Silhouette Coefficient metrics in each approach to assess and analyse the clustering performance. These metrics offer valuable insights into the quality and cohesion of the resulting clusters obtained from the various techniques used. To provide an overview and facilitate a comparative analysis, we will present the results again in this section to summarize the outcomes of each approach, as we can see in Table 8.

Table 8 - ML Approaches Performance

Approach	WCSS	Silhouette Coefficient
PCA	4754.24	0.3772
FAMD	45060.39	0.2255
Feature Selection Methods	538953.114	0.2280
Expert's Feature Selection	6230.614	0.1764

The results' analysis turns evident that the PCA approach has yielded the best results in terms of both WCSS and Silhouette Coefficient. This approach effectively reduces the dimensionality of the feature space, capturing the most important information and retaining the underlying structure of the data, leading to more compact and well-defined clusters, as reflected by the lower WCSS values. Additionally, the higher Silhouette Coefficient suggests that the clusters obtained through PCA are well-separated and distinct. Following closely behind PCA in terms of performance is the feature selection approach implemented by the expert, which involved selecting a subset of features based on domain knowledge and expertise. Although the feature selection approach did not achieve the same level of improvement as PCA, it still resulted in lower WCSS values compared to the other approaches. This indicates that the expert's insights and selection of relevant features contributed to the formation of more coherent and distinguishable clusters.

After this analysis, we recognized the importance of conducting a thorough expert analysis to complement our findings, so, in collaboration with the expert in the field, a detailed examination of the results generated by the two best approaches was conducted: PCA and the Expert's Feature Selection.

6.2. Performance of PCA

Starting with the first approach that revealed the best results in the Table 8 above, in order to gain deeper insights into the clusters produced, we employed statistical analyses like t-tests on the numerical features. This involved comparing the means of the numerical variables between different clusters, providing valuable information regarding the significance of differences in means, and we summarized the results in tables that included the computed p-values. The p-values allowed us to assess the probability of observing differences in the means by chance alone. By considering both the magnitude of the differences and the statistical significance, we were able to identify meaningful variations in the numerical features across clusters. Among these variables, three stood out as having statistically observed differences between the three standard clusters since they had a p-value below 0.05 (significance level), as we can see in Table 9: age (AGE), Diastolic Blood Pressure in the ICU (D_AD_ORIT), white blood cell count (L_BLOOD) and Alanine Aminotransferase (ALT_BLOOD), highlighting the importance of them in characterizing and distinguishing the phenotypes of each group.

Table 9 - ANOVA table - PCA

Feature	Cluster #1	Cluster #2	p-value
AGE	0	1	0,00
AGE	0	2	0,00
AGE	1	2	0,00
K_BLOOD	0	1	0,00
K_BLOOD	0	2	0,00
K_BLOOD	1	2	0,32
NA_BLOOD	0	1	0,00
NA_BLOOD	0	2	0,00
NA_BLOOD	1	2	0,99
ALT_BLOOD	0	1	0,00
ALT_BLOOD	0	2	0,00
ALT_BLOOD	1	2	0,01
AST_BLOOD	0	1	0,00
AST_BLOOD	0	2	0,00
AST_BLOOD	1	2	0,08
L_BLOOD	0	1	0,00
L_BLOOD	0	2	0,00
L_BLOOD	1	2	0,01

ROE	0	1	0,00
ROE	0	2	0,17
ROE	1	2	0,00
S_AD_ORIT	0	1	0,00
S_AD_ORIT	0	2	0,00
S_AD_ORIT	1	2	0,31
D_AD_ORIT	0	1	0,00
D_AD_ORIT	0	2	0,00
D_AD_ORIT	1	2	0,01

Among the remaining quantitative variables examined, the analysis did not yield sufficient evidence to reject the equality between means when comparing clusters 1 and 2. Specifically, variables K_BLOOD, NA_BLOOD, AST_BLOOD and S_AD_ORIT did not exhibit statistically significant differences between these two clusters. This finding suggests that, in terms of these variables, clusters 1 and 2 share similar mean values and may not represent distinct phenotypic profiles. However, it is important to note that this lack of significant differences between clusters 1 and 2 does not imply that these variables are irrelevant for distinguishing between clusters or understanding the studied condition. It simply suggests that, in terms of the mean values of these specific variables, clusters 1 and 2 do not exhibit statistically significant disparities, implying that the associated p-values for the comparisons between these clusters were greater than the significance level (0.05) as it is shown in Table 9.

For the categorical features, we used bar charts to see the frequency distribution of each category within the clusters, as we can see in the examples of Figure 23. By comparing the proportions of different categories among the clusters, we can identify any noticeable trends or discrepancies that provided valuable insights into the underlying patterns and associations.

