TÉCNICO LISBOA

Conceptualization of a Clinical Decision Support System for the Management of Type 2 Diabetes Mellitus: An Integrated Approach Between Patient, Hospital and Pharmacy

Ana Rita de Aguiar Fontes

Thesis to obtain the Master of Science Degree in Biomedical Engineering

Integrated Master in Biomedical Engineering

Supervisor(s): Prof. Leonardo Azevedo Guerra Raposo Pereira
Carolina Sofia Lisboa Sequeira

Examination Comitee

Chairperson: Prof. Ana Luísa Nobre Fred
Supervisor: Carolina Sofia Lisboa Sequeira
Members of the Committee: Prof. Ana Catarina Lopes Vieira Godinho de Matos

November 2021

Preface

The work presented in this thesis was performed at the company Glintt (Quinta da Beloura, Portugal), during the period March-Septembter 2021, under the supervision of Carolina Sequeira. The thesis was co-supervised at Instituto Superior Técnico by Prof. Leonardo Azevedo.

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Abstract

The work presented in this thesis was developed in the context of an internship at Glintt and is a part of the ConnectedHealth project, that aims to create an interoperable health information system highly focused on the individual, towards a more preventive and value-based medicine. One of its goals is to incorporate a Clinical Decision Support System, which became the focus of this work. Having chosen the study case of Type 2 Diabetes Mellitus, one of the most prevalent chronic diseases worldwide, the objectives became: i) to understand how Clinical Decision Support is improving health outcomes of these patients nowadays; ii) to propose a solution for its application in the context of this work. The first objective was accomplished by an intense phase of research that served as a basis for the second objective. After following an action research methodology, two sort of Clinical Decision Support solutions were proposed: a knowledge-based and a non-knowledge-based solution. The first one is based on predefined rules, mostly for warning of harmful events/situations. It also included a proposal of how ConnectedHealth could be used in practice for the management of these patients, considering its goals and the functionalities of existent Glintt solutions. The second one is a proposal of how machine learning could be used to bring value for disease management, using the data that will be collected by the system to train algorithms for predictions. The first solution received positive feedback from healthcare professionals and there are intentions to put it into practice. The second will need further studies when real data is collected.

Keywords

Clinical Decision Support Systems; Type 2 Diabetes Mellitus; Value-Based Medicine; Interoperability; Machine Learning

Resumo

O trabalho apresentado nesta tese foi desenvolvido ao longo de um estágio na Glintt, no âmbito do projeto ConnectedHealth, cujo principal objetivo é criar um sistema informático de saúde interoperável e focado no indivíduo, tendo em vista uma medicina mais baseada na prevenção e na criação de valor. Um dos seus objetivos é a incorporação de sistemas de Apoio à Decisão Clínica, que é o foco deste trabalho. Escolhendo como caso de estudo a Diabetes Mellitus Tipo 2, uma das doenças crónicas mais prevalentes globalmente, definiram-se como objetivos do trabalho: i) Perceber como sistemas de Apoio à Decisão Clínica estão a ser utilizados atualmente para melhorar os resultados clínicos destes pacientes; ii) Fazer uma uma proposta de uma solução para ser aplicada no contexto deste trabalho. O primeiro objetivo foi alcançado através de uma fase intensa de pesquisa, que serviu de base para o segundo objetivo. Seguindo uma metodologia de investigação/ação, dois tipos de soluções para Apoio à Decisão Clínica foram propostas: a primeira baseada em conhecimento e a segunda não baseada em conhecimento. A primeira tem por base regras pré-definidas, maioritariamente para alertar para eventos/situações adversas. Também inclui uma proposta de como o ConnecedHealth pode ser posto em prática para gestão destes pacientes, tendo em conta os seus objetivos e as funcionalidades de algumas soluções já existentes da Glintt. A segunda é uma proposta de como *machine learning* pode ser usado para trazer valor para estes pacientes, usando os dados que se planeiam recolher com o sistema para treinar algoritmos de forma a fazerem previsões. A primeira solução recebeu feedback positivo por parte de profissionais de saúde e há intenções de a colocar em prática. A segunda solução necessita de mais estudos quando dados reais foram obtidos.

Palavras-Chave

Apoio à Decisão Clínica; Diabetes Mellitus Tipo 2; Medicina Baseada em Valor; Interoperabilidade; Machine Learning

Agradecimentos

Em primeiro lugar queria agradecer à Carolina Sequeira o voto de confiança neste projeto e todo o apoio que me deu ao longo dos últimos meses. Obrigada também por seres um exemplo de que, por trás de grandes profissionais, também estão grandes pessoas.

A toda a equipa da Glintt também um obrigada gigante pela boa disposição e pelo apoio incondicional, nunca deixando para trás o trabalho que tem de ser feito. Vocês fizeram-me acreditar que o ditado "*Faz o que gostas e não tens de trabalhar um único dia na tua vida*" pode ser real. Queria agradecer em particular ao Nuno Oliveira, à Laura Saldanha e à enf. Pereira Lopes que deram contributos essenciais para este trabalho.

Queria também agradecer ao professor Leonardo Azevedo a ajuda imprescindível e por ter aceite este desafio, mesmo não sendo um *expert* na área da saúde. Obrigada por ter apoiado todas as ideias e pelas sugestões valiosas, que nunca me fizeram sentir desamparada.

Por fim, à minha família que sempre esteve lá para tudo e me atura todos os dias, e aos amigos que fazem da vida um sítio melhor. Era tudo mais difícil sem vocês.

Content

List of Figures

List of Tables

Abbreviation List

Abbreviation	Meaning	Page
DM	Diabetes Mellitus	1
CDS	Clinical Decision Support	1
CDSS	Clinical Decision Support System	1
ML	Machine Learning	1
CDR	Clinical Data Repository	2
WHO	World Health Organization	2
T1DM	Type 1 Diabetes Mellitus	3
T2DM	Type 2 Diabetes Mellitus	3
ICHOM	International Consortium for Health Outcomes Meaasurement	4
EHR	Electronic Health Record	5
CPOE	Computerized Physician Order Entry	5
HL7	Health Level 7	8
SAGE	Shareable Active Guideline Environment	9
SEBASTIAN	System for Evidence Based Advice Through Simultaneous Transaction with an Intelligent Agent Across a Network	9
FHIR	Fast Healhcare Interoperability Resources	9
API	Application Programming Interface	9
CHO	Carbohydrate	12
HbA1c	Glycated Haemoglobin	13
BMI	Body Mass Index	20
SBGM	Self Blood Glucose Monitoring	22
CGM	Continuous Glucose Monitoring	22
RF	Random Forest	23
SVM	Support Vector Machine	23
ANN	Artificial Neural Network	23
GDPR	General Data Protection Regulation	72

1. Introduction

1.1 Motivation

Healthcare systems in developed countries currently face the challenge of an ageing population with a burden of chronic illnesses, resulting in a decrease in the quality of life and an increase in the expenses of public health systems due to inefficiencies in its management [1]. There is a critical need to adapt and provide better time and cost-effective solutions to deal with this issue.

One of the most prevalent chronic illnesses worldwide is Diabetes Mellitus (DM), affecting 9.1% of the population across all countries [2]. Briefly, diabetic patients can be managed by controlling lifestyle variables, getting the adequate treatment (insulin and/or hypoglycemic medication) and doing the necessary exams to control the appearance of associated complications, in order to achieve the best health outcomes possible [2]. This requires a coordination of efforts from both the patient and different healthcare professionals. Such an endeavor could be enhanced with the use of the right technology.

Living in the era of technology, and building upon the current digitalization of our society, health is one of the sectors that can be the most benefited. Technology can be used in the different stages of healthcare, including prevention, diagnosis and treatment [3]. One of the most prominent examples of the application of technology for the better management of health is Clinical Decision Support (CDS). A Clinical Decision Support System (CDSS) is a system that performs CDS and aims to improve healthcare delivery by supporting medical decisions or actions, using the correct information and knowledge at the right time [4].

CDS can be achieved either from simple if-then programable rules or from systems that generate their own knowledge from the data, in a data-driven approach [4]. One motivation for the latter is the fact that nowadays so much information is generated daily, namely health-related information, that it is possible to analyze that information with appropriate models, usually Machine Learning (ML) models, to collect patterns and generate new knowledge, a concept known as data mining, that ultimately intends to improve care delivery [5].

1.2 Context

1.2.1 The ConnectedHealth Project

This dissertation was developed along the course of an internship in Glintt, a Consulting and Technological Services company, mostly focused in the healthcare sector. The scope of this work is inserted in ConnectedHealth, an European-funded project coordinated by Glintt, in collaboration with some key partners, including Fraunhofer, NOS and Lusíadas Saúde Group. The overall goal of the project is to create a set of integrated, interoperable and personalized tools

and services with the ability to improve the way healthcare is delivered. It intends to have an impact on prevention, detection, treatment, monitoring and management of disease.

Specifically, the work developed is inserted in *Subproject 1: A Smart Hospital Information System Using a Patient Driven Approach,* that aims to create an interoperable health information system highly focused on the individual and towards a more preventive medicine. Such solution should also support lifestyle and allow the care, monitoring and treatment both inside and outside the hospital, with participation of the individual. Finally, patient's security needs to be maximized, both when treated inside a health unit and when transitioning between different health units or at home care. Therefore, the solution should be highly integrated and interoperable. In order to accomplish it, some technological features are planned to be developed, including a Clinical Data Repositorium (CDR), gathering the data from different sources, and a Global Rules Engine that receives information from the CDR and generates alerts, based on configurable rules, which can be considered CDS.

Despite being inserted in one subproject, the work developed has to consider the goals of the others. Namely, the creation of an open Internet of Things platform connecting devices, sensors and systems and the creation of a cloud-based pharmacy solution focusing on personalized care. The final objective is to validate the solutions proposed in real life contexts, with 6 use case pilots planned for exploration. The work developed for this dissertation is focused on *Study Case 1: Providing Care at Home for Patients with common chronic diseases: Diabetes Mellitus Type 2.*

Despite the division into subprojects, ConnectedHealth also defined some general goals, namely: to convert the information collected into knowledge, in order to provide new insights to the medical community. Additionally, it aims to transform Portugal in an international reference for value-based healthcare.

1.2.2 Diabetes: An Epidemic

According to the World Health Organization (WHO), DM is a chronic disease characterized either by lack of enough insulin production by the pancreas or by lack of efficient use of the insulin produced by the body [2]. Insulin is the hormone that regulates the sugar (glucose) that circulates in the blood. If poorly regulated, it leads to an increase in glucose levels, a condition known as hyperglycemia. If the body is in a hyperglycemic state for too long, it can have a very negative impact, leading to damage of nerves and blood vessels, with the possibility of developing blindness, kidney failure, heart attacks, strokes and lower limp amputation. Additionally, under certain circumstances, for example too much intake of anti-hyperglycemic drugs, the body can also enter in an acute hypoglycemic state that might be lethal.

Statistically, the WHO estimates that, in 2014, about 422 million people worldwide had the disease and, in 2019, about 1.5 million deaths were directly caused by diabetes [2]. These numbers are worrying and rising, making diabetes an "epidemic" [6] and a challenge for all the health systems in the world. There are two main types of Diabetes Mellitus:

- Type 1 Diabetes Mellitus (T1DM): an autoimmune disease, where the immune system attacks and destroys the insulin producing beta cells in the pancreas. It has an early onset in life and the patients are completely insulin-dependent, requiring daily administrations of the hormone since their bodies are not able to produce it [2].
- Type 2 Diabetes Mellitus (T2DM): It usually appears at some point in adulthood, associated with lifestyle factors such as overweight and lack of physical activity. This condition is usually characterized by insulin resistance [7]: the cells in the body resist the effect of insulin to drive glucose in the blood to their inside. As a result, glucose accumulates in the blood and the pancreas. In an attempt to maintain the normal levels of blood sugar, the pancreas reacts to this by producing even more insulin. Eventually, the pancreas reaches its maximum capacity and there is a point when the beta cells production is compromised due to exhaustion.TD2M is by far the most common type of diabetes, accounting for more than 90% of all cases [8].

In Portugal, the numbers are disturbing. According to a report from *Observatório Nacional da Diabetes*, built in accordance with directives from the Portuguese Directorate-General for Health (Direção Geral da Saúde), in 2018 the prevalence of diabetes in the Portuguese population aged 20-79 years was 13.6% [9]. This means that more than one million Portuguese have diabetes. This trend gets even worse with ageing, since more than one quarter of the population within the age group from 60-79 years presents the disease. In terms of lethality, about 26.6% of deaths in hospitals were of diabetic patients, which means that in Portugal more than 1 in every 4 patients that die in hospitals have diabetes [9].

Some basic domains for diabetes care can be identified. In the context of this work, it makes sense to make two distinctions:

- Clinical Management of Diabetes: in terms of efforts from the healthcare teams to maintain patient's health over time. It will be explained in greater detail in Section 2.3.1.
- Daily-Life Management: one encompasses all the lifestyle variables that have to be managed daily by the patient. It will be explained in greater detail in Section 2.4.1.

1.2.3 Considerations for the Management of Chronic Diseases

1.2.3.1 Limitations of Traditional Healthcare Systems

Chronic diseases are influenced by several complex factors, thus being a dynamic process to deal with. This characteristic represents a challenge to traditional healthcare model systems, as these were mostly built to handle acute symptoms, often failing to manage patients with chronic diseases [10]. Thus, there is a need to create better management models for the disease and for the overall well-being of the chronic patient. In opposition to acute diseases, where patients are treated in a physician-centered mode, with short appointments and little instruction on their condition, treatment of chronic diseases requires a long-term management,

that aims to first stabilize the patient's condition and then prevent long-term complications [11]. Consequently, healthcare is being directed towards a Patient-Centered Care Approach, that ensures all health-related decision are valued by the patient and respect their personal preferences and needs [12]. That should also include health maintenance of individuals at home, to guarantee their appropriate management and avoid burden in hospitals: home-centered care [10]. An Australian study, for example, claims that chronic diseases are associated with more than one third of avoidable hospitalizations in the country [13].

Implementing an approach for the management of the disease that is centered on the patient requires an integration, communication and data interoperability between the different institutions where the person might have records, so that the data can follow the person in the different points of care instead of being sectioned across distinct health information systems. However, this does not seem to be the case in Portugal, where ninety different information systems have been identified in the various health systems [14]. These either do not communicate at all or communicate with delays and in unreliable ways. Consequently, analysis and tests might be replicated, which results in higher costs for institutions, affects the patient and represents inefficiency for the healthcare system in general.

However, information systems usually only contain the data that the patient provides in healthcare units, whereas the information on his/her overall health can be spread outside those units. Namely in pharmacies and at home (e.g., eating habits, sleep cycles, physical activity, etc). The transition to a patient-centered and home-centered care approach implies a need for the various healthcare professional to access and register patient data from anywhere, obviously with the appropriate consent from the latter.

1.2.3.2 Value-Based Healthcare

Traditionally, healthcare has been in a Fee-for-Service Model, where hospitals are reimbursed for the amount of services they provide (the number of appointments, medical exams and interventions performed, etc.). However, this does not always mean better health outcomes. In fact, healthcare professionals may be so focused in meeting certain indicators and targets that they might fail to give each individual patient the necessary time to assess their needs and what would really represent the best health outcomes individually. This is where the concept of Value-Based Healthcare comes: it measures the improvement in a patient's health outcomes for the cost of achieving that improvement [15]. The goal is to maximize the value of care for patients, minimizing the cost of healthcare and, in the end, recalibrate the way healthcare is measured, by prioritizing patient's health outcomes instead of services provided [16].

New efforts have been put into practice to measure patient's reported outcomes in the best way possible. An example is the International Consortium for Health Outcomes Measurement (ICHOM), that aims to find global standards to measure outcomes that matter the most to patients, in order to fully unlock the potential of value-based healthcare worldwide [17].