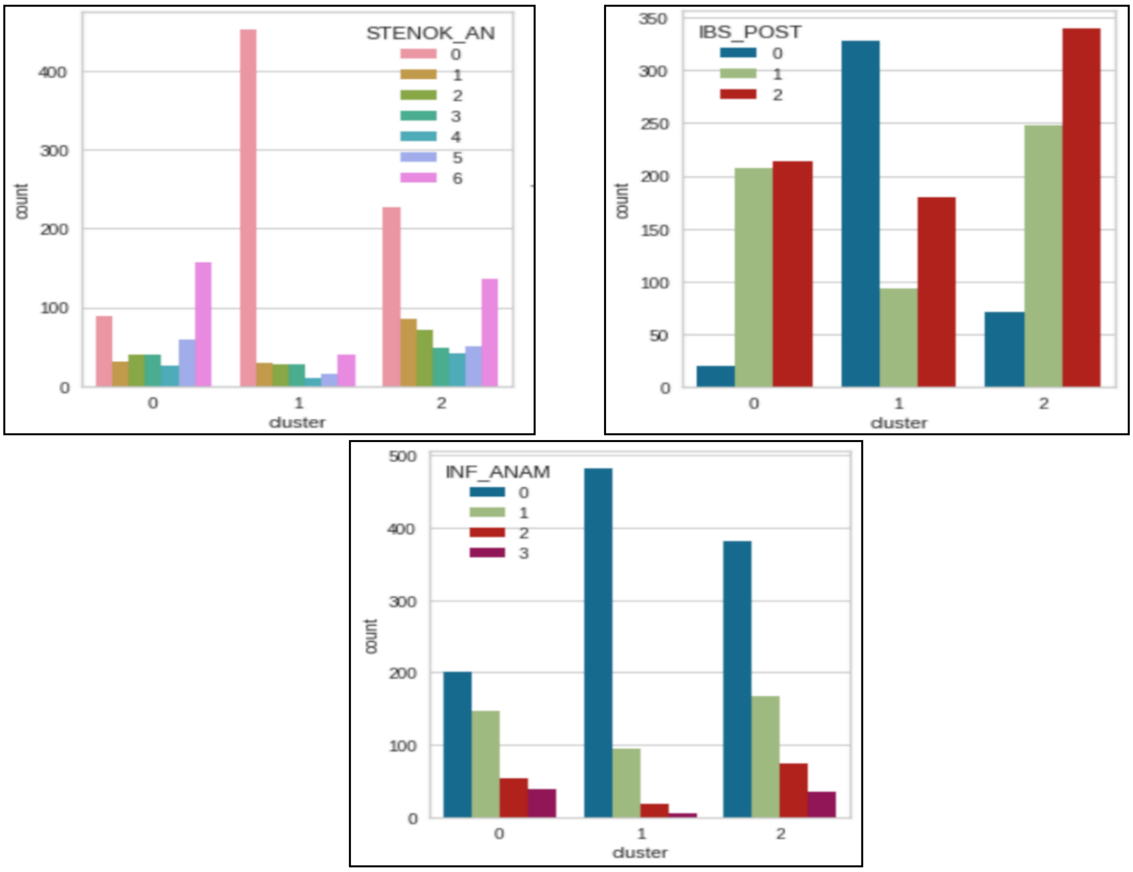

Figure 23 - Bar Chart Categorical Features in PCA

The analysis of the data reveals that patients who do not experience angina on exertion (IBS_POST) are predominantly found in cluster 1, which also contains a larger proportion of patients without a history of infarction before admission (INF_ANAM) and those who have never reported symptoms of angina pectoris (STENOK_AN). Additionally, cluster 1 tends to consist of younger patients. Based on these observations, the expert could draw the conclusion that cluster 1 could be identified as the "Non-Ischemic" group from a phenotypic perspective. This cluster exhibits distinct characteristics compared to the other clusters, suggesting a different underlying physiological profile associated with coronary artery disease.

On the other hand, cluster 0 predominantly consists of patients who exhibit symptoms of pulmonary edema upon admission. It is within this cluster that a higher proportion of patients admitted to the ICU with cardiogenic shock (K_SH_POST) are found. These observations suggest a clinical scenario characterized by a disrupted flow of blood, resulting in blood volume accumulation outside the heart. Moreover, cluster 0 includes the oldest patients among the clusters. Considering these characteristics, the phenotypic profile that emerges from this analysis aligns with a "Congestion" phenotype from the expert perspective.

Finally, cluster 2 encompasses patients of intermediate age who exhibit a higher prevalence of previous infarctions and greater severity of Coronary Artery Disease (CAD), specifically in the form of unstable angina, as shown in Figure 23.

These distinctive characteristics strongly suggest an "Ischemic" phenotype for this cluster based on expert analysis. The findings emphasize the significance of age, prior cardiac events, and CAD severity in understanding the unique profile of patients within this cluster.

To assess the quality and accuracy of the clusters, we also sought the expert's input and requested their evaluation of a representative sample from each cluster, ensuring a minimum of 10% coverage for each cluster. The expert was asked to determine whether the samples were appropriately positioned within their respective clusters. By analyzing these expert evaluations, we were able to assign an accuracy measure to each cluster. The results are shown in Table 10. These accuracy measures provide valuable insights into the performance and reliability of the clustering algorithm, validating the effectiveness of the clustering approach in capturing meaningful patterns and grouping similar samples together.

Table 10 - Accuracy of each cluster - PCA

Cluster	Accuracy
0	91%
1	71%
2	91%

6.3.Performance of Expert's Feature Selection

To gain deeper insights and expert input, we employed box plots to examine the distribution and variation of numerical features, allowing a better understanding of how the clusters differ in terms of these variables and bar charts to visualize the distribution of categorical features. Besides these visualization methods, we also made a t-student table for this approach in order to evaluate the statistical significance of the observed differences between cluster means in a methodical way so we wouldn't miss any important information when analysing the feature selection clustering.

The figures presented in Figure 24, display box plots of the features 'AGE' and 'ROW' that showed statistically significant differences between pairs of clusters. These differences have been determined through the application of a t-student test, as done before and shown in Table 11 for this approach, where a p-value lower than 0.05 indicates a significant distinction.

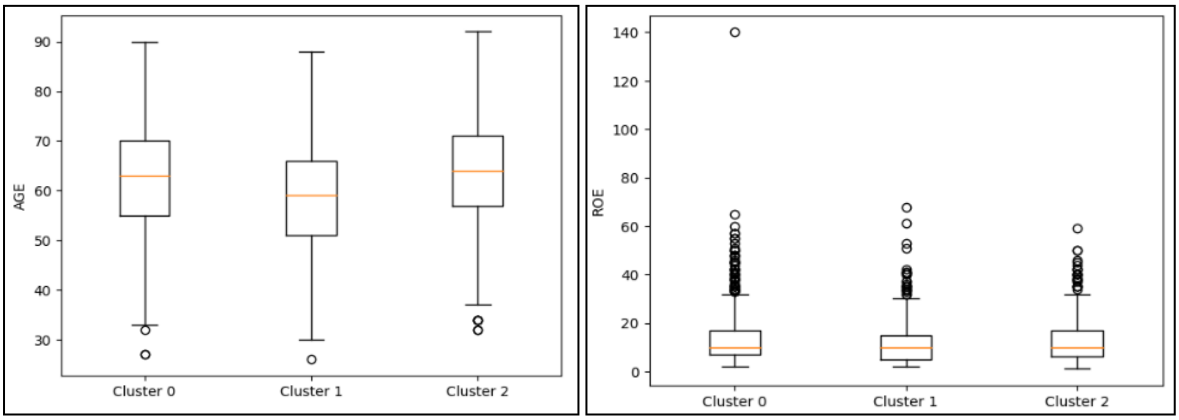

Figure 24 - Box Plots of Numerical Features – Expert's Feature Selection

When examining the results obtained from the t-student table, it was found that the approach PCA did not yield satisfactory outcomes compared to the approach of Expert's Feature Selection. In the t-student table of the latter approach, it was observed in Table 11 that most of the pairs for each feature presented high p-values, implying that the statistical relationship between the features and clusters was not strong. In contrast, in the t-student table of the approach PCA as we can see in Table 9, even if a feature had a high p-value, it would only be significant in one of the pairs and not across all pairs.

Based on these findings, we concluded that the approach PCA provided more meaningful and interpretable results when comparing each pair of cluster. The t-student table of the approach Expert's Feature Selection did not yield desirable outcomes as most of the pairs showed high p-values, indicating weak statistical significance, as we can see in Table 11.