1.2.4 Clinical Decision Support Systems

CDS is a process that aims to enhance decisions and actions related to health by using the right available clinical knowledge and patient information. The receivers of CDS can be patients, doctors or anyone else involved in the delivery of healthcare to the patient. CDS can include delivery and guidance by providing clinical knowledge when needed or it can simply include the filtering of data in an intelligent way. The formats in which the information is delivered can also vary: they can be data and order entry facilitators; filtered data displays; referencing of information; alerts and others. [18]

Living in the era of digitalization and big data, the role of CDSS acquires an increasing importance. In the last few decades, the digitalization of patient records, in the form of Electronic Health Records (EHR), combined with Computerized Physician Order Entry (CPOE) has been the norm. However, the full potential of both systems can only be achieved with a quality clinical decision support. This is due to the fact that, many times, a physician needs to be aware of an immensity of information regarding the patient when deciding and this decision might need to be made within seconds or minutes. Besides that, medical knowledge is constantly being updated and the doctor is not always informed about the most recent insights and/or does not have the time to search for them [19]. The physician can also be tired and not able to remember all the important information at all times. Fatigue does not happen to computers. Therefore, the integration of CDSS within health informatic systems, if well performed, had the initial aim of helping health care professionals in their practice by including patient safety alerts, health maintenance reminders or help in the diagnosis [19].

More specifically, CDSS can improve healthcare delivery in the following aspects [4]:

- Patient Safety: reducing the errors of prescription and medication that cause adverse events. For example, by detecting harmful drug interactions in prescription or something that the patient is allergic to.
- Clinical Management: follow up and treatment reminders, making sure there is an adherence to clinical guidelines.
- Cost Containment: reducing test and order duplication or suggesting cheaper treatments or medications.
- Automation of Administrative Functions: Automatically providing diagnostic code selection or note auto-fills.
- Diagnostic Support: providing diagnostic suggestions based on patient's data or even by interpreting medical images and laboratory results.
- Patient Decision Support: decision support provided directly to the patient through personal health records/other systems.

There are several ways to classify and subdivide CDSS, but the most relevant for this dissertation is the distinction between knowledge and non-knowledge based CDSS [4][20]. The architectures for each are described in Figure 1.1. Knowledge-based systems are composed of

programmed rules (i.e., if-then statements). The system verifies the data that is given against the rule and produces an action or output. Rules can be created according to literature-based, practice-based or patient-directed evidence and guidelines. These systems require all possibilities to be programmed explicitly. Non-knowledge-based systems also require a data source, but they apply artificial intelligence algorithms that are able to learn and establish their own rules based on the data. The main disadvantage of the latter is the fact that they are "black-box systems", which means that doctors cannot understand the logic they use to produce recommendations, making them often unreliable. This is also the reason why they are not so widespread, despite seeming to be a promising solution [21].

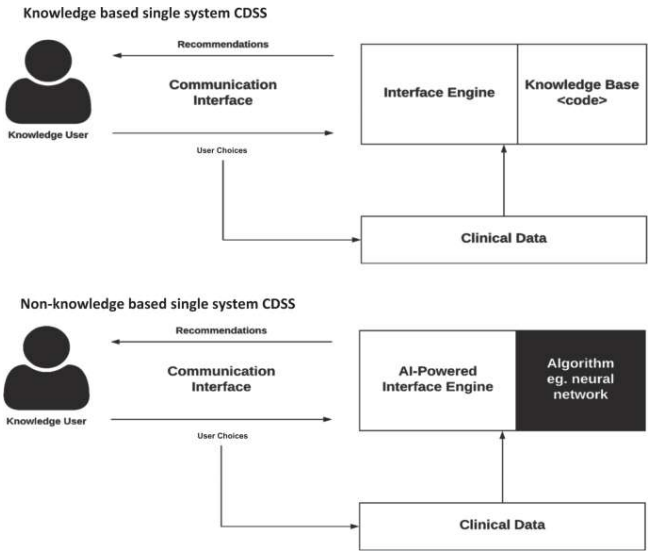


Figure 1. 1 Architectures of knowledge and non-knowledge based CDS systems [4]

Ideally, a CDSS should deliver information and knowledge specific of the patient to the stakeholders involved (i.e., clinicians, patients and/or others), with the ultimate goal of enhancing healthcare delivery. However, nowadays these systems are more focused in delivering information to the clinician rather than the patient [22]. This is not aligned with the direction that health systems are taking to position themselves in a more patient-centered spectrum. To date, when developing these systems, most researchers only consider medical determinants of health. Often, patient-related factors are ignored, including social determinants, patient-specific values, comorbidities and preferences. Taking some of these into account, especially for chronic diseases where self-management seems to be so important, is an important step towards more efficient CDSS and a step towards value-based healthcare.

1.3 Objectives of the Dissertation

At the beginning of the internship at Glintt, the ConnectedHealth project was still at very early stages: the main ideas of what was planned to be developed were there, some are explained in Section 1.2.1, but still nothing concrete had came forward. This work is one of the kickstarts of the project, aiming to propose more concrete and feasible solutions, with a special focus on the CDSS that subproject 1 intends to incorporate.

Several study cases for the application of the ConnectedHealth project had already been thought and proposed. Picking up the use case for the management of T2DM, the concrete objectives of this dissertation became:

i) Understanding how CDS are currently being used to improve health outcomes of T2DM patients

ii) Understanding how CDS, both knowledge and non-knowledge based, could be applied in practice, considering the objectives of the project, and propose a concrete solution for its application.

It is important to note, however, that the work described is a part of a 3-year project ending in 2023 and counts with a multi-effort from different companies. It was clear from the beginning that the solution proposed would be more conceptual than practical and the implementation of the ideas would take some time, since they need to be approved and to follow the appropriate bureaucracies and, afterwards, a technical development process.

1.4 Structure of the Dissertation

This work is divided in 5 chapters:

- Chapter 1 (Introduction): Presents the main ideas and concepts behind the solution that will be proposed.
- Chapter 2 (State of the Art): The disease is explored, including how its most important aspects are currently being managed. Current CDS uses for diabetes are discussed, as well as the state of other aspects crucial to be integrated in the ConnectedHealth project, namely: interoperability between CDS and health information systems; value-based healthcare and the role of the pharmacy for chronic diseases. Chapter 2 is an important basis to understand the rest of the work, thus being quite extensive.
- Chapter 3 (Methodology): The methodology that was used to develop the work and arrive to the proposal of the solution is explained.
- Chapter 4 (Proposed Solution and Discussion): The solution proposed is presented and discussed. It is divided in two main parts: the knowledge-based solution (4.1) and the non-knowledge-based solution (4.2). Legal Considerations are also made on the solution (4.3), as well as what can be done in the future (4.3).
- Chapter 5 (Conclusions): Presents the main conclusions that can be drawn from this work, focusing mostly on how it can bring value to the current spectrum of T2DM management.

2. State of the Art

2.1 Interoperability Considerations on CDSS

In this section, the evolution of architectures for delivery of clinical decision support will be described, in order to understand the current spectrum and the need for standards and interoperability. 4 main architectures can be identified and are presented in chronological order next:

1) Standalone Decision Support Systems, beginning in 1959 [23]

These systems had no integration with the clinician workflow, running independently of any other system: the users had to access the system and enter the necessary data with their own front-end interface to obtain clinical decision support.

The systems were accurate, but had low adoption, mostly because of this lack of integration, not allowing proactivity: usually the people are unaware of their lack of knowledge and their need for external aid. Besides that, when using it, clinicians lost some time entering the necessary data in the system's own standards and terminology.

2) Decision Support Integrated into Clinical Systems, beginning in 1967 [23]

In this phase, CDS was integrated into CPOE and EHR systems. This integration brought advantages compared to the standalone systems: the user did not have to enter information that was already stored electronically. This also allowed the system to be proactive and emit instant alerts when detecting a dangerous situation such as drug-drug interactions.

However, these sorts of systems had some downsides: they were directly built and integrated into a larger clinical system, with no easy way to share or reuse their content. Besides that, the knowledge bases were difficult to maintain: for example, if a clinical guideline was updated, the entire source code probably had to be reviewed.

3) Standards for Sharing Decision Support Content, beginning in 1989 [23]

To overcome the difficulty in sharing decision support systems, a number of standards were created. For example: Arden Syntax, a standard language for encoding, representing and sharing medical knowledge across different information systems through Medical Logical Modules, each one containing knowledge and logic for a single decision/alert [26]; GELLO: a standard query and expression language providing a framework to manipulate clinical data for decision support in healthcare [27]. Both these standards are approved by Health Level 7 (HL7), an organization that provides a framework and standards for the exchange, integration, sharing, and retrieval of electronic health information.

Overall, these standards allow the separation of the code for clinical decision support from the code of the clinical information system, which makes it more shareable.

There are some disadvantages, namely: too many standards to choose from; the standards constraining what a user can encode, which did not happen it the previous phases, and the issue with terminology: either CDS and clinical information systems use the same terminology or a mapping effort needs to occur.

4) Service Models, beginning in 2005 [23]

Recently, the efforts have been directed to separate the clinical decision support and the information systems, recombining them through standard application programming interfaces (APIs), i.e. standard interfaces are used to enable them to operate together. The earlier examples include the Shareable Active Guideline Environment (SAGE) and System for Evidence Based Advice through Simultaneous Transaction with an Intelligent Agent across a Network (SEBASTIAN) systems, which use opposite approaches. SAGE places the standardized interface in the clinical system's side: the CDSS can access this information in a standardized way. SEBASTIAN, on the other hand, uses the standard interface in front of the clinical decision support, which can be called in a standardized way from the clinical system.

These systems are just described as the first service models and recent information on their use could not be found. However, it seems clear that Services Models are the best approach to achieve interoperability in healthcare. In order to fully accomplish that, a strong standard has to exist for communication between the institution's information systems and the CDSS. This is where HL7 FHIR, towards which the healthcare sector in transitioning to, can have an important role.

2.1.1 FHIR

FHIR stands for Fast Healthcare Interoperability Resources and is a standard for healthcare data exchange, created by HL7. HL7 has had a work of 20 years in the production of standards for data exchange. FHIR leverages the best practices and aims to simplify implementation without sacrificing integrity of information [24]. FHIR solutions are easily applicable to several contexts, including mobile phone apps, cloud communications, EHR-based data sharing, server communication in institutional healthcare providers, etc [25].

FHIR contains two primary components at its core [26]:

- Resources: collection of information models that defines the data elements, constraints and relationships for the "business objectives" most relevant to healthcare. Examples of resources are "patient", "procedure", "order" and "observation".
- APIs: collection of well-defined interfaces for interoperability between two applications.

Overall, FHIR allows to: define these healthcare business objects, relate them, implement them in a computable form and share them across well-defined interfaces. It also provides specifications for the implementation of servers capable of requesting and delivering FHIR business objectives.

FHIR is a promising solution to achieve real interoperability in healthcare and is being adopted by Glintt and implemented in the latest solutions.

2.1.2 FHIR-based CDS Hooks API

The CDS Hooks API [27] is a specification built on FHIR and describes how a client (usually a EHR) can automatically invoke external CDS services during certain events, named "hooks" in the normal clinical workflow. The output of the CDS service is presented to the client in the form of a card. The flow between the hook and the CDS Services is briefly described in Figure 2.1.

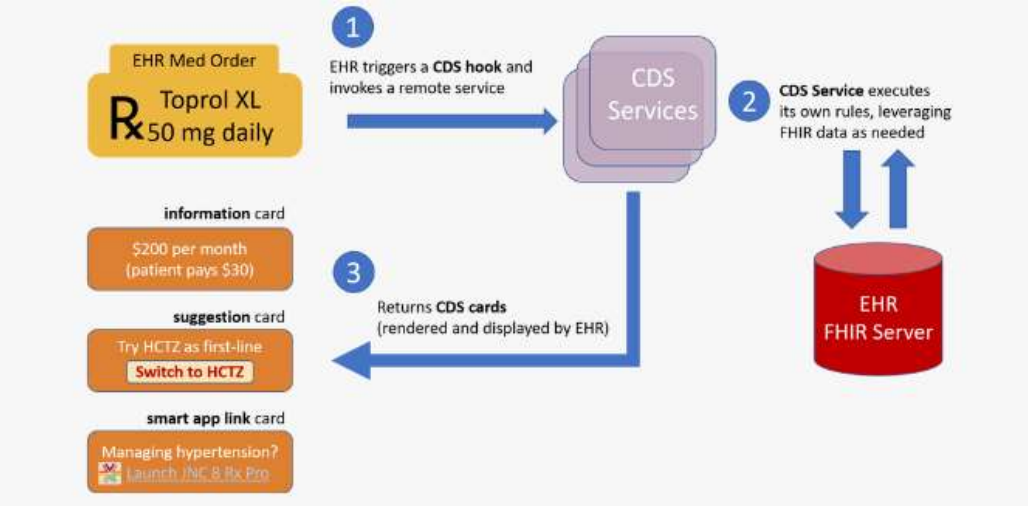


Figure 2.1 Architecture of the CDS Hooks Specification [27]

During the normal clinical workflow, for example when a clinician is initiating a prescription, some hooks can be triggered, in the case of the figure the "medication-prescribe" hook. When such a trigger occurs, the CDS client notifies each CDS service associated with that "hook". Each service gets basic details about the context of the hook, as well as service-specific data. The services must provide nearly real-time feedback on the event, in the form of cards (each service can produce multiple cards in response to the hook). In the example of Figure 2.1, the user, in this case the doctor, enters a prescription for a medication. It automatically invokes a CDS service that receives the associated code and searches for medications with a similar action and a lower price. The cards returned to the user by the CDS Hooks specification can have different forms:

- Information cards, just providing reading text for informative purposes.
- Suggestion cards, with a specific suggestion that the user can accept by clicking a button or decline.
- App link card, providing a link to an app adequate to the issue in hands.

Even though it is a relatively recent standard, a few examples in literature can be found of the application of CDS Hooks in practice. A group of researchers presented a FHIR-based CDSS platform [28] that has two different systems: a Rule Engine (knowledge-based) and a Bayesian

Engine (non-knowledge based). The former consists in an inference system based on rules manually formalized based on clinical guidelines. In the form of if-then rules, this system receives patient information as input and returns conclusions about the patient with recommendations for further action in the form of artifacts (free text). Regarding the Bayes Engine, based on Bayesian networks, it is able to make conclusions regarding diagnosis under conditions of high certainty by applying the algorithm. The models were created having clinical guidelines and scientific publications as basis. The CDS Hooks specification was used for the invocation of both engines within predefined moments in the workflow within the hospital information system.

2.2 Clinical Data Repositories

Clinical Data Repositories are secondary databases that receive data originally input into other sources. It intends to facilitate organization and querying of data [29].

An example is the commercial Smile CDR. Smile CDR was designed around the HL7 FHIR standard and is used to store health records. It is highly flexible and scalable. It can serve as a centralized repository inside a hospital, for health exchange of data from different sources, or as a powerful backend for an health app development [30].

The smile CDR can receive data from different sources: apps, EHRs, other data sources (different hospital data sources, for example) and store it, based on the FHIR specification. That data can then be queried and used for different purposes [31]. The architecture is described in Figure 2.2, where one can visualize what the components of the CDR are and the possible origins of the data (EHRs, apps, other data sources):

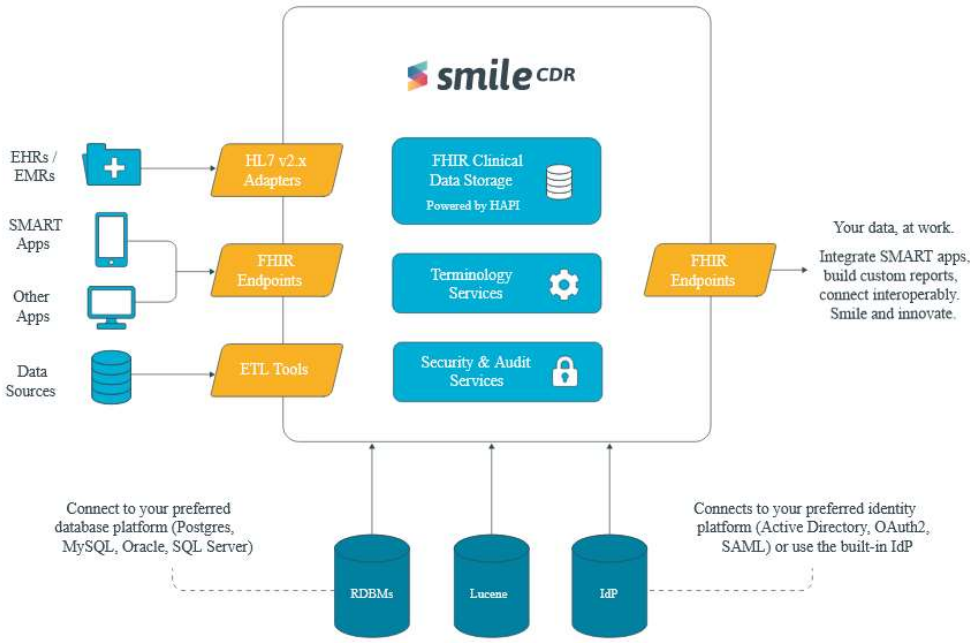


Figure 2. 2 Architecture of Smile CDR [31]

2.3 Clinical Management of Diabetes

In this section, considerations will be made about how diabetes is currently managed clinically, in terms of efforts from the healthcare teams to maintain patient's health over the time. Furthermore, one will look at how CDS can be used to improve it.