Table 11 - ANOVA table – Expert's Feature Selection

Feature	Cluster #1	Cluster #2	p-value
AGE	0	1	0,00
AGE	0	2	0,03
AGE	1	2	0,00
K_BLOOD	0	1	0,49
K_BLOOD	0	2	0,15
K_BLOOD	1	2	0,66
NA_BLOOD	0	1	0,21
NA_BLOOD	0	2	0,22
NA_BLOOD	1	2	0,81
ALT_BLOOD	0	1	0,09
ALT_BLOOD	0	2	0,69
ALT_BLOOD	1	2	0,24
AST_BLOOD	0	1	0,04
AST_BLOOD	0	2	0,68
AST_BLOOD	1	2	0,03
L_BLOOD	0	1	0,00
L_BLOOD	0	2	0,23
L_BLOOD	1	2	0,01
ROE	0	1	0,00

ROE	0	2	0,02
ROE	1	2	0,04

During the analysis of the categorical features, we observed significant differentiation among the clusters, particularly in the features 'INF_ANAM' and 'IBS_POST' as we can see in Figure 25. Upon consulting with the domain expert, it was confirmed that these features play a crucial role in separating the patients based on their medical history. In particular, the variable 'IBS_POST' represents the presence or absence of symptoms related to angina pectoris weeks or days before admission to the hospital. The expert's analysis further supported the notion that patients within different clusters could be distinguished primarily based on their history of experiencing angina pectoris symptoms.

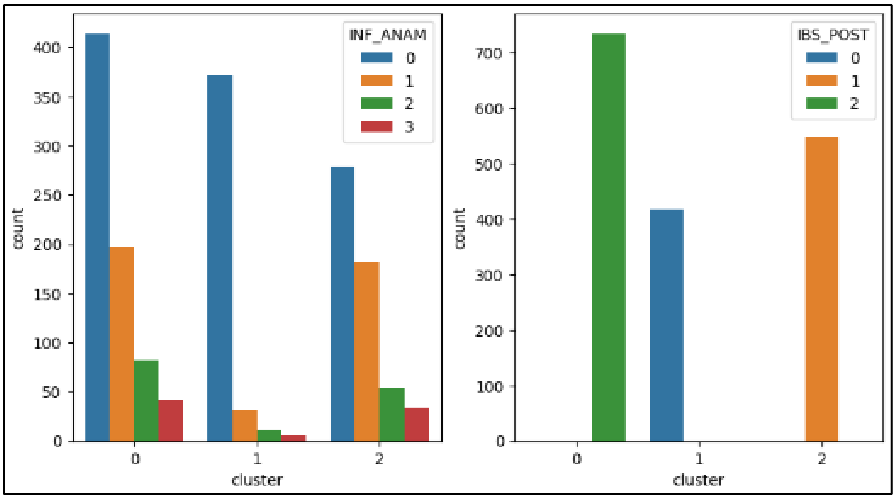

Figure 25 - Bar Chart of 'INF_ANAM' and 'IBS_POST' - Expert's Feature Selection

Similarly, feature 'INF_ANAM' also emerged as a differentiating factor, indicating the number of myocardial infarctions in the anamnesis. The combination of these variables, along with the expert's insights, suggests that the division of patients into distinct clusters is heavily influenced by their previous experiences and symptoms related to angina pectoris.

To conclude the analysis and evaluate the quality and accuracy of the clusters, we again asked the expert to review and assess whether the samples were appropriately placed within their respective clusters, similarly to what we followed in the approach PCA, ensuring a minimum coverage of 10% for each cluster. The results of the expert evaluation are presented in Table 12. It becomes apparent that the outcomes are not as favorable as those achieved with the PCA approach, as it yielded more reliable and accurate clusters. Although the approach of Expert's Feature Selection had its merits, the findings suggest that the other approach is more suitable for capturing the inherent patterns and characteristics within the dataset.

Table 12 - Accuracy of each cluster – Expert's Feature Selection

Cluster	Accuracy
0	31%
1	97%
2	16,5%

The expert, upon reviewing all the results shown above, agreed with the observations and conclusions drawn from the analysis, concluding that, even if the selected features successfully contributed to the separation and differentiation of the clusters, when attempting to interpret the clusters in terms of clinical phenotypes, he encountered challenges, since the selected features did not provide sufficient discriminatory power or domain-specific relevance to define unique and clinically meaningful phenotypes for each cluster. As a result, the expert found it difficult to discern distinct patterns or characteristics that could be easily linked to specific medical conditions or characteristics of interest. As opposed, the approach PCA method showed surprising effectiveness in capturing the underlying variability and structure in the data, which made possible to express the variance and intrinsic relationships of the data more completely, allowing the expert to identify distinct phenotypes for each cluster, as explained before, since the principal components effectively captured the most important sources of variability and served as a solid foundation for interpretation. The collaboration and insights provided by the expert played a pivotal role in guiding our evaluation and analysis of the clustering outcomes.

7. Conclusions and Future Work

This chapter summarizes the work carried out, the main conclusions, and some possibilities of future work within the scope of this project as well as the limitations on this project.

7.1. Summary

The originally defined objectives for this project were focused on investigating and implementing healthcare solutions in the field of cardiology. Specifically, our aim was to develop a decision support system (DSS) that would empower doctors with enhanced decision-making capabilities, reducing the probability of errors, and optimizing physical exercise and diet patient prescriptions regarding a cardiac rehabilitation programme. Our goal was to provide doctors with a tool that would aid them in delivering personalized and effective care to their patients. Once the initial study phase was completed, our primary goal shifted towards the development of a decision support system that could seamlessly integrate with a broader healthcare platform. This digital platform – 2ARTs – developed in parallel with the DSS, aimed to create an environment where doctors could prescribe treatments tailored to each patient's unique health profile. Moreover, it was sought to facilitate seamless communication and efficient patient monitoring, enabling doctors to offer prompt and effective support throughout the cardiac rehabilitation process.

To achieve these objectives, our project involved working on the development of a cardiac rehabilitation platform, designed to empower doctors with the capability to oversee their patients' progress seamlessly, transcending geographical boundaries and traditional hospital visits. Additionally, we were committed to the creation of a DSS that would be integrated into the backoffice, improving the decision-making skills of healthcare professionals. With its integration, these doctors are better able to provide tailored prescriptions, resulting in a more targeted and efficient method of patient treatment. With a dataset consisting of approximately 1,700 patient profiles and over 112 variables, the dataset used in the development of this DSS underwent a process of analysis and data pre-processing, including feature selection and reduction, to extract the most relevant and informative insights. Multiple modelling approaches, such as the use of PCA of Feature Selection analysis, and evaluation metrics like WCSS and Silhouette Coefficient were explored and considered, in consultation with a domain expert, to identify the most effective feature reduction techniques. Through these efforts, we developed a decision support system that not only provided doctors with valuable information and guidance but also enabled them to identify distinct clinical phenotypes within the patient population. These phenotypes, meticulously identified by an expert with the use of the approach with the best evaluation results, designate as Ischemic, Non-Ischemic, and Congestion, serving as essential indicators in patient management. By categorizing patients into specific phenotypic groups, doctors could make more informed decisions regarding treatment plans and interventions, ensuring a more tailored and effective approach to cardiac rehabilitation.

Overall, our project encompassed the creation of a healthcare platform, integrated with a decision support system, to empower doctors with the necessary tools and insights to deliver personalized care, monitor patient progress, and improve outcomes in the realm of cardiac rehabilitation.

7.2. Limitations and Future Work

The development of this project creates possibilities for the investigation and application of novel ideas and innovative functionalities. To optimize the performance of the decision support system, it would be highly valuable to gather real-world data from actual patients using the health platform developed in our study. This would enable us to create a tailored model specifically designed to meet the unique needs and objectives of our research.

The dataset used in our previous analysis encompassed a wide range of features, some of which were not directly relevant to our specific focus. However, by acquiring a more comprehensive dataset that aligns closely with the objectives of our study, we can enhance the precision and relevance of our results, leading to more refined and customized outcomes. This would empower healthcare professionals with accurate and insightful decision support, enabling them to provide personalized care and support to patients undergoing cardiac rehabilitation. We can also optimize the effectiveness and impact of our developed platform, promoting better patient outcomes and enabling efficient cardiac rehabilitation programmes, by using the full potential of a comprehensive dataset. Looking ahead, improving patient care through prescriptions that are specifically crafted to meet the unique needs of each patient group also presents a promising area for further investigation. The prescription process for doctors would be sped up in this way, and it would also guarantee that the suggestions remain to be optimally aligned with the particular needs of each patient group. Additionally, adding past prescription data to the DSS model would significantly improve its capacity for prediction, having a more in-depth understanding of patient profiles for each phenotype by incorporating a wider range of previous prescriptions. As a result, the conclusions and suggestions generated would be improved, increasing the system's capability to precisely direct medical professionals in the optimization of patient care techniques.

Finally, submitting the information gathered to the rigorous validation procedure provided by microneurography would be an advantageous route for further exploration. This advanced technique entails the insertion of fine electrodes into peripheral nerves, enabling the direct recording and analysis of neural activity. By leveraging microneurography, we could establish a direct link between the predicted clinical phenotypes and the actual physiological responses within the patient population. This validation approach would offer a higher level of confidence in the accuracy and applicability of the DSS' predictions.

Bibliographic References

[1] J.-F. Levesque, M. F. Harris, and G. Russell, "Patient-centred access to health care: conceptualising access at the interface of health systems and populations," *Int J Equity Health*, vol. 12, p. 1, 2013, doi: 10.1186/1475-9276-12-18.