2.3.1 Current Management of T2DM

The main goal of diabetes management is to control the glycaemic values of the patient, in order to prevent the most life-threatening outcomes of diabetes: both short-term and long-term complications. Some of the information that will be presented was provided directly by healthcare professionals from the Lusíadas Group and the rest was obtained from literature.

Short-Term Complications

In terms of short-term complications, **hypoglycemias** are worth being discussed. An hypoglycemia occurs when there is not enough glucose in the blood, with levels below 70 mg/dL. It can happen due to the effects of medication, if the person if taking too much insulin or certain medications such as sulfonylureas or glinides; if there is a meal skip or delay; if not enough carbohydrates were consumed in the previous meal; if the person does too much physical exercise and exaggerates on intensity or if there is an alcohol excess.

Early signs of low blood sugar are: sweating, tiredness, dizziness, hunger, trembling, among others. If it is not quickly treated, it can lead to: weakness, blurred vision, difficulty in concentration, and, eventually, seizures, collapsing, coma in the worst scenarios.

The patients should always be aware of these signs in order to act quickly and bring glycemic values back to normal. This sort of episodes can vary from mild to severe. Mild hypoglycemias can be treated by the own patient, by eating or drinking 15-20g of fast acting carbohydrates (CHO). Severe hypoglycemia, on the other hand, requires hospitalization if there is a loss of consciousness or seizures for more than 5 minutes. If not treated quicky, it can potentially be lethal.

There are not always warning signs of hypoglycemias, so the key to handle them is to understand their causes and prevent them. The regular checking of glycemic values is also important to evaluate if the values are within normal ranges.

Long-Term Complications

Over time, high blood glucose can damage both small (microvascular) and large (macrovascular) vessels. In T2DM, it is possible that some of these complications are already occurring at the time of diagnosis, since it is a disease that progresses slowly. Getting a control of glucose levels is important to prevent worse outcomes from those complications.

Microvascular complications affect the following organs [33][34]:

- Eyes: it can lead to cataracts and **retinopathy** and both can lead to blindness. It is important that the diabetic patient has at least a yearly eye check, including a dilated eye examination. It is estimated that a diabetic person aged 50-69 years is four times more likely to develop visual impairment than a non-diabetic person in this age group [35].
- Kidneys: diabetic **nephropathy** can cause the kidneys to fail. To prevent it, certain values should be evaluated in blood/urine tests regularly, especially to detect the appearance of microalbuminuria, an early sign of kidney failure. Once it is detected, patients can start specific medication. More specifically, the values that should be obtained regularly include [36]: Urine albumin to creatinine ratio (UACR), Blood Creatinine and Estimated Glomerular Filtration Rate (eGFR).
- Nerves: this sort of nerve damage is called diabetic **neuropathy** and happens when the blood vessels that feed the nerves get damaged, damaging the nerves as well. The most common form is diabetic peripheral neuropathy, affecting the nerves that go to the hands and feet. The most concerning situation is usually related to sores that appear in the foot. They can get infected and, if the infection spreads, the foot may need amputation. So, it is important to have regular foot exams by a podiatrist, but in the follow up appointments with a nurse, explained next, a foot examination should also take place. It is estimated that an individual in need of lower leg amputation is 22 more times likely to be diabetic than non-diabetic [35].

Macrovascular complications [33][34] occur when the large blood vessels are affected: plaques accumulate and eventually it can lead to heart attack and stroke. Therefore, it is important to keep blood pressure and cholesterol under control.

To sum up, some exams should be performed regularly in diabetic patients, namely: eye and foot checks. Blood and urine checks are also important to analyse kidney-related levels, cholesterol and also glycated haemoglobin (HbA1c). A HbA1c test gives an estimation of the average blood sugar levels in the blood in the previous two or three months, since it measures how much glucose is attached to erythrocytes circulating in the blood [36].

HbA1c should be obtained regularly, at least twice a year, but it can vary on the doctor's decision, since it is the biggest control of how the disease is truly progressing. The target is usually 7% in adults [36], but it can also be personalised and individualised for each patient regarding his/her characteristics. Note that, even though it reflects the average blood glucose over the last few months, blood glucose should also be measured daily for proper management of the disease and to handle acute situations.

Glycaemic values are controlled first by an adequate lifestyle and by the use of the adequate medication: usually, for type 2 patients, they start with anti-hyperglycaemic medication and it can evolve to the need for insulin.

Usually, these patients are accompanied by a doctor, often an endocrinologist or a physician from internal medicine, who defines the medication and makes the laboratory requests.

They are also followed by a specialist nurse with an important role in assessing some aspects to manage patient's lifestyle, following a pre-defined script, that gathers information related to patient's biometric data, smoking/drinking habits, psychosocial status, updated vaccination, physical activity and eating patterns, among others.

These patients are also followed regularly by an ophthalmologist and, ideally, a podiatrist to assess the risk or control retinopathy and neuropathy complications. The frequency they should visit these specialists depends on their stage of the disease and whether or not they have already been diagnosed with the complication.

In the following sections, one is going to look into how CDS is being deployed to ensure the adequate clinical management of diabetes, both through knowledge and non-knowledge based systems.

2.3.2 Knowledge-Based CDS for Clinical Management of T2DM

In this section, one will look at examples of how knowledge-based CDS can be used to improve the clinical management of diabetic patients. Three examples will be analyzed: An implementation of the SEBASTIAN web service, COMPETE II Project and Diabetes Wizard.

USING SEBASTIAN [37]

SEBASTIAN is a standards-based web service for Clinical Decision Support. Implemented as a web service (phase 4 of the architecture, as was previously discussed in Section 2.1). It uses XML messages to communicate with client systems [37].

SEBASTIAN uses a HL7 standard to model patient information. As explained in Figure 2.3, its main feature is Executable Knowledge Modules (EKMs), that capture medical knowledge in XML documents. Each module specifies the data it requires to assess a patient and the logic it requires to generate conclusions using patient-specific data. The primary objects returned to the client's system are the EKM results, following evaluation of a patient.

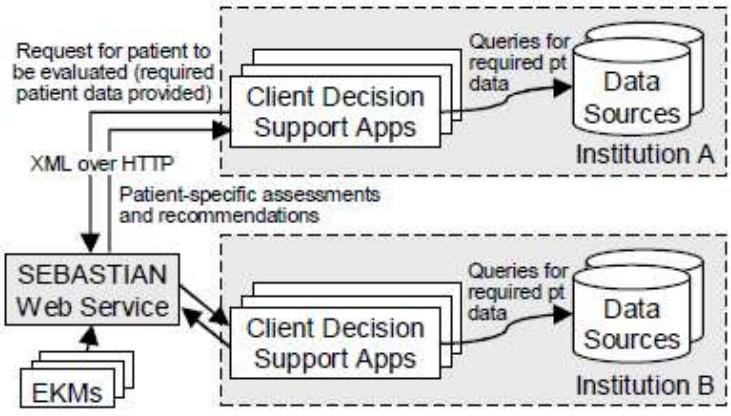


Figure 2. 3 Overview of SEBASTIAN Architecture [37]

One of the implementations of SEBASTIAN was for an Outpatients Diabetes Reminder System (DRS) at Duke Family Medicare Center [37]. When a diabetic patient checks at the clinic,

a nurse requests a diabetes care reminder system for the patient. This works the following way: there is a Duke Common Clinical Data Repository that contains all available information regarding the patient. The DRS controller requests the necessary information through web service interfaces and then makes a request to SEBASTIAN to evaluate the data using 13 EKMs for diabetes management, making use of information available in the CDR, including: prior encounter diagnoses, procedures, past and scheduled encounters, allergies and lab results. In the end, the results are translated into a PDF file that contains:

- Relevant previous clinical values.
- A section where clinicians can enter data that was not previously collected in a coded format.
- A section providing decision support on needed care, i.e, a reminder system to perform what is required in the necessary frequency. The aspects englobed by the care reminder system are displayed in Table 2.1, in pages 17 and 18.

Note that, unlike the other two projects that will be analyzed, in this one the focus was to describe how it was implemented with SEBASTIAN and not the results that came from the implementation of Clinical Decision Support. No information could even be found on how many patients participated on this or even if there was a control group to study if it had positive effects on health outcomes of diabetic patients. The examples provided next will be more focused on the results rather than the way they were implemented.

COMPETE II PROJECT [38]

In this study, several type 2 diabetic patients from primary care clinics in Ontario, Canada were recruited: 253 were put in the intervention group and 258 in the control group and the follow-up was of about half a year. The intervention consisted in shared access between patients and the primary care provider to a web-based “diabetes tracker”. This diabetes tracker not only provided sequential monitorization for several aspects of diabetes that need control (and can be consulted in Table 2.1) but also provided information on their clinical targets and had prioritized messages of advice. The tracker interfaced with the electronic medical record in the primary care provider side and with an automatic telephone reminder system for the patient. Monthly, the patient received reminders related to medications, laboratory, and physician visits.

Two sort of outcomes were measured and compared between control and intervention group:

- Process Outcomes: a score to measure whether the patients were taking the monitorizations with the advised frequency (the frequency was defined in accordance with American and Canadian guidelines for the management of diabetes). For example, they were advised to measure blood pressure quarterly and, if they did it, they increased 2 points in their process outcome score. The maximum score was 10.

- Clinical Outcomes: a score to assess whether the monitorizations taken were within their clinical target. For example, if blood pressure was below 130/80 mm Hg, they would add 1 point to their clinical outcome score. The maximum score was 8.

In terms of results, the intervention group presented better outcomes in both scores. The process score was of 5.19, on average, in the 6 months that preceded the study and changed to 6.52 in the intervention group and only to 5.25 in the control group. Regarding the clinical score, better outcomes were also verified in the control group, with statistically significant improvement in blood pressure and glycated hemoglobin. The patients in the intervention group also reported greater satisfaction in terms of ease of disease management, relationship with primary care providers and the quality of care.

As a final note, even though it is defined as a web-service, it is mentioned that the authors were unable to completely integrate the tracker's decision support system with the electronic health records. The monitoring data had to be entered twice to update both systems. Therefore, in terms of architecture, it is classified as a standalone CDSS, phase 1 of the 4-phase architecture described in Section 2.1.

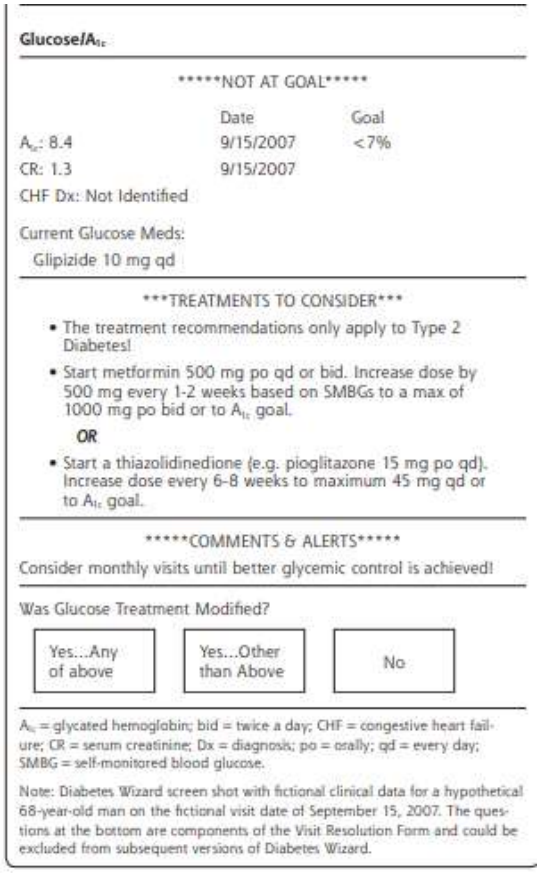
DIABETES WIZARD [39]

Diabetes Wizard is an electronic record-based diabetes clinical decision support developed mostly for the control of 3 variables: A1c, blood pressure and LDL. The study for its implementation took place in Minnesota, with the participation of 2556 patients that were randomly assigned to receive the intervention or not. The follow up lasted about 7 months.

Diabetes Wizard was developed with the premise that one of the biggest barriers to proper diabetes care is the lack of pharmacotherapy intensification at the appropriate time in patients that have not achieve their clinical goals. This could be improved by providing patient-specific and drug-specific CDS at the time of the clinical encounter.

Diabetes Wizard not only suggests specific medication changes in patient not at target for A1c, blood pressure and LDL, but also suggests changes for patients with contraindications for certain medications. Besides that, it suggests obtaining overdue lab tests, specified in Table 2.1. Finally, it suggests diminishing the time between follow-up visits when the clinical targets are not met.

The recommendations are based on clinical algorithms developed by the research time in accordance with clinical guidelines. An example of recommendation is presented in Figure 2.4.



Glucose/A_{1c}

*****NOT AT GOAL*****

	Date	Goal
A_{1c}: 8.4	9/15/2007	<7%
CR: 1.3	9/15/2007	

CHF Dx: Not Identified

Current Glucose Meds:
Glipizide 10 mg qd

TREATMENTS TO CONSIDER

- The treatment recommendations only apply to Type 2 Diabetes!
- Start metformin 500 mg po qd or bid. Increase dose by 500 mg every 1-2 weeks based on SMBGs to a max of 1000 mg po bid or to A_{1c} goal.

OR

- Start a thiazolidinedione (e.g. pioglitazone 15 mg po qd). Increase dose every 6-8 weeks to maximum 45 mg qd or to A_{1c} goal.

*****COMMENTS & ALERTS*****

Consider monthly visits until better glycemic control is achieved!

Was Glucose Treatment Modified?

Yes...Any of above | Yes...Other than Above | No

A_{1c} = glycated hemoglobin; bid = twice a day; CHF = congestive heart failure; CR = serum creatinine; Dx = diagnosis; po = orally; qd = every day; SMBG = self-monitored blood glucose.

Note: Diabetes Wizard screen shot with fictional clinical data for a hypothetical 68-year-old man on the fictional visit date of September 15, 2007. The questions at the bottom are components of the Visit Resolution Form and could be excluded from subsequent versions of Diabetes Wizard.

Figure 2. 4 Example of Diabetes Wizard Use [39]

Overall, the results of this project were positive. The intervention group patients showed improvement compared to the control group in: HbA1c (–0.26%) and in terms of systolic blood pressure, patients in intervention group were more likely to remain in target than the control group (80.2% vs 75.1%). No significant positive or negative impact on diastolic blood pressure and LDL cholesterol values in the intervention group compared to control. However, in the end, 94% of physicians in the intervention group were satisfied with the intervention and after 1 year of the study some were still using it.

Diabetes Wizard is described as EHR-based but there is no mentioning of the use of standards or interoperability to the EHRs of other clinics. It is probably in the second phase of the CDS architecture, as specified in Section 2.1.

Overall Analysis: Comparison Between the Projects

Table 2.1 sums up the overall monitorizations that the 3 mentioned projects aimed to target.