[2] I. A. Pap and S. Oniga, "A Review of Converging Technologies in eHealth Pertaining to Artificial Intelligence," *Int J Environ Res Public Health*, vol. 19, no. 18, p. 11413, Sep. 2022, doi: 10.3390/IJERPH191811413.

[3] B. H. Dobkin and A. Dorsch, "The promise of mHealth: Daily activity monitoring and outcome assessments by wearable sensors," *Neurorehabil Neural Repair*, vol. 25, no. 9, pp. 788–798, Nov. 2011, doi: 10.1177/1545968311425908.

[4] O. Ben-Assuli, "Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments," *Health Policy (New York)*, vol. 119, no. 3, pp. 287–297, Mar. 2015, doi: 10.1016/J.HEALTHPOL.2014.11.014.

[5] "Advances in Healthcare Technology: Shaping the Future of Medical Care - Google Livros." Accessed: Jul. 04, 2023. [Online]. Available: https://books.google.pt/books?hl=pt-PT&lr=&id=q7MH6QWJ4gMC&oi=fnd&pg=PR9&dq=healthcare+technology+advances&ots=FI2LNAfEjE&sig=PU9xSMIJ-P6uHsb1ov7V5UW_W-s&redir_esc=y#v=snippet&q=hie&f=false

[6] F. Lopez-Jimenez *et al.*, "Artificial Intelligence in Cardiology: Present and Future," 2020, doi: 10.1016/j.mayocp.2020.01.038.

[7] "Addressing the Problem of Cardiac Rehabilitation Program Dis... : Journal of Cardiopulmonary Rehabilitation and Prevention." Accessed: May 12, 2023. [Online]. Available: https://journals.lww.com/jcrjournal/fulltext/2005/03000/addressing_the_problem_of_cardiac_rehabilitation.7.aspx

[8] "Cardiac Rehabilitation Communications Toolkit | Million Hearts®." Accessed: Jul. 31, 2023. [Online]. Available: https://millionhearts.hhs.gov/partners-progress/partners/cardiac-rehab-toolkit.html

[9] A. Abreu *et al.*, "Mandatory criteria for cardiac rehabilitation programs: 2018 guidelines from the Portuguese Society of Cardiology," *Revista Portuguesa de Cardiologia*, vol. 37, no. 5, pp. 363–373, May 2018, doi: 10.1016/J.REPC.2018.02.006.

[10] B. M. Ferreira Neves, O. Drª, C. M. Graça, P. Gomes, D. Mário, and S. Santos, "Fatores preditores de incumprimento do programa de reabilitação cardíaca na insuficiência cardíaca," 2019.

[11] H. M. Dalal, P. Doherty, S. T. J. McDonagh, K. Paul, and R. S. Taylor, “Virtual and in-person cardiac rehabilitation,” *BMJ*, vol. 373, Jun. 2021, doi: 10.1136/BMJ.N1270.

[12] J. Horsky, J. Aarts, L. Verheul, D. L. Seger, H. van der Sijs, and D. W. Bates, “Clinical reasoning in the context of active decision support during medication prescribing,” *Int J Med Inform*, vol. 97, pp. 1–11, Jan. 2017, doi: 10.1016/J.IJMEDINF.2016.09.004.

[13] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, “Big data in healthcare: management, analysis and future prospects,” *J Big Data*, vol. 6, no. 1, pp. 1–25, Dec. 2019, doi: 10.1186/S40537-019-0217-0/FIGURES/6.

[14] M. W. L. Moreira, J. J. P. C. Rodrigues, V. Korotaev, J. Al-Muhtadi, and N. Kumar, “A Comprehensive Review on Smart Decision Support Systems for Health Care,” *IEEE Syst J*, vol. 13, no. 3, pp. 3536–3545, Sep. 2019, doi: 10.1109/JSYST.2018.2890121.

[15] M. W. L. Moreira, J. J. P. C. Rodrigues, V. Korotaev, J. Al-Muhtadi, and N. Kumar, “A Comprehensive Review on Smart Decision Support Systems for Health Care,” *IEEE Syst J*, vol. 13, no. 3, pp. 3536–3545, Sep. 2019, doi: 10.1109/JSYST.2018.2890121.

[16] K. Y. Ngiam and I. W. Khor, “Big data and machine learning algorithms for health-care delivery,” *Lancet Oncol*, vol. 20, no. 5, pp. e262–e273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.

[17] “Cardiovascular Disease”.

[18] “WORLD HEALTH ORGANIZATION Regional Office for the Eastern Mediterranean ORGANISATION MONDIALE DE LA SANTE Bureau regional de la Mediterranee orientale,” 1998.