Table 2. 1 Summarization of the monitorizations of the different projects [10][42][43]

	SEBASTIAN	COMPETE II	Diabetes Wizard
BP measurement	**x**	**x**	**x**
Weight measurement	x	x	

Foot Exam	x	x	
HbA1c	**x**	**x**	**x**
Kidney Values	**x**	**x**	**x**
Cholesterol	**x**	**x**	**x**
Eye Exam	x	x	
Influenza Vaccine	x	x	
Pneumococcal Vaccine	x		
Aspirine Therapy	x	x	
Smoking		x	
Physical Activity		x	
ACE inhibitors		x	
Potassium lab results			x
Liver function (lab)			x
4-phase architecture integration (Section 2.1)	4th	1st	2nd

Overall, the different projects were CDS efforts to ensure that the following variables were being assessed/taken with the necessary frequency, according to clinical guidelines: laboratory values (HbA1c, kidney values, cholesterol, among others); vital signs (blood pressure and weight); diabetic-related exams for long-term complications (foot and eye exams); immunizations (influenza and pneumococcal vaccines); some lifestyle aspects (physical activity and smoking status) and to check if the patients were under other therapies beneficial for their condition (aspirine and ACE inhibitors, an antihypertensive drug). The projects tried to assess how disease outcomes varied when these aspects were being managed properly.

The different CDSS analyzed had different ways of action. SEBASTIAN and COMPETE II were both programmed with the frequency each event should be performed and were associated with reminders for it. COMPETE II had an additional assessment of whether the values obtained were meeting their clinical targets. Diabetes Wizard also has a component related with frequency, since it warned to obtain some overdue lab exams, but it was more directed to detect whether the prescribed therapy was not working, suggesting changes. It is also interesting to integrate those projects in the 4-phase architecture model of CDSS and to understand that they were all in different phases, which makes some of them more interoperable than other.

2.3.3 Non-Knowledge Based CDS for Clinical Management

In this section, the goal is to focus on how Artificial Intelligence (AI) and Machine Learning (ML) algorithms are contributing to decision support for disease management and risk prediction in T2DM.

AI represents a shift in paradigm since it is changing the way diabetes is prevented, detected and managed [40]. AI methods, combined with some of the latest technologies, including medical devices, mobile computing and sensor technologies can create better management services to deal with the disease [41].

Something important to consider when dealing with diabetes is that it is such a prevalent disease in the population and led to such an extensive research, that huge amounts of data have been generated. The application of Machine Learning and data mining techniques is an efficient approach to use these large available datasets related to the disease, to extract knowledge for the diagnosis, its management and other aspects [42]. Data Mining is a part of AI that can discover useful information within a database, namely interesting relationships between the items and patterns in the data set.

Machine Learning and Data Mining have been applied in diabetes for the following [42][43]:

- Disease prediction and diagnosis;
- Interpretation and prediction of blood glucose levels: use of data mining techniques to identify which might be the best predictors, rules and trends associated with glycemic control and also the identification of factors that can influence it. Ideally, the goal is to detect adverse glycemic events before they occur.
- Prediction of diabetes associated complications: As explained in Section 2.3.1, one of high glycemic values can lead to harmful health conditions, namely micro and macrovascular complications. Several machine learning and data mining techniques have also been used for the prediction of one or several of these complications.

2.3.3.1 Prediction of Diabetes Microvascular Complications

As mentioned previously, one of the greatest concerns associated with diabetes is the fact that uncontrolled hyperglycaemia can cause harmful consequences in small bloods vessels, defined as retinopathy (affecting eyes), nephropathy (affecting kidneys) and neuropathy (associated with neural damage).

There is some controversy regarding the appearance of complications: their pathogenesis is not fully understood and sometimes it is difficult to understand why it happens to some patients, that might even have their glycaemic values more controlled, and not others. Additionally, early recognition of complication signs helps to prevent its progression [44]. This is the premise for the use of machine learning techniques to predict the appearance of microvascular complications in diabetic patients. The literature that was found addresses this issue in two different ways: one that focuses on identifying the risk factors that influence the appearance of complications, using simple logistic regression to find associations [35][45]. The other way already has a defined set of features defined as risk factors for complications and the goal is to study which machine learning algorithm is more adequate for prediction [46][47]. Both ways can be interesting in the context of this work and will be analysed, as well as the main results

that were obtained. Table 2.2 summarizes all the variables that were under study analysis under the four different studies considered.

Table 2. 2 Variables Studied for the Influence of Microvascular Complications in the Different Studies

Input Variables	[35]	[45]	[46]	[47]
Age	X	X	x	X
	(50-64; 65-74; 75+ years)	(per 10 years increase)		(25-35;36-45; 46-55; 56-65; >65 years)
Gender	X	x	x	x
	Female(F)/Male(M)	F/M		F/M
Education level	X	X		
	(primary/less; secondary; third/higher)	(less or higher than high school)		
Duration of Diagnosis	X	X	x	X
	(0-4 years; 5-9 years; >10 years)	(less or higher than 10 years)		(<4; 4-10; >10 years)
Ever smoked	X		x	
	(yes/no)		(yes/no)	
HbA1c levels		X	x	
		(per % increase)		
Blood Glucose Level				X
				(<200; 200-400; 401-600 mg/dL)
Physical activity	X			
	(low/medium/high)			
Hypertension	X	X	x	X
	(history of hypertension: yes/no)	(history of hypertension: yes/no)		(<120; 120-139; 140-159; 160-180; >180 mm Hg)
Cholesterol	X	X		
	(history of high: yes/no)	(LDL levels per mmol/L increase)		
Body Mass Index (BMI)		X	x	X
		(per 5kg/m2)		(<18.5; 18.5-24.9; 25 to 29.9; >30)
Creatinine		X	x	
		(>106 vs <106)		
Family History of DM		X		X
		(yes/no)		(yes/no)
DM treatment		X		
		(insulin vs drug)		
Algorithms Used	Log-binomial regression	Multivariable logistic regression	Logistic Regression, Naïve Bayes, Support Vector Machines, Random Forest	k-means clustering; Naïve Bayes; Decision Tree

Note that, in the first study [35], the goal was to assess underlying factors for complications in general, while the others focused in understanding what was beneath the appearance of each complication individually. The third study [46] is related to the MOSAIC project, an EU-project with the goal of changing the way diabetes is handled to an approach more focused on patient follow-up and prevention of worsening of the condition, rather than treatment of acute complications [48]. The features that were used as inputs were mentioned, but not the way they were handle (for example how the age was organized in categories and delivered to the algorithm was not explained). This study also aimed to assess appearance of complications in different time horizons: 3, 5 and 7 years, building different prediction models for each. The 3-year horizon model provided the best results.

Overall, the goal of the different studies was to use the ML algorithms to gain knowledge about the underlying causes of complications, in some cases generating "rules". An example of a rule is "if hypertensive and female then retinopathy" [47], meaning that being hypertensive and female might increase the risk of developing that complication. The main conclusions that each study drew regarding both the underlying causes and the adequacy of the different algorithms to make predictions are presented in Table 2.3.

Table 2. 3 Main Conclusions of the Different Studies on Microvascular Complications [50][52][53][54]

Study	Main Conclusions on the Variables Contributing to Complications	Considerations on the Algorithms
1 [35]	-Participants with type 2 diabetes for more than 10 years were twice as likely to have complications than those who were diagnosed for less than 4 years. - patients who had ever smoked were one and a half more likely to develop complications that those who never did it. -Risk is lower in patients with 3rd level of education compared to 1st. -High physical activity means the patient is half less likely to develop it compared to low physical activity. -Hypertensive patients are one and a half times more likely to develop it. Note: In this study the goal was to find underlying factors to develop any complication. The others studied each complication individually.	-
2 [45]	- HbA1c was the only variable significantly associated with all 3 complications. (-Smoking needs further exploration) -Age at onset of diabetes is important (earlier age, higher severity) -Other risk factors: Hypertension, patient age, family history of diabetes, duration of diabetes, insulin use, BMI -Retinopathy influenced by: age, **duration of DM**, **HbA1c**, DM treatment, creatinine, **hypertension,** creatinine -Nephropathy influenced by: education, **HbA1c**, DM treatment, creatinine, **hypertension,** creatinine -Neuropathy influenced by: **gender (female)**, **duration of DM**, family history, **HbA1c level**, **BMI**	-
3 [46]	-**HbA1c** is a risk factor for all complications -**Duration of diabetes** and **BMI** are risk factors for **retinopathy** and **neuropathy** -**Hypertension** is a risk factor for **retinopathy** and **nephropathy**	SVMs and RFs had the highest performance. However, LR was adopted. This was due to the fact that the first two are harder to interpret (black boxes), whereas LR is of easy interpretation and more user friendly for clinicians, supporting some of the

		features they already use for graphical visualization, making it the most suitable
4 [47]	The following risk factors were identified for each microvascular complication: -Retinopathy: female and **hypertension** -Neuropathy: **female** and **BMI** -Nephropathy: Duration of Diabetes	-clusters of 6 work well for grouping the population, but there is no knowledge of the desired class -Naïve Bayes and Decision Tree were used together to define the rules

2.3.3.2 Non Knowledge-Based CDS: Prediction of Hypoglycemic Events

As discussed in Section 2.3.1, hypoglycemias can come without warning and have very negative consequences, which is why efforts have been put in building algorithms that might prevent them in advance.

Several projects have applied ML to predict blood glucose, in general, while others focused on hypoglycemia in specific. The focus here will be on hypoglycemia since it can be lethal, causing fear in the life of diabetic patients [49]. Mechanisms to predict it in advance, giving time to take action in order to prevent worse consequences are a source of safety in the life of diabetic patients [50]. This is especially important considering that some hypoglycemic situations can be practically asymptomatic and unrecognized, giving the patients barely time to take action (like taking fast-acting CHO or stopping anti-hyperglycemic/insulin medication).

This type of prediction escalated with the increased use of continuous monitoring devices (CGMs), especially in type 1 patients, as opposed to simple SBGM [51]. CGM includes wearing a small device underneath the skin, that measures glycemia continuously throughout day and night [52], while SBGM includes using a fingerstick to measure glycemia at certain times of the day [53]. The difference is that the former can collect data every 5 minutes and is able to detect glycaemia fluctuations caused by intense physical activity or food/insulin, whereas the latter has limited readings throughout the day.

Increased use of mobile health, sensors (including CGM sensors) and apps has allowed automated and continuous collection of diabetic data, that can be used in ML algorithms to find rules for prediction of such events. Though physicians can, obviously, predict a certain risk of hypoglycemia, there are certain underlying factors that are difficult to understand and computers can help with that. If such prediction mechanisms were associated with a system that would inform/warn the patient and guide him/her in what to do to prevent its occurrence, a lot of value could be provided to the life of diabetic patients [50].

Regarding the data collected, blood glucose is by far the most important for this task. However, some works use time series data of other values: insulin or other medications, CHO and physical activity. In general, the accuracy of prediction increases when more data is used [50]. Another factor to consider is the prediction horizon: prediction should be given when there is still enough time to act (eat CHO or stop insulin intake), but the earlier it happens, the less

accurate it is, so there should be a trade-off [49]. Most studies have a prediction horizon of 15 minutes, but studies can range from 5 minutes to 1 week.

Regarding the type of algorithms, in the past physiological models were the most common. These usually have 4 modules: meal dynamics, insulin dynamics, exercise models and glucose dynamics [49]. They are usually accurate, but they require a considerable knowledge of physiological constants. Data-driven models, on the other hand, are becoming more popular, using pattern recognition techniques. There is, however, a challenge of inter and even intra subject variability. Thus, two different approaches can be taken: individualized models, or general models, trying to make good enough predictions that can fit everyone.

This task can either be a classification or a regression task. The former is a binary problem (yes/no for the occurrence of a hypoglycemic event in a certain period of time) and the latter predicts the value of glycemia the patient will have and, if it is in the hypoglycemic interval, an alert is issued. Overall, the most used ML algorithms for these tasks include Random Forests (RFs), Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) [50].

Hypoglycemia Prediction Using ML Models for Patients with Type 2 Diabetes

Predicting diabetes in type 2 patients is more complicated, since these patients sometimes just check for their blood glucose once or twice a day using SGBM, which does not provide much data to train an algorithm.

However, there are examples in literature that just used SBGM data and obtained good results. There was a study that used data from another trial that collected SBGM values from 163 type 2 patients over a 1-year period. They made the prediction of hypoglycemia a classification problem, and the goal was to predict whether hypoglycemia would occur or not in the eight day, taking into account SBGM data from the seven previous days [54]. Two approaches were studied:

- First approach: using only Blood Glucose (BG) data, the goal was to predict a hypoglycemia in the eight day (throughout the entire day, not a specific hour). They used the values of BG and their timestamps. Several methods were tested and RF and SVMs got the best results. The authors ended up choosing the RF and compared their performance with the predictions of real human doctors. RF had a higher sensitivity (proportion of actual positives for occurrence of hypoglycemia predicted as positive) than humans (91.7% vs 52.7 ± 16%), but a lower specificity (proportion of actual negatives predicted as negative) than the latter (69.5% vs 79.8 ± 5%).
- Second approach: using BG data, together with medication data (timestamp of administration, class and dosage), which was the only data they had access to from the study besides BG, the goal was to predict the hour in which the hypoglycemia would occur in the eight day. The specificity increased by around 30%. Both the specificity and sensitivity were around 90% in this data set.

The authors concluded that these models have potential to be employed in real life context in an automated way to guide intervention and prevent hypoglycemia. Besides that,

adding medication data increased the performance of the models. Further studies with more data, such as nutritional and physical activity data, could provide very favorable results and they should be explored.

2.4 Daily-Life Management of Diabetes

By daily-life management, one encompasses all the lifestyle variables that have to be managed daily by the patient. In Section 2.4.1, all the aspects that the T2DM patient needs to control will be explored and in Section 2.4.2 different CDS solutions will be analyzed.

2.4.1 General Considerations on Lifestyle Management

To control the disease in the best way possible, it is important to maintain an adequate lifestyle, since sedentary living, stress, nonadherence to medication, lack of regular medical examinations and bad habits can lead to harmful outcomes for these patients [41].

A good control of glycemic values can help in the prevention of micro and macro vascular complications of diabetes. However, it is estimated that around 50% of adults with T2DM are not meeting the desired targets. The blame can be attributed to inadequate blood glucose (BG) monitoring; nonadherence to medication or inadequate lifestyle, namely in terms of nutrition and physical exercise; lack of proper education or access to the right healthcare provision [55].

Telemedicine, which is the provision of remote patient care or even appointments via telecommunication technologies [56] is a way to tackle this issue, with the proposition of providing better education, compliance, quicker access to providers and monitoring of BG readings and complications. Many applications exist for this purpose, but with limitations. Most are only meant to be used by the patient, without participation of any healthcare professionals, and many others do not take into account important diabetes-related influent factors like physical exercise or cannot be easily personalized [55].

Current technologies are handling these issues, by having some sort of monitoring component for the physician that accompanies the patient. Many are also enabling the monitorization of data related to diet, physical activity, medication, glucose measurements, etc, and combining that data with tools that make quality decisions for the patient (decision support systems), with the ultimate goal of enhancing therapeutic outcomes. The control of data related to those aspects is important for diabetic patients, since it can improve the overall outcomes of the disease and the reasons for this are explained next:

- **Food intake** [57]**:** An adequate and healthy diet, with low caloric and carbohydrate intakes, has the potential to keep an adequate metabolic function and prevent further complications. The goal with the right nutrition is to: keep blood sugar levels in normal ranges for the prevention and control of diabetes complications; reduce the risk of macrovascular complications by keeping an adequate lipidic and lipoproteic profile and

diminish the risk of cardiovascular disease by keeping adequate blood pressure levels, associated with the ingestion of salt.

- **Physical Exercise** [58]**:** Aerobic activity, as simple as 30-minute walks daily, can have a very positive impact in the patient with type 2 diabetes. It improves blood glucose control since it diminishes insulin resistance (by stimulating its production and its transport to the cells) and improves triglycerides and blood pressure levels, thus reducing cardiovascular risk factors. Finally, it contributes to weight loss and improvement of well-being.
- **Medication Management** [59]: Diabetic patients take anti-hyperglycemic medication to help maintain blood glucose levels as normal as possible. When this medication is no longer enough, as the disease progresses, they start taking insulin doses.

As a disease characterized by the poor regulation of glucose in the blood, daily measurement of these values and occasional measurement of Hb1Ac in laboratory analysis are important to control the disease and evaluate how well it is being managed. For daily measurement, either SBGM or CGM are used, according to the doctor's decision.

In Portugal, even though many apps can be installed by users to help them manage certain aspects of their lifestyle, those are not integrated with healthcare professionals. According to one of the nurses that provided insights to the project, one of the most advanced practices regarding blood glucose levels monitorization takes place in Cascais Hospital. Diabetic patients insert their blood glucose measurements in specific platforms that can also be consulted by nurses. However, they are not associated with any sort of alarmistic when there are urgent values and they are also not associated with the patient's EHR: the nurse has to open that "standalone" platform to consult the values.

Regarding nutrition and physical activity, in the nurse follow-ups, the script that is usually followed assesses the current habits of the patient and then there are moments for advice, both in conversation or the nurse can even write "prescriptions" or give brochures with advice. But there is no digitalization whatsoever for consultation and, since these changes in lifestyle can improve the disease outcomes as much as medication, for instance, there might be room to handle these aspects in a different way.

2.4.1 Examples of Knowledge and Non-Knowledge Based Systems for Lifestyle Monitorization

In the following sections two projects will be discussed: METABO and glUCModel. Both are examples of lifestyle monitoring for diabetic patients, integrated with healthcare professionals, and both incorporate knowledge and non-knowledge based CDS. Knowledge-based is in the form of simple alarmistics and non-knowledge based is in the algorithms each incorporates for prediction of adverse glycemic events.

2.4.1.1 METABO [11][60][61]

METABO is an EU-funded project, with the aim of assisting diabetic patients, of both types 1 and 2, in the management of their disease on a day-to-day basis. By controlling how different factors and real-life situations affect blood glucose concentration, it can provide structured information and therapeutic decision support to patients and care givers. METABO has a professional and a patient end-user profile.

The patient monitoring device works in a mobile-phone-based interface and allows patients to check the treatment that was prescribed in terms of medication, diet, physical activity, frequency that glucose should be measured and even goals and education that should be achieved. The patient monitoring device also has an important function of data collection, both by acquiring it from a device or by manual input. It collects data related to food (manual input of carbohydrate intake per meal) and medication intake, physical activity (using a pedometer) and glycemic values, either discrete values using a fingerstick or with CGM [60].

Based on the treatment plans and the inserted data, it has an automated system of feedback that provides alerts, reminders and motivation messages, which can be defined as knowledge-based CDS. The feedback system has 3 layers (L1, L2 and L3) [60]: L1 is for urgent and health critical warnings, for example if a critical glycemic value is detected. They are generated within the platform and must not fail to be delivered. L2 is for reminders related to patient behaviors in a period of 3 to 7 days. For example, to remind them to measure their blood pressure if they are supposed to do so weekly and have not introduced that data yet. Finally, L3 is related to the GOAL, a module that defines goals the patients must meet. If the patient has the goal of losing a certain amount of weight in a determined period of time and is not accomplishing it, the system can send a message to remind him to follow his exercise and meal plan.

Education about the disease is very important for the patients, since there are so many variables to handle. With this in mind, METABO was developed with two features for education [62]:

- ENCYCLOPEDIA: when users want to search for a specific piece of information, they can do it in this database.
- QUIZ: when users just want to improve their general knowledge, QUIZ works like a game. Not only it is more interactive, but a score can also be given to estimate the knowledge of the user.

Both features cover educational content related to diet and nutrition, hypoglycemia episodes, glycemia, insulin, food intake, physical activity, blood pressure and weight.

On the side of the physician, there is a Control Panel, which allows for the definition of the patient's prescription; to enter data generated in hospital context and to retrieve information about the patient's behavior and health situation according to the data inserted and collected by the latter. It also receives feedback from the METABO decision support system and allows direct message contact between the doctor and the patient.

Data produced within the system is mined at different levels to be used for distinct applications to provide decision support: non-knowledge based systems [11][61]. The most

interesting example is the creation of an individualized metabolism model for each patient that allows the prediction of glycemic excursions (hypo/hyperglycemic events).

The (individualized) Metabolic Modelling System, described in Figure 2.5, runs in the patient monitoring device and models patient's metabolism in order to predict hypo/hyperglycemic events, providing decision support in the form of short and long term alerts to the patient. It can provide immediate critical feedback or suggestions for modifications of diet, activity or medication in order to avoid future critical events.

The system takes as input the insulin and food intake, glycemic data, physical activity and other clinical data. The data is preprocessed and then goes to compartmental models that consist in an insulin regulatory system and a predictive model of glucose (physiological model). These models simulate: the ingestion and absorption of carbohydrates (meal model), the absorption and pharmacokinetics/pharmacodynamics of subcutaneously administered insulin (insulin model) and the impact of exercise on the glucose-insulin metabolism (exercise model). Additionally, support vector regression is used to provide individualized glucose predictions.

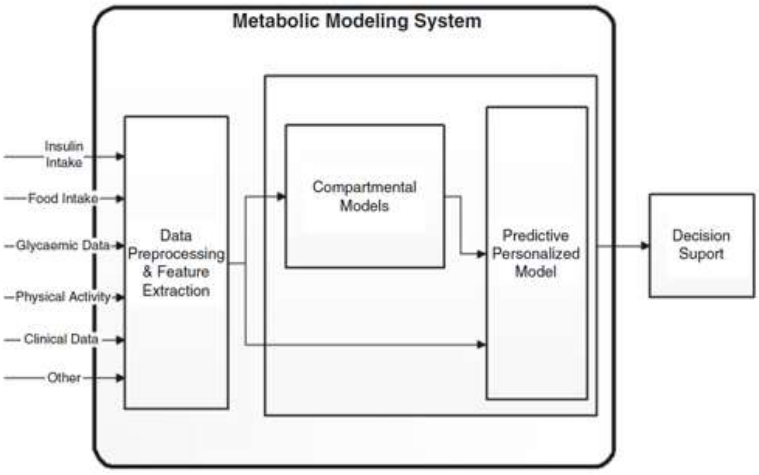


Figure 2. 5 Architecture of METABO metabolic modelling system [11]

2.4.1.2 glUCModel [62]

This project, developed in the Universidad Complutense de Madrid, was also created with the goal of helping diabetic patients and physicians to improve the control and management of the disease. It is especially directed to insulin-dependent patients, both type 1 or 2. But it can also be useful for type 2 patients only taking oral anti-hyperglycemic drugs.

It is a telemedicine system functioning as a web application, that can be consulted and used by both the patient and the doctor. It is constituted by 5 modules:

-**Data Interface**: It allows the patients to check and update all the relevant data, namely data related to: glycemia levels, injected insulin, ingested carbohydrates, physical activity, weight measurements and records of important tests for diabetes control. The data can either be introduced automatically, from the glucometers for example, or manually. Doctors can also access the data interface and track the patient's evolution

-**Database**: Responsible for the storage of the user's information (glycemic levels, insulin intakes, physical activity, weight, and other medical tests) and also the data that the recommender system, described next, needs in order to function.

-**Recommender System (g-RS)** (knowledge-based CDS): it evaluates the patient's data in the database and creates recommendations or suggestions, through Case Base Reasoning (the recommendations are mostly not based on rules, even though there are initial rules containing initial knowledge, but on previous cases that the system has seen). There are two sort of recommendations: one related to education, because the patient takes an initial test and the system recommends which topics need further study in e-learning. The other sort of advice is related to the data inserted in the system. g-RS can recommend changes in lifestyle or, again, additional study of a topic. The recommendations are sent via e-mail.

-**e-learning:** it aims to provide patients with all the necessary information related to the disease. It contains PDF files for each relevant topic (diet, complications, insulin tuning, among others). There are also tests to evaluate the level of knowledge over a certain topic.

-**Glucose Model Module** (non-knowledge based CDS)**:** Through evolutionary computation techniques, using the data stored in the database, the system can customize a blood glucose model for each patient by obtaining an equation describing glycemia as a function of a set of parameters that best fit each user. This model can be used to forecast short, medium and long-term glucose levels.

2.5 Current Glintt Software Solutions

Being a Glintt project, it made sense to make an analysis of some of the solutions already developed and in the market. This way, when proposing something new, some knowledge will exist regarding what required functionalities can already be met easily by existing solutions and do not need development from scratch.

Globalcare [63]

Globalcare is a hospital management software that aims to respond to the different needs of healthcare institutions, supporting and orienting the activity of healthcare professionals. To meet the needs of the different teams, it covers virtually all hospital activities: outpatient and inpatient care, emergency, day hospital and specific areas including diagnostics and therapeutics, drug circuit, patient safety, etc. Globalcare is also able to be integrated and communicate with other systems.

One of the ways to assure Patient Safety is through **Clinical Pathways**. Clinical Pathways are an implementation of clinical protocols and guidelines specific to each care setting. They identify risk quicker and anticipate intervention. Nowadays, Critical Pathways, for emergency context, and Care Pathways, for inpatient context, are already implemented. Critical Pathways consist in "fast tracks" and "vias verde". For example, in the Sepsis Via Verde, if certain signals of sepsis are detected, there is a direct activation of certain protocols for exam's request

and antibiotic prescription. It also activates the constant monitorization of specific values. One of the goals of Glintt is to implement Clinical Pathways for chronic disease management, namely in the context of the ConnectedHealth project.

Another interesting module of Globalcare is **Bedside Nurse**. It is directed at nurses, allowing them to manage their tasks in a remote and ambulatory way. They can have real time access to all clinical information regarding the patient and they can register information easily and quickly. Bedside Nurse is also an interesting solution for chronic patients in the context of ConnectedHealth, if there are outpatient visits, for example.

Viewer [64]

Viewer aims to be the "focus that health needs". In an era where data is digitally dispersed across so many different platforms, Viewer is designed to aggregate and link all the relevant data that the person that is using it needs, in an user-centric approach. This way, if the user is a doctor, by having the necessary data displayed at the right time, in the right place, the doctor can have more time with the patient and focus on what is really important. Viewer is highly interoperable, using the FHIR standard, and allows integration with all the relevant applications along the care path. Initially, it is being developed with doctors as end users in mind, but eventually it intends to be scaled to nurses and even the patient.

Sifarma [65]

Sifarma is the software leader in the market for management and customer service in the Portuguese pharmacy sector.

Nowadays, there is a risk assessment of diabetes called Findrisk that is already performed in the community pharmacies. If a patient is considered to be in high risk of DM through that scale, Sifarma communicates with the Health Data Platform of the Ministery of Health and those patients are referenced for a primary care appointment.

Another interesting feature of Sifarma is the existence of refill alerts. By simple calculations, it can predict when the medication box ends and the patient needs to go to the pharmacy. For example, if it is a box of 60 pills to be taken twice a day, it will send a refill alert every 30 days via e-mail or text message.

HELIOS [66]

Helios is an interoperability platform, allowing hospital's information systems to integrate and interoperate. It also performs semantic integrity of HL7 messages and real-time integrations monitoring.

2.6 The Community Pharmacy in Chronic Disease Management

A systematic review of the relevance of the interventions of clinical pharmacists in the management of T2DM has the premise that, even though there is a tighter control of blood glucose and cardiovascular risk factors nowadays, the recommended targets are not being met [67]. To accomplish that, new healthcare models, with integrated teams and an active participation of the patient, should be implemented. They also consider that the role of pharmacists is still undervalued.

One of the goals of ConnectedHealth is to integrate the pharmacies in the process of care for the patient as much as possible, since they have a proximity to the population that hospitals and clinics do not.

In that review, integrated team efforts with pharmacists from different healthcare systems were analysed: nine from North America, five from South America, three from Europe, one from Africa, and twenty-one from Asia. Overall, they included medication management, educational interventions and referrals to other healthcare professionals and services. The results were positive: almost all studies showed improvement in HbA1c, blood glucose, blood pressure, lipid profile and BMI in the intervention groups [67].

In Portugal, it was estimated in 2017 that pharmacists contribute to a gain of 8.3% in the quality of life of the general population in their roles of: providing pharmacotherapeutic counselling, monitoring therapy outcomes, contributing to decrease the risk for unintentional adverse events and preventing harmful drug interaction. It is estimated that, in the future, there may be an additional increase of 6.9% [68]. These future services might include higher integration in primary and secondary care services. In primary care services it might include a higher participation in the care plan of the patient and in screening for diseases such as diabetes, especially in rural areas, where there is not so much covering by hospitals and primary care units. In secondary care services, it could include dispensing medicines currently dispensed exclusively at hospital pharmacies or the adjustment of dosing regimens.

Although the involvement of the pharmacy in the process of care seems to be increasing, there is still an issue with their digital communication: there is still no solution able to aggregate information collected from pharmacy and hospital systems. The ConnectedHealth project identifies this as a market gap and the proposed solutions intend to bridge it, in order to increase interoperability and integration between both.

2.7 Value-Based Healthcare for Diabetes

The **International Consortium for Health Outcomes Measurement (ICHOM)** conducted the first multinational effort to recommend a standard list of the outcomes that matter the most for type 1 and type 2 diabetic patients [69]. The authors stated that the current efforts focus mostly on improving HbA1c values, but that does not necessarily always translate into better health status all the time. Therefore, through a Delphi method, they tried to understand together

with patients what were the most important outcomes to measure, in their perspective. They approved the following outcomes, divided into different categories:

- Diabetes Control**:** The working group defined the following variables should be measured regularly: HbA1c (every 6 months) and blood pressure, lipid profile, BMI (annually). The targets were not specified.
- Acute Events: They defined that the frequency of the following events should be reported: diabetic ketoacidosis (more common in type 1), Hyperosmolar Hyperglycemic Syndrome and Hypoglycemia.
- Chronic Complications: to check annually for conditions related to long term micro and macrovascular complications, mostly to check for its presence. The conditions include: autonomic neuropathy, peripheral neuropathy, Charcot's foot, lower limb ulcers, peripheral artery disease, ischaemic disease, chronic heart failure, chronic kidney disease and dialysis, cerebrovascular disease, periodontal health and erectile dysfunction. Some specify the guidelines that should be followed.
- Health Services: They considered it would be important that some aspects of hospital utilization are measured: number of hospitalizations and emergency room attendances per years; number of discharge diagnoses of some diabetes related conditions (cardiovascular acute kidney injury, foot and lower limb-related complications, acute metabolic diagnoses,…). They also considered it was important to assess financial barriers and difficulties in paying for healthcare.
- Survival: The group considers there is an issue in the data recorded in death certificates. There should be an effort to report the cause of diabetes and attribute the primary cause to diabetes when that is the case.
- Patient-Reported Outcome Measures (PROMs): They identified a set of key domains that should be captured using PROMs and decided to prioritize assessing annually well-being, depression and diabetes related emotional distress. In order to do so, they selected two generic and a diabetes specific tool to do these "measurements":

  -WHO well-being index, to measure well-being in general
  -Patient Health Questionnaire-9 (PHQ-9) to measure depression
  -Problem Areas in Diabetes (PAID) scale, that measures diabetes-specific emotional distress and problems specific of diabetics.

3. Methodology

The method that was carried throughout this work can be inserted in the spectrum of Action Research.

In Action Research, the researcher and the client work together to "diagnose" a problem and to develop a solution for that problem [70]. In the case of this work: the problem is related to what aspects of diabetes management can be improved and proposed solutions for it. The client encompasses the diabetic patients and the healthcare team that follows them, that was represented by a Lusíadas team of 2 nurses, 1 endocrinologist and 1 physician from internal medicine. The researcher is Glintt and the other companies working together for the ConnectedHealth project. This work aims to give a contribution to that research.

The Action Research approach followed a cycle of four steps: plan, act, observe and reflect, described in Figure 3.1:

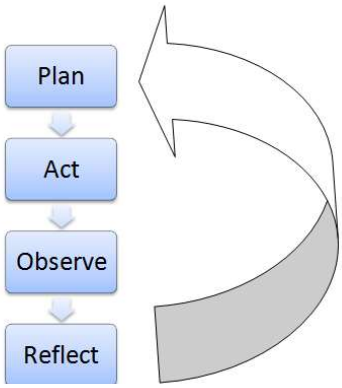


Figure 3. 1 Steps of the Action Research Cycle [71]

The knowledge-based and non-knowledge based CDS solutions, that will be proposed in Chapter 4, had each their own cycles, only sharing the first step of the first cycle (initial planning). Furthermore, it can be said that the proposal of the knowledge-based solution followed two cycles, whereas the non-knowledge only had one. This is explained in greater detail in Table 3.1.