[19] J. Tessler and B. Bordoni, “Cardiac Rehabilitation,” *Essentials of Physical Medicine and Rehabilitation: Musculoskeletal Disorders, Pain, and Rehabilitation*, pp. 678–683, May 2023, doi: 10.1016/B978-0-323-54947-9.00123-1.

[20] R. S. Taylor, H. M. Dalal, and S. T. J. McDonagh, “The role of cardiac rehabilitation in improving cardiovascular outcomes,” *Nature Reviews Cardiology 2021 19:3*, vol. 19, no. 3, pp. 180–194, Sep. 2021, doi: 10.1038/s41569-021-00611-7.

[21] M. F. Piepoli *et al.*, “E U R O P E A N SO CIETY O F CARDIOLOGY ® Challenges in secondary prevention after acute myocardial infarction: A call for action,” *Eur J Prev Cardiol*, vol. 23, no. 18, pp. 1994–2006, 2016, doi: 10.1177/2047487316663873.

[22] C. A. Zanetti, A. George, R. A. Stiegmann, and D. Phelan, “Digital Health,” *Handbook of Research on the Efficacy of Training Programs and Systems in Medical Education*, pp. 404–426, May 2023, doi: 10.4018/978-1-7998-1468-9.ch021.

[23] P. Kostkova, “Grand Challenges in Digital Health,” *Front Public Health*, vol. 3, p. 147199, May 2015, doi: 10.3389/FPUBH.2015.00134/BIBTEX.

[24] H. Han, M. Huang, Y. Zhang, and J. Liu, “Decision Support System for Medical Diagnosis Utilizing Imbalanced Clinical Data,” *Applied Sciences 2018, Vol. 8, Page 1597*, vol. 8, no. 9, p. 1597, Sep. 2018, doi: 10.3390/APP8091597.

[25] “Artificial Intelligence (AI) In Healthcare & Hospitals.” Accessed: Nov. 12, 2022. [Online]. Available: https://www.foreseemed.com/artificial-intelligence-in-healthcare

[26] “What is artificial intelligence in healthcare? | IBM.” Accessed: Nov. 12, 2022. [Online]. Available: https://www.ibm.com/topics/artificial-intelligence-healthcare

[27] “16 Machine Learning in Healthcare Examples to Know | Built In.” Accessed: Nov. 12, 2022. [Online]. Available: https://builtin.com/artificial-intelligence/machine-learning-healthcare

[28] I. Haq, K. Chhatwal, K. Sanaka, and B. Xu, “Artificial Intelligence in Cardiovascular Medicine: Current Insights and Future Prospects,” *Vasc Health Risk Manag*, vol. Volume 18, pp. 517–528, Nov. 2022, doi: 10.2147/VHRM.S279337.

[29] L. Long *et al.*, “Exercise-based cardiac rehabilitation for stable angina: systematic review and meta-analysis,” *Open Heart*, vol. 6, no. 1, p. e000989, Jun. 2019, doi: 10.1136/OPENHRT-2018-000989.

[30] B. Bjarnason-Wehrens *et al.*, “Exercise-based cardiac rehabilitation in patients with reduced left ventricular ejection fraction: The Cardiac Rehabilitation Outcome Study in Heart Failure (CROS-HF): A systematic review and meta-analysis,” *Eur J Prev Cardiol*, vol. 27, no. 9, pp. 929–952, Jun. 2020, doi: 10.1177/2047487319854140.

[31] S. Ojha, “Recent Advancements in Artificial Intelligence Assisted Monitoring of Heart Abnormalities and Cardiovascular Diseases: A Review,” 2022.

[32] T. Vromen, N. Peek, A. Abu-Hanna, M. Kornaat, and H. Kemps, “A computerized decision support system did not improve personalization of exercise-based cardiac rehabilitation according to latest recommendations,” *Eur J Prev Cardiol*, vol. 28, Nov. 2020, doi: 10.1093/eurjpc/zwaa066.

[33] A. Triantafyllidis *et al.*, “Computerized decision support for beneficial home-based exercise rehabilitation in patients with cardiovascular disease,” *Comput Methods*

Programs Biomed, vol. 162, pp. 1–10, Aug. 2018, doi: 10.1016/J.CMPB.2018.04.030.

[34] R. della Méa Plentz, “Relevance and Limitations of Decision Support Systems for Outpatient Cardiac Rehabilitation,” *Biomed J Sci Tech Res*, vol. 12, no. 4, Jan. 2019, doi: 10.26717/BJSTR.2019.12.002295.