Table 3. 1 Description on the Steps Followed in the Action Research Method

Nº of Cycle	Phase of the Cycle	Knowledge-Based Solution Proposal	Non-knowledge Based Solution Proposal
1	**Plan**	This phase consisted in understanding the main goals of the project, by reading the necessary documents and having internal meetings. That was followed by intense research, presented in Chapters 1 and 2. The research allowed to: -Get acquainted with the disease (etiology, symptomatology, possible consequences and prevalence); understand how it is normally managed, both clinically and in terms of lifestyle; -Understand how CDS is being used to handle the disease. METABO and gluCModel can be defined as benchmarks here, since they already have a very integrated approach for its management. Additionally, it was also analysed how ML algorithms are being used to turn information into knowledge, one of the goals of the project.	

		This research was crucial to develop the rest of the work, which explains why Chapters 1 and 2 are so extensive. In this phase, one can also include the bi-weekly meetings that took place with Fraunhofer, where some of the findings from literature were presented and discussions occurred regarding which tools and technologies would be necessary in the context of the disease.	
	Act	After the insights provided by the research, there was already enough knowledge of the disease to propose a solution for the project. That was initially done in the form of an use case, presented in Section 4.1.1 (note that the version presented is much more complete that the initial one). The use case described how the different stakeholders (patient, doctor, nurse and pharmacist) would interact and benefit with the system to be developed. It was the basis for the entire solution	Considering the research that was done, data mining was proposed to handle two aspects of the disease: prediction of microvascular complications and prediction of hypoglycemias. The algorithms and input variables were proposed for each. A "proof of concept" was developed for the prediction of microvascular complications: a synthetic dataset was built, consistent with literature, and a Python library with SVM and RF (the proposed algorithms) was used to analyse the data.
	Observe	This phase was marked mostly by a validation meeting with the Lusíadas team, where the use case was presented. They gave their feedback and suggestions on what should be included or excluded.	The results obtained from the previous phase were observed.
	Reflect	The previous meeting not only had the goal of understanding if the proposals made sense clinically, but also if they brought value in the perspective of a clinical team. Therefore, there was a reflection of the considerations they made and how they could be incorporated to improve the use case.	In this phase, relevant conclusions were drawn from the results, together with how the data analysis process could be applied to the real data planned to be collected.
		Knowledge-Based Solution Proposal: Continuation	
	Plan	Having the use case overall validated in the previous cycle, it was necessary to understand and plan how it could be brought to life. In this phase, there was another research, but this time of existing technologies, inside and outside Glintt, to be incorporated. It was also necessary to understand what aspects the solution requires had to be developed from scratch.	
2	**Act**	This phase consisted in building the functional requirements: detailing which aspects would have to be put into practice and the function of each component of the system. A dataflow diagram was also built, explaining how data is supposed to flow within the system.	

	Observe	This phase consisted mostly in internal feedback, to understand within Glintt if the functional requirements were feasible. Additionally, with the functional requirements and the dataflow, the designer of the team had enough material to build the initial mock-ups of what the solution would look like, which will also be presented. It was important to observe if all the requirements made sense when put in the context of a software application.
	Reflect	After looking at the initial mock-ups and having a more practical perspective of the solution, further refinements were made for the functional requirements, in order to improve them.

Table 3.1 and Figure 3.2 present the scheme of the macro plan followed to arrive to the work that will be presented. It is worth mentioning that, in a micro level, for every aspect of the proposed solution (each aspect of the use case, each functional requirement, the mockups of the app, etc), the same cycle was followed.

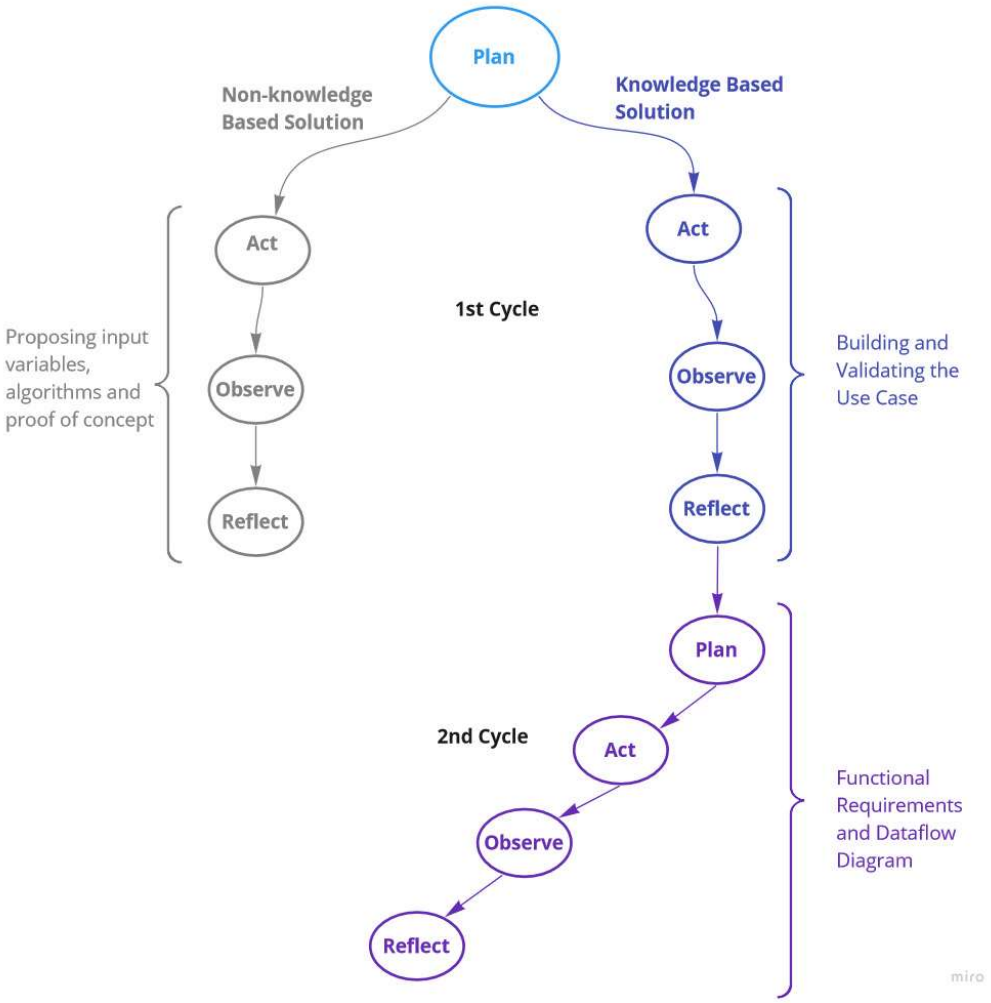


Figure 3. 2 Scheme of the Action Research Cycles Followed

4. Proposed Solution and Discussion

4.1 Knowledge-Based CDS: An Integrated System for the Management of Diabetic Patients

4.1.1 Presentation and Discussion of the Use Case

The use case, basis for the solution, was built in the form of a patient journey, describing a patient named António that is diagnosed with diabetes and is referred to use the proposed solution. The patient journey is described in the following steps and then discussed:

Use Case: Patient Journey

1) *António has been feeling very thirsty, with an excessive urge to urinate and tingling in some parts of his body. He goes to his pharmacy to ask for help. A risk assessment for diabetes is performed at the pharmacy. According to the score obtained, which was high, the pharmacist referenced him to a specialist doctor from Lusíadas Hospital.*
2) *He goes to a first medical appointment in which the doctor requests additional laboratory exams to confirm the diagnostic. After getting the results, the diagnostic for type 2 diabetes mellitus is confirmed. The doctor also identifies a plan of action, referencing António to other specialists for further control of the disease, including an ophthalmologist and a podiatrist. He was also referenced to a specialist nurse, with whom he has regular follow-up appointments. The doctor schedules the further visits and laboratory exams.*
3) *The doctor also gives António a prescription, with a medication component but also a non-medication component, related to attitudes. The attitude prescription can also be performed by the nurse. Attitudes concern how he has to behave to manage the disease, in terms of measurements (how often to measure glycemia for example), physical activity and nutrition.*
4) *The doctor has a conversation with António, pointing out that it is important that he starts having more control over his health, advising him to use an app for that process. António installs the app and allows all health parties involved (hospital and pharmacy) to have access to his data.*
5) *António has a glucometer with which he takes measurements of his blood glucose (SBGM), in the frequency previously defined by the doctor, and the values are transmitted to the app. Ideally, they are transmitted automatically but, if there is an error, it can be done by manual input. He receives urgent alerts if a harmful hypoglycemic or hyperglycemic event is detected. He also receives alerts if a deterioration of his condition is detected over time.*
6) *In one of the follow-up appointments with the nurse, she alerted him to pay attention to the symptoms of hypoglycemia (including sweating, dizziness, palpitations, among others). If he starts feeling any of these, he should immediately take a blood glucose*

measurement. If it is a hypoglycemia, he should follow the normal plan of action (eat fast-acting carbohydrates and measure again in 15-20 minutes). The outlier inserted in the app also alerts the nurse about what is happening, and this professional can contact António, to make sure he is handling the situation properly. If António is out of reach, the nurse can call an ambulance to his home, if there is a suspicion a worse consequence of hypoglycaemia affected António.

7) *If the healthcare professional realises that the normal values of a specific patient are different from those originally defined as general targets, he can adapt and define personalised targets for the patient directly in the system. This is a step towards a personalised medicine and can eventually diminish a possible alert fatigue from the system.*
8) *The app has a record of the meals and the time they are meant to be taken in a food diary. When António has doubts about what to eat, he can consult the healthy suggestions.*
9) *António usually takes 30 minute daily walks, because he was told by his doctor how important it is to stay active, and wears a "fitness bracelet" to monitor his physical activity, as prescribed by the doctor. This bracelet also communicates with the app. The bracelet monitors heart rate, giving an immediate alert if it detects arrythmias.*
10) *António can consult his prescription, both related to drugs and attitudes, on the app. He also has summaries of what actions he has to take throughout the day (regarding medication, measurements, nutrition, physical activity, next doctor visits,…), to facilitate his life. Additionally, he can leave daily notes in the app. For example, he can say that in a certain day he had a birthday party and committed some nutritional excesses. That way, the nurse is aware of this if there are hyperglycemic outliers in that day.*
11) *He can also consult what medication he has to take and at what hours every day. He has to put simple checks in the app when he takes it. If checks are missed, at the end of the day an alert is issued reminding him of the importance of taking the medication properly.*
12) *The doctor receives an alert when the medication prescription ends to issue a new one or, if he considers necessary, schedule a new appointment with the patient.*
13) *António is reminded in advance of the appointments and exams he has to attend, which were scheduled in the last medical visit. If he skips them, the nurse that follows António receives an alert informing he did not go to his appointment or did not perform a certain exam/medical procedure around the time he was supposed to and can communicate with him.*
14) *In the case of laboratory analysis, the doctor can, in advance, prepare the laboratory request and it becomes available in the app.*
15) *When the exam is performed, it can be integrated by the system and received in the app as well. If something is considered to be an urgent outlier, an alert is issued to the specialist to identify an action plan in advance: anticipate the appointment; prescribe SOS medication, etc.*

16) *His nurse should remotely monitor and regularly check if everything is okay with António, namely by checking the evolution of his monitorizations, intervening when necessary.*
17) *There is a patient/nurse communication channel. The latter can send a message when he/she detects something wrong with the blood glucose evolution pattern of António or when he/she receives an alert that António is not going to appointments when he is supposed to. António can also send a photo of suspicious wounds to be checked. If the nurse considers them suspicious, he/she can make a home visit to António.*
18) *In this process, there is also an integration with the community pharmacy. It can anticipate the need for medication, sending refill alerts to the app so that António goes to the pharmacy to get a new stock.*
19) *Every time medication is dispensed, that information is sent to the patient's app. That is actually how it can be perceived by the system that one prescription is over, allowing the alert to be sent to the doctor (as mentioned in point 12).*
20) *When António goes to the pharmacy to get a new stock, his weight and blood pressure are also measured. This data is either introduced manually or automatically sent to the app, where it can be consulted by everyone. If an outlier is detected in blood pressure, for example, the pharmacist can send a note to the doctor in Lusíadas. If the pharmacist understands the patient is not adhering to medication, he can also send a note to the hospital so the healthcare professionals can think of a plan of action.*
21) *At least annually, the nurse makes sure that António fills questionnaires related to his well-being and how he perceives his health is, both mentally and physically, in order to understand whether the treatment his bringing value to his life.*

For a more interactive visualization of the patient journey and the proposal of value, consult Annex 1. This Annex also presents in what points data should collected for the knowledge and non-knowledge based CDSS and what wearables, devices and other existing Glintt solutions each point requires. These latest aspects will be discussed in greater detail throughout the next sections, but for now some features of the use case will be analysed, both in terms of their integration with what was discussed in the state of the art and the feedback they received by the team of healthcare professionals that were consultants in this project.

4.1.1.1 Use Case: The Clinical Management of the Patient

As referred in Chapter 2, one important aspect of the initial research was to understand how diabetes is managed clinically, in terms of which specialists should follow the patient and what exams should be done regularly, and adapt the solution to provide some help in that aspect. As summarized in Table 2.1 from Section 2.3.2, CDS can be applied in diabetes to assure the collection of the necessary documents, exams and values, including: eye check, foot check, kidney values, blood pressure, cholesterol, HbA1c, weight, immunizations and use of medications.

The research that ICHOM took to define global standards for measuring diabetes outcomes, presented in Section 2.7, defined what aspects should be assessed regularly to handle chronic and acute complications and the proposed solution is in accordance with many of those.

Table 4.1 summarizes how different aspects of clinical management are proposed in the use case. There is a definition of which point of the use case they are contemplated in and how it can be performed.

Table 4. 1 Summary of Clinical Management Aspects in the Proposed Use Case

Variable	**How it can be collected**	**Contemplation in the Use Case**
Blood Pressure	Pharmacy visits	Point 20
BMI (weight)	Pharmacy visits	Point 20
Glycated Haemoglobin	Blood Tests (prescribed by the specialist doctor)	Points 2, 14, 15
Foot Check	Podiatrist/Nurse (referenced by the specialist doctor)	Points, 2, 13
Eye Check	Ophthalmologist (referenced by the specialist doctor)	Points 2, 13
Kidney Values	Blood/Urine Tests (prescribed by the specialist doctor)	Points 2, 14,15
Lipid Profile (cholesterol)	Blood Tests (prescribed by the specialist doctor)	Points 2, 14,15

All these aspects aim to assess the patient's overall health status, but also to control the appearance of the associated complications. Blood pressure and lipid profile are important to control macrovascular health; foot check, eye check and kidney values to assess the appearance of neuro, retino and nephropathy, respectively. Finally, HbA1c controls how the disease is progressing and, together with BMI, can have an impact on the overall appearance of complications.

Some of the CDS projects discussed in the state of the art, namely SEBASTIAN and COMPETE II (Section 2.3.2) had a big focus in controlling exactly when a certain appointment or exam had to occur. The approach here was different, mostly because one of the doctors from the Lusíadas team pointed out that each patient is a different case: a patient in an early stage of the disease might only need one eye exam per year, whereas in a more advance stage it would make sense to perform it every six months, for example.

The way to handle that issue in the proposed use case was to ensure compliance to what the "central" specialist doctor that follows the patient determines: in point 2, that doctor defines a plan of action and determines what other specialists the patient should see, with the possibility of

directly scheduling the appointment. That information becomes associated with a calendar that the patient can consult. In point 13, if the patient did not attend that appointment, an alert is sent to some of the intervenients. The same happens to blood/urine tests: the doctor defines when they should be performed (for example, before the next appointment with him/her) and can previously prepare the laboratory request for the necessary values (glycated hemoglobin, cholesterol, renal function values, etc), as contemplated in point 14. Even though this doctor in specific might not be the responsible for other specialist-related exams, like foot and eye checks, by referencing the patient to these other doctors, ophthalmologist and podiatrist, he is making sure the right follow-up is guaranteed. The results of the exams should also become available and he/she can check that they were, indeed, performed.

The results of the laboratory exams can, if there is a possibility of integration from the place where they are performed, be updated into the app. As mentioned in point 15, in the case of value outliers, alerts should be sent to the doctor so that, if he/she considers necessary, can identify a plan of action.

Another interesting consideration is the script followed by nurses in the follow-up appointments. It includes the assessment of certain aspects, including immunizations and smoking status, that were mentioned in the studies from the state of the art and are presented in Table 2.1, but not included in specific in the use case. The goal is that the script is still followed in the appointments with the nurse and it could even be implemented in Clinical Pathways. That idea will be discussed in greater detail further on.

Diabetes Wizard, the other project discussed in Section 2.3.2, had a focus on medication intensification when clinical targets were not being met. The proposed solution should be able to detect patterns of deterioration in the patient over time, for example by increasing number of weekly hyperglycemias, and communicates it with a healthcare professional. The latter can, from there, decide to change or not medication.

4.1.1.2 Use Case: The Daily Follow-up of the Diabetic Patient

Regarding this aspect, the telemedicine projects presented in Chapter 2, METABO and gluCModel, were considered benchmarks for the daily management of diabetic patients, since they already offer an integrated solution aiming to control the different aspects of lifestyle that contribute to a good control of diabetes. Table 4.2 summarizes the aspects contemplated in both and also integrated in this solution, indicating the point of the use case where they are mentioned.

Table 4. 2 Summary of how different lifestyle aspects are managed in METABO, gluCModel and in the proposed solution

	METABO [12][63]	**gluCModel [55]**	**Proposed Use Case**
Nutrition	Recommended daily calories and CHO intake per meal can be consulted.	Patient uploads date, hour, moment of day (breakfast, midmorning lunch,…) and number of CHO.	Point 8
Glycemic Values	SBGM and CGM data (not specified how it can be incorporated/consulted).	The patient has to measure blood glucose and store it indicating date, hour, moment of day, value and optionally a comment. Data could also be imported automatically from some glucometers.	Point 5
Physical Activity	Duration, intensity and frequency collected by a pedometer and a metabolic holter.	Authors designed a scale from 1 to 10 to indicate level of exercise.	Point 9
Medication	Current administrations (ATC code of the drug); strength; time and dosage.	Directed to insulin-dependent patients: they specify the date, hour of injection, moment of the day, insulin type and number of units.	Point 11
Clinical Decision Support System	-Feedback System in 3 layers, explained in section 2.4.2 -Metabolic Model can forecast adverse glycemic events	-It generates recommendations under certain predefined conditions -Like METABO, the metabolic model can make forecasts.	(knowledge-based: Section 4.2.1.6; non-knowledge solution: Section 4.2)

Note that gluCModel and METABO were mostly designed for patients taking insulin rather than anti-hyperglycemic medication. As later explained by the doctors, in early stages of the disease T2DM patients only take medication and they start insulin as it progresses, usually some years later. In early meetings, it was decided that, for now, only patients taking medication, rather than insulin, would be contemplated for the solution. However, it could eventually be adapted to insulin-dependent patients, but an effort would have to be made in that sense, since the system would require additional features, for example the calculation of the insulin dosage according to food intakes.

Another aspect that both METABO and gluCModel focus on is patient's education, which makes sense for a chronic disease, since they need to learn a lot regarding their condition and

how to manage it. The app does not present concrete proposals regarding that aspect, at least in this initial use case. However, education on the disease is contemplated in the initial conversations with the healthcare professionals (points 3 and 6). Additionally, a communication channel is proposed between patient and nurse, through which any doubt should be clarified.

As mentioned in point 7, there is also the possibility of personalizing some values at some point in the future. For example, if the doctor considers that it is normal for a certain patient to have glycemic values a little above the predefined targets, there should exist a possibility to change it. Together with personalizing the need for attendance of appointments and exams, one considers this solution a step towards a more personalized medicine.

4.1.1.3 Use Case: The role of the Pharmacist

Overall, literature reports positive outcomes when the community pharmacist in involved in the care for the chronic patient.

One of the goals of ConnectedHealth was exactly to have a bigger integration of the community pharmacy in the journey of healthcare delivery. This also makes sense to be adopted and incorporated in a Glintt project, since Glintt is the leader in the Portuguese pharmaceutical sector in terms of software.

Integration and collaboration with the community pharmacy is achieved in some points of the use case:

- Point 1: taking advantage of something that already exists in some Portuguese pharmacies for risk assessment of diabetes, there is a referral to the hospital in case of high risk of diabetes.
- Point 18: the patient receives alerts from the pharmacy in his app to refill his medication stocks, which is something that already exists as well.
- Point 19: the information on the disposal of medications is sent from the pharmacy to the central system to be cross-checked with the initial prescription, in order to know when the latter is finished.
- Point 20: Some measures that the patient has to take regularly, such as blood pressure and weight, can also take place there, with certified medical devices, including a sphygmomanometer and a scale weight, respectively.

In terms of communication between hospital and pharmacy, which was one of the goals of ConnectedHealth and mentioned in point 20, it was defined that the easiest solution would be sending notes with written information between the pharmacist and the doctor.

It is however worth mentioning that the initial proposal aimed to give an even bigger role to the pharmacist, including questionnaires with the goal of identifying early symptoms of diabetes-related complications. But this did not pass in the validation from the Lusíadas Group doctors, since they claim the pharmacist may not have enough medical knowledge to perform this and some symptoms could be misunderstood. According to them, pharmacists can help in the

other proposed aspects, but complications should be assessed in the specific exams mentioned previously, with the adequate healthcare professional performing them.

4.1.1.3 Use Case: Validation and Feedback by Healthcare Professionals

The proposed solution was adapted given the feedback of the healthcare professionals. These are some of the aspects worth mentioning, since they brought useful insights:

- Glycemic Values Alert: First of all, there was a suggestion to send an alert every time there was an hyperglycemic event. However, this was immediately seen as non-feasible since there are many uncontrolled patients and it would probably create an alert fatigue that could be counter-productive to the goals one aims to achieve. Besides that, a patient could have a hyperglycemic event solely related to a punctual food excess, which is not significative to his/her health condition. So, they pointed out that the most valuable thing would be a way to control how the patterns of the patient are evolving, for example alerts in weekly blocks if a certain number of outliers is detected or even the possibility to analyse patterns in clear visualizations provided by the solution, in order to evaluate if the disease is deteriorating or not. The doctors also pointed out that the normal values of glycemia for one patient might be different than the normal values for others. So, even though some initial general values for what outliers might exist, the solution can help to understand what the normal values for each patient are and personalize alerts from there.
- Medication: In terms of medication, the initial proposal was also different. There was a suggestion to send an alert every time it was time to take a pill. The doctors also found this overwhelming, since this sort of patients usually take a lot of medication, so there was another risk of fatigue. With a simple check throughout the day and a reminder just in case the patient did not put the check, that burden might be avoided. It is worth mentioning, however, that one of the doctors pointed out that sometimes the patients get confused about which medications to take at what hours, so a simple list providing an easy consultation of those aspects could bring a lot of value to the solution, which is why is was explicitly incorporated in point 11.
- Nutrition and Physical Exercise: Regarding these aspects, the doctors and nurses pointed out the need for individualization and to take every patient's preferences into account, so that benefits can be maximized. For example, there was an initial suggestion to send an alert to the patient if he had not exercised for a day. But the medical team mentioned that there should more flexibility should exist, considering what the patient is willing or not to do. The same applies to nutrition: diabetic patients should follow a strict diet. But if they have too many restrictions they might lose their motivation, which is counter-efficient. So, overall, there was an advice to allow some flexibility in these aspects of the solution and not have alerts that are too strict.
- Prescription of Non-Drug Therapies: During some discussions, the idea of digitally prescribing "attitudes" in complementation with medication also came up. As mentioned

throughout this work, chronic diseases are dealt with the adoption of the adequate lifestyle practices. Nowadays, as one was informed by a nurse, those aspects are discussed during appointments, both with nurses or doctors, and the best practices are advised. At most, some educational papers are provided to the patient. However, during the discussion it came clear that, if medications can be prescribed and consulted digitally, so should the other sort of prescriptions. This solution can be a kickstart for that. Namely, the patient would benefit if he could easily consult in his app: how often and at what times of the day he should measure glycaemia (if not wearing a CGM device); an adequate meal plan and physical activity plans to follow. This is similar to what happens in METABO, as well, and is contemplated in point 10 of the use case.

- Other Considerations: Overall, the initial use case received a positive feedback. The doctors pointed out that such an integrated solution has the potential to bring value to the way chronic diseases, including diabetes, are managed. On the one hand, it allows to monitor the lifestyle aspects that influence the outcomes of the disease and could provide some insights into analysing how the disease is progressing. The alerts regarding the attendance of appointments and exams also allow a closer monitorization that the prescribed care recommended for the patient is being followed, with the possibility of directly contacting him/her if he is not attending them as supposed. The doctors pointed out that, nowadays, after an appointment they have no way to control if the patient is ever going to appear again or not and a system of alerts can bring some help in this aspect.

4.1.2 Functional Requirements

The use case was just the initial way to structure what the proposed solution would do. The next step was to specify what is required to make it happen. It was important to understand, in terms of technology, what solutions already exist inside the company and only need to be adapted, what has to be bought and what solutions have to be developed from scratch. Afterwards, the functional requirements were defined, describing what the different components of the system must do to make the solution described happen.

Having the use case as a base, one comes up with the following technological components required to put the solution into practice:

4.1.2.1 Glucometer

This is described in point 5 of the use case and, obviously, makes sense for a patient with diabetes, a condition in which glycemic values are poorly regulated in the blood, to control this value regularly.

There were two options for measuring the blood glucose values: a glucometer, so the patient performs SBGM, and a CGM device. Neither option was excluded from the beginning. But the ultimate goal was to find one that could be integrated, via an Application Programming Interface (API), with a central system.

Fraunhofer, that has an important role in the search for the adequate wearables and sensors in the ConnctedHealth project, ended up proposing a glucometer: Meditouch 2 Blood Glucose Monitor by Medisana [72], that calculates the blood glucose by obtaining blood in a fingerstick and already has the ability to send the data via Bluetooth to an app, where it is displayed.

In principle, the data obtained can be send to the app (that will be discussed further on) via APIs, but the app should also be as agnostic as possible and have the possibility to insert the values manually, in case there are interoperability errors between the app and the device or in case the patient uses other sort of glucometers.

4.1.2.2 Fitness Bracelet

As described in point 9 of the use case, the patient is advised to embrace a healthier lifestyle that includes the practice of physical exercise. In order to monitor it, he can wear a fitness bracelet that monitors his steps, distances and calories. Many of these bracelets also measure the heart rate and that would also be a plus for the solution, since it can allow the detection of arrythmias.

Again, Fraunhofer chose a solution from the market, Fitbit Inspire 2 [73], with the ultimate goal, of allowing interoperability of the collected data with the central system, in mind. But again, the hypothesis of having manual input of data and making it agnostic of the equipment should be considered.

4.1.2.3 Nutrition Control

Initially, the nutrition module, described in point 8, was not necessarily seen as a technological necessity. But many questions were raised on how to do that control, for several reasons: first, diabetic patients benefit from both CHO and calory counting, since CHO directly influences blood glucose and calories influence the patient's overall health. However, while type 1 patients are used to do this sort of CHO counting in their daily lives every time they have a meal, since it is necessary to know which insulin bolus to take, type 2 patients that do not take insulin are not used to do that. There were some concerns that having them do that would raise the issue of fatigue and overburden and lead them to drop the solution.

There were some discussions of simply having some healthy meals that the patient could consult when he did not know what to eat, but eventually Fraunhofer presented an app developed by them in the past that suits the needs for this solution [74]. This app takes user information such as: anthropometric data (weight, height, age, activity level and estimated nutritional requirements); eating habits; food preferences and restrictions. The idea is to have a professional, such as a nutritionist or even a nurse, making adequate meal plans from there. The meal plans become available for consultation in the app, as well as information regarding them, including ingredients, quantities and their nutritional information.

This would make it possible to have information regarding the CHO and calories that the patient is ingesting available. This would also facilitate his/her life, in the sense that there would be no effort in planning meals: they are already available, with the guarantee that they are healthy and suitable for a diabetic patient. There has to exist, however, some trust that the patient strictly follows the meal plans. Therefore, the professional that plans them should make an effort in respecting the patient's preferences.

This app opens the door to include a nutritionist as another player in the journey of the patient, if considered necessary.

4.1.2.4 The App

In the use case, there is a constant mentioning of the app that the patient uses. In truth, that app can incorporate the CDR initially thought to be developed in the project to store the patient's health data. It would even make sense that the app has different end-users: the patient, the nurse and the doctor in ways that will be described next. The app should include the following features, besides basic personal information on the patient (name, age, residence, etc.):

- **Data from sensors (glucometer and fitness bracelet):**

- Glucometer: the blood glucose data should be collected and displayed in the central system, in order to alert for urgent outliers, such as concerning hypoglycemic values and to analyse patterns of evolution of glycemia to infer how the disease is progressing. That data should also be organized in weekly or monthly charts, to allow easier consultation of the evolution by the patient and the nurse.
- Fitness bracelet: the data related to physical activity (distance, steps, calories) is collected so that conclusions can be drawn regarding whether the patient is exercising properly considering the recommendations he/she was given. The bracelet also collects the heart rate and is able to detect arrhythmias. That data should also be intuitively organized in charts to understand if the patient is being active.

- **Meal Plan (from the nutrition app)**

The simplest solution is to have a link directly connecting the central app to the nutrition app. In this way, the patient can consult the meals planned and the healthcare professionals can consult, if they wish, the CHO and calories the patient has been ingesting, for example.

- **Drug Prescription**

The drugs prescribed by the doctor should be available in the app and, from there, be easily accessed from the pharmacy's informatic system via APIs. The drug prescription should also have the necessary organized information of which medications to take and when, as will be described in the Care Plan.

When the pharmacy dispenses the medication, that information should be sent again, via interfaces, to the app. This is the only way to know when the prescription is finished.

- **Attitudes Prescription/ Care Plan**

Attitudes could, in principle, be prescribed at the same time and in the same way as drugs by the physicians. The nurse should also have the possibility to prescribe some attitudes. These attitudes should include: at what times of day blood glucose should be measured; the physical activity plan and even when to go to the pharmacy or to further appointments and medical exams. All that information should be available in the system, so that the patient can consult them in case of doubt. The use case also contemplates the existence of alerts if certain aspects prescribed are not fulfilled (for example, nonattendance of appointments).

It is still being discussed how the attitude component of prescriptions should be defined. The main proposal was to make it a component of the external prescription of medications (from Globalcare).

- **Laboratory Requests**

Before a laboratory analysis, the doctor can prepare the request with advance and make it available automatically in the system, so that the patient can easily take it.

- **Laboratory Results**

The original idea is to have the system receiving the laboratory results and have alerts if outliers are detected, so that a plan of action can be anticipated by the doctor. However, to do this, the place where the patient takes the analysis has to be able to interoperate with this system. For now, there is a collaboration with Lusíadas patients, but they can perform their exams elsewhere. So, either it had to be guaranteed that patients perform their exams in Lusíadas and the results are sent automatically or, if they do it elsewhere, they could simply attach the PDF file of their exams. However, in this last possibility, it would be more complicated to have alerts associated.

- **Pharmacy Measurements**

The use case proposes that, when collecting their medication, the patients also took some measurements in the pharmacy, including weight (for BMI calculation) and blood pressure. This data should also be stored and it would be interesting to consult how it is evolving over time, in the same way as the data from the sensors.

- **Communication Channel**

It would be interesting to have a communication channel where all members of Viewer, including patients, nurses and doctors, could communicate with each other individually or even

by creating groups. This is also in accordance with what was proposed in the use case, where close communication between nurse and patient, at least, is mentioned in several points (in the case of an adverse glycaemic event is detected; in case a patient skips something, etc). It is also what happens in the case of METABO.

Using Existing Glintt solutions for the App

For the previous technological necessities, glucometer, fitness bracelet and nutrition app, solutions from the outside market were considered. But for the app in specific and all the features it should have, Glintt Solutions can be adapted.

The **Viewer**, briefly explained in Section 2.5, could be used. The Viewer could aggregate the data that was described, since it is highly interoperable and has the ability to communicate with sensors and others softwares.