[35] M. T. Ishraque, N. Zjalic, P. M. Zadeh, Z. Kobti, and P. Olla, “Artificial intelligence-based cardiac rehabilitation therapy exercise recommendation system,” *2018 IEEE MIT Undergraduate Research Technology Conference, URTC 2018*, vol. 2018-January, Oct. 2018, doi: 10.1109/URTC45901.2018.9437568.

[36] M. A. Rahman and A. Tumian, “Variables Influencing Machine Learning-Based Cardiac Decision Support System: A Systematic Literature Review,” *Applied Mechanics and Materials*, vol. 892, pp. 274–283, Jun. 2019, doi: 10.4028/WWW.SCIENTIFIC.NET/AMM.892.274.

[37] L. S. Pescatello, Y. Wu, G. A. Panza, A. Zaleski, and M. Guidry, “Development of a Novel Clinical Decision Support System for Exercise Prescription Among Patients With Multiple Cardiovascular Disease Risk Factors,” *Mayo Clin Proc Innov Qual Outcomes*, vol. 5, no. 1, pp. 193–203, Feb. 2021, doi: 10.1016/J.MAYOCPIQO.2020.08.005.

[38] G. P. Diller *et al.*, “Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients,” *Eur Heart J*, vol. 40, no. 13, pp. 1069–1077, Apr. 2019, doi: 10.1093/EURHEARTJ/EHY915.

[39] A. Lin *et al.*, “Artificial Intelligence in Cardiovascular Imaging for Risk Stratification in Coronary Artery Disease,” *Radiol Cardiothorac Imaging*, vol. 3, no. 1, Feb. 2021, doi: 10.1148/RYCT.2021200512.

[40] M. Barrett *et al.*, “Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care,” *EPMA Journal 2019 10:4*, vol. 10, no. 4, pp. 445–464, Nov. 2019, doi: 10.1007/S13167-019-00188-9.

[41] A. Manresa-Rocamora *et al.*, “Exercise-based cardiac rehabilitation and parasympathetic function in patients with coronary artery disease: a systematic review and meta-analysis,” *Clin Auton Res*, vol. 31, no. 2, pp. 187–203, Apr. 2021, doi: 10.1007/S10286-020-00687-0.

[42] J. Casonatto, T. Tinucci, A. C. Dourado, and M. Polito, “Cardiovascular and autonomic responses after exercise sessions with different intensities and durations,” *Clinics*, vol. 66, no. 3, p. 453, 2011, doi: 10.1590/S1807-59322011000300016.

[43] M. J. Pearson and N. A. Smart, “Exercise therapy and autonomic function in heart failure patients: a systematic review and meta-analysis,” *Heart Fail Rev*, vol. 23, no. 1, pp. 91–108, Jan. 2018, doi: 10.1007/S10741-017-9662-Z/FIGURES/6.

[44] J. Saltz, I. Shamshurin, and C. Connors, “Predicting data science sociotechnical execution challenges by categorizing data science projects,” *J Assoc Inf Sci Technol*, vol. 68, no. 12, pp. 2720–2728, Dec. 2017, doi: 10.1002/ASI.23873.

[45] “Welcome to Flask — Flask Documentation (2.3.x).” Accessed: Jun. 26, 2023. [Online]. Available: https://flask.palletsprojects.com/en/2.3.x/

[46] “Cloud Run documentation | Cloud Run Documentation | Google Cloud.” Accessed: Sep. 28, 2023. [Online]. Available: https://cloud.google.com/run/docs

[47] “Myocardial infarction complications Database.” Accessed: Jun. 23, 2023. [Online]. Available: https://figshare.le.ac.uk/articles/dataset/Myocardial_infarction_complications_Database/12045261/3?file=22803572

[48] “Mean, Mode and Median - Measures of Central Tendency - When to use with Different Types of Variable and Skewed Distributions | Laerd Statistics.” Accessed: Jun. 16, 2023. [Online]. Available: https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php

[49] S. Sarkar, U. Taraphder, S. Datta, S. P. Swain, and D. Saikhom, “Multivariate Statistical Data Analysis-Principal Component Analysis (PCA),” *International Journal of Livestock Research*, vol. 7, no. 5, p. 60, 2017, doi: 10.5455/ijlr.20170415115235.

[50] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis: A Global Perspective.* . Pearson, 2014. Accessed: Jun. 23, 2023. [Online]. Available: https://www.drnishikantjha.com/papersCollection/Multivariate%20Data%20Analysis.pdf

[51] “prince/prince/famd.py at master · MaxHalford/prince · GitHub.” Accessed: Jun. 24, 2023. [Online]. Available: https://github.com/MaxHalford/prince/blob/master/prince/famd.py