The Viewer aims to quickly present the user with the relevant information needed. It would require different features according to the end user: for example, a patient is probably more interesting in checking for his/her care plan, the nurse to check the monitorizations and evolutions of some parameters and the doctor to see the laboratory results, but in theory all the information described should be available there. Besides that, each intervenient is thought to receive different alerts according to their role, as will be described in Section 4.1.2.6. Therefore, the Viewer should be adapted for each user in that sense.

Other Glintt solutions are relevant to bring the proposed use case to practice. For example, there should be an integration with **Globalcare** for the following purposes:

- Appointment Requests: In the appointment with the "central" doctor that follows the patient, when he/she makes a referral to other specialist, for example an ophthalmologist, he can immediately request the appointment (point 2).
- Diagnostic Record: Used when the laboratory results confirm the diagnostic of T2DM (point 2).
- Laboratory Order (point 14).
- Prescription: the external prescription of medications already exists in Globalcare and the idea is to extend it to non-medication content (point 3).
- Bedside Nurse can be used in an ambulatory context, to get access to patient's records remotely and do annotations in an efficient way, if the nurse goes to the patient's house to check some wound or similar.

Regarding the pharmacist, this healthcare professional is not a contemplated end-user of Viewer, unlike all the other intervenients (doctor, nurse, patients). Therefore, the only way to bring the use case to life is to have the pharmacist use **Sifarma**, the pharmacy's information system, and the required data is acquired and sent via APIs to Viewer. The notes/messages that have been described for communication between pharmacy and hospital would also have to be sent via these APIs.

Glintt has the goal to apply **Clinical Pathways,** also described in Section 2.5, for the management of chronic diseases. So, additionally to everything that has been proposed already, thinking how Clinical Pathways could be applied in the use case was necessary:

- They can be applied in Sifarma, in point 1: if a certain risk score of Diabetes is detected, it should activate an immediate protocol to advise the pharmacist to send a note for referencing to a specialist in the hospital. Nowadays, Sifarma is already doing this referencing to the public health data of Ministery of Health, but it could be done to hospitals using Globalcare as well.
- In Globalcare, in point 2: if the patient is directed from the situation above, a Clinical Pathway can be used to follow a protocol of what to do. That could include further exams, confirmation or not of diagnosis, and, if the diagnosis is confirmed, another protocol with the steps to follow (schedule appointments with other specialists, other exams, prescription of medications and attitudes, etc)
- Everytime there is a follow-up appointment, the Clinical Pathway can also be helpful to cover everything that is necessary, both for the doctor and nurse appointments. In the case of the nurse appointments, it could follow the script these healthcare professionals usually follow and it could include additional features: for example if the nurse registers that the patient did not have his last flu shot, there could be an immediate link to schedule it.
- Again in Sifarma, in point 20, another clinical pathway could be used regarding the measurements taken in the pharmacy. If there are outliers, it could again activate some referencing to the hospital, similar to the first case.

Note that these proposals are very preliminary and further discussions will have to take place to understand if this sort of application of Clinical Pathways is feasible.

4.1.2.5 Clinical Data Repositorium

In the previous section, the macro functional requirements of the app were described. Viewer is supposed to serve as a "central hub" of health, where clinical data is stored and owned by the patient and, under his authorization, other healthcare professionals can consult it.

To make that happen, the data needs to be stored in a structured way: all the data collected in the app should be stored in the CDR, based on the FHIR specification, since it is the standard Viewer uses and the standard the healthcare to is transitioning to, in general.

Table 4.3 specifies the FHIR resources and elements of the resources of the data necessary for each of the features predicted:

Table 4. 3 FHIR Resources and Elements of the data stored in the CDR

Macrofunctionality	FHIR Resource	Elements of the Resource
Measurements of the Sensors/ Pharmacy Measurements	Observations [75]	Code; Value

Meal Plan	Nutrition Order [76] (if a healthcare professional makes a recommendation of the meal plan)	foodPreferenceModifier ; excludeFoodModifier; NutritionOrder.oralDiet.type; oralDiet.nutrient.amount
	Nutrition Intake [77] (gathering the information of what the patient has eaten	code; consumedItem.nutritionProduct; consumedItem.amount; ingredientLabel.nutrient; ingredientLabel.amount
Drug Prescription	Medication Request [78]	medication[x]; dosageInstruction
Prescription of Attitudes (Home Glucose Measurements/ Physical Activity); Laboratory Request	Service Request [79]	Intent; category; orderDetail; quantity[x]; occurrence[x];
Prescription of Attitudes (Next Appointments)	Appointment [80]	serviceCategory; appointmentType; start; participant.status
Laboratory Results	Diagnostic Report [81]	Category; code; result

In terms of how the implementation of the CDR will be, it was still not defined. Smile CDR, discussed in Section 2.2, seems like a promising option, but it is commercial and Glintt usually prefers open-source solutions, where people from all around the world can contribute to the development of the software. Therefore, internal decisions need to occur on whether to buy Smile CDR or to develop a similar open-source FHIR-based database from scratch.

4.1.2.6 Clinical Decision Support System

The knowledge-based CDSS shall receive information from the CDR via APIs and is responsible for all the rules and alerts described in Table 4.4, that are in accordance with what was proposed in the use case:

Table 4. 4 Triggers and Suggestions of Alerts Proposed for the Knowledge-Based CDSS

Trigger	**Suggestion of Alerts to the Patient**	**Suggestion of Alerts to the Nurse**	**Suggestion of Alerts to the Doctor**
1. Glycaemia < x	An hypoglycaemia was detected. Take your SOS plan and measure it again in 15 minutes. If values do not increase, go to the hospital	Patient *(*insert name of the patient with link for his/her the profile*)* is experiencing an hypoglycaemia	
2. Glycaemia > x	A critical hyperglycemic value was detected. Take your SOS plan	Patient *(*insert name of the patient with link*	

		for his/her the profile)* is experiencing a hyperglycaemia above value x	
3. Number of hyperglycemias in the past 7 days > x ***Additional Suggestion*:** Number of hyperglycemias in the past 7 days > previous week (to detect deteriorations easily)	You had x hyperglycemias registered last week. It is important to follow your plan of care: take your meals properly, do the prescribed physical exercise and take your medication properly to keep your health under control! (Note: a comparison with the previous weeks could be presented)	Patient *(*insert name of the patient with link for his/her the profile*)* had x hyperglycemias registered last week (Note: a comparison with the previous weeks could be presented)	
4. Number of calories or number of steps in the last 3 days < what was defined by the healthcare professional	In the last 3 days, the sensors detected a physical activity *(*insert value**)% below what was expected. Keep in mind that physical activity improves blood glucose control and diminishes insulin resistance		
5. Medication check was not inserted	Did you take your medication properly today? It is important to keep your glycemic levels controlled		
6. Arrythmia detected through the fitness bracelet	Arrythmias were detected. Go to the hospital.	Arrythmias were detected in patient *(*insert name of the patient with link for his/her the profile*)*	
7.1. 5 days prior to a scheduled event in the calendar (appointment, exam) **7.2.** If the date passed and the person did not attend it	**6.1** Do not forget about your appointment on day *(*insert date and hour of the appointment with link for further information*)* **6.2.** You did not attend your appointment! Please reschedule as soon as	**6.2** Patient (**insert name of the patient with link for his/her profile*)* did not attend appointment *(*insert date and*	

	possible. It's important to follow your care plan	*hour of the appointment with link for further information*)*	
8. Medication stock will be over in 5 days (Note: Sifarma can already do this tracking)	In 5 days, go to the pharmacy to refill your stock of medication *(*insert name of the medication*)*		
9. Pharmacy dispensed the amount of medication that was in the prescription	Your prescription is over. If your doctor does not issue a new one, contact him.		Prescription of patient *(*insert name of the patient with link for his/her the profile*)* came to an end. Call him for an appointment or prescribe a new one
10. Outliers of the laboratory results (ex: HbA1c>8%)		Patient *(*insert name of patient with link for his/her profile*)* had a HbA1c value above 8%	Patient *(*insert name of the patient with link for his/her the profile*)* had a HbA1c value above 8%. You want to call him for an appointment or prescribe SOS medication?

The x values should be case specific, as different patients will have different thresholds. However, for the initial programming of the rules, the values could be adapted from clinical guidelines. In general, these alerts were approved by the healthcare professionals, who will be responsible for providing those initial values.

Regarding the first 3 alerts, the considerations of the healthcare professionals after presenting the use case were respected. Hypoglycemias should be immediately addressed because they can be potentially lethal. In the case of single hyperglycemias, they are only concerning above certain values. Otherwise, a single hyperglycemia might not be very meaningful. Thus, the control should exist in weekly blocks, since the goal is to have a general perspective on the overall health status of the patient and how it is evolving.

Considerations on the CDSS interoperability

By the time this work was finished, the architecture for the CDSS and its integration with the other components was still under development, like the case of CDR. However, it was left clear that it will be a service model (phase 4 as specified in Section 2.1), using the FHIR CDS

Hooks specification, since the goal is to achieve as much interoperability as possible across different systems.

Another point worth explaining is the fact that two distinct CDSS will have to be developed: a rules engine, i.e., a knowledge-based system for the implementation of the rules, which was detailed above and a different non-knowledge based system with ML algorithms for the data mining that will be proposed in Subchapter 4.2. This seems to have been achieved in the study described in Section 2.1 [33], where a Rules Engine and a Bayes Engine were developed, both invoked from the hospital information system by using the CDS Hooks specification. Therefore, that platform can be used as a benchmark for the development of a CDS system in the context of this work.

4.1.3 Dataflow

The next step was to build a dataflow scheme, to specify the flow of data through the different components of the proposed solution, which is shown in Figure 4.1.

All the fulfilled arrows represent APIs. In the center, there is a specification of the data that is predicted to exist in the CDR and the associated macro functionality of Viewer (the app) on the left. The provenience of that data is also detailed:

Globalcare:

- Globalcare has the external prescription of medications feature, where doctors can do the drug prescription for the patient. If, in the future, it will be developed as it is initially being planned, it will also incorporate the prescription of non-drug therapies through this feature, namely indications for glucose measurements and for physical exercise.
- Doctors can prepare the laboratory requests in Globalcare.
- Globalcare is also the software to request for the next appointments and exams.

Sifarma:

Via APIs, Sifarma can access the drug prescription in Viewer. It also sends information to Viewer everytime medication is dispensed. Additionally, Sifarma sends the measurements of weight and blood pressure performed in the pharmacy.

The CDR also receives information from the **sensors** worn by the patient, in this case the glucometer and the fitness bracelet. Regarding the nutrition app and the meal plan, it was still not defined if Viewer will simply have a link directly to the nutrition app or if the information regarding the meals will be directly available in Viewer, even though in the dataflow scheme it is predicted that the CDR gathers all the information regarding nutrition. It is still not defined which professional will prepare the nutrition orders and which professional will elaborate the meal plans. It would make sense to be either a nurse or a nutritionist.

Finally, it is time to discuss the **CDSS** and its alerts. The pink arrows indicate the data stored in the CDR that need to be delivered to the CDSS to generate all the alerts specified in Table 4.4. The CDSS receives that information and compares it to its configurated rules. If, according to the rule, an alert is to be generated, it communicates with an external Alert Notification System. The Alert Notification System is then responsible to send the alerts to the respective intervenient, the yellow arrows, who receives it in Viewer. Those alerts are also the ones specified in Table 4.4.

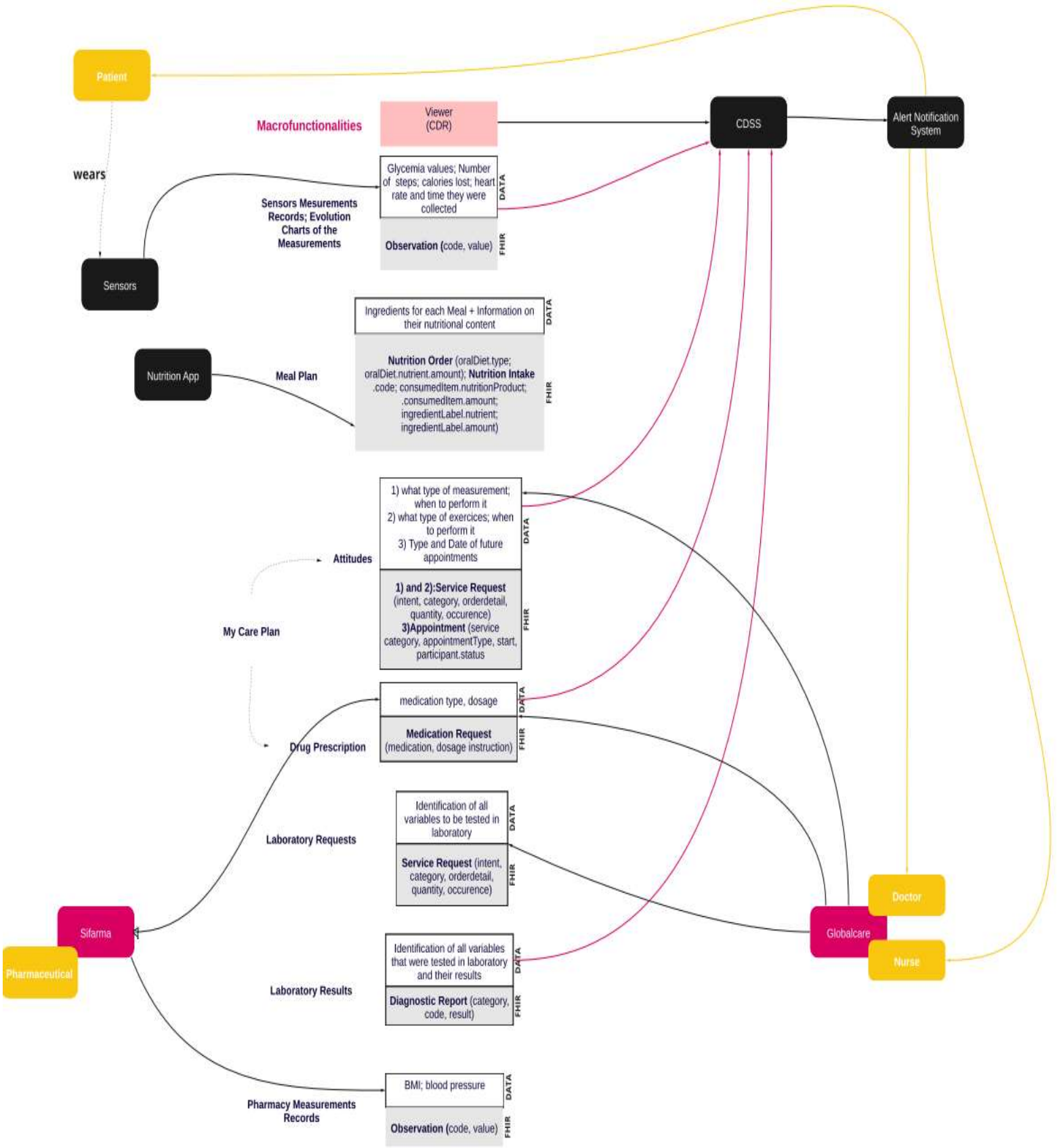


Figure 4. 1 Dataflow Diagram of the Proposed Solution

4.1.4 A preliminary look at the App

After the specification of the functional requirements, it was possible to design what the Viewer can look like on the patient's side, as an app installed on a smartphone. The following mock-ups were designed. They are in Portuguese, but each one will have an explanation.

Figure 4.2 shows the "main screen". On the upper left corner there is information about the patient and on the upper right corner there are three symbols: one for messages, one for alerts and notifications (the bell) and the other one will be explained later. In this screen, the patient can choose to look either at "Highlights" or "My Care Plan", with the Highlights being presented here. In the highlights one has:

- The last glycemic measurement, the hour of the day and how much it varied compared to the last measurement (in the image: a))
- A chart of the evolution of glycemic data (in the image: b)). It is not specified if it represents the last measurements or averages from previous days or weeks. The goal is to have several intuitive ways to analyze patterns related to these values. Note that one of the values is red, to allow quick identification of outliers that have been registered (hypo or hyperglycemias).
- A quick view of the next appointment and exam, with date and hour (in the image: c)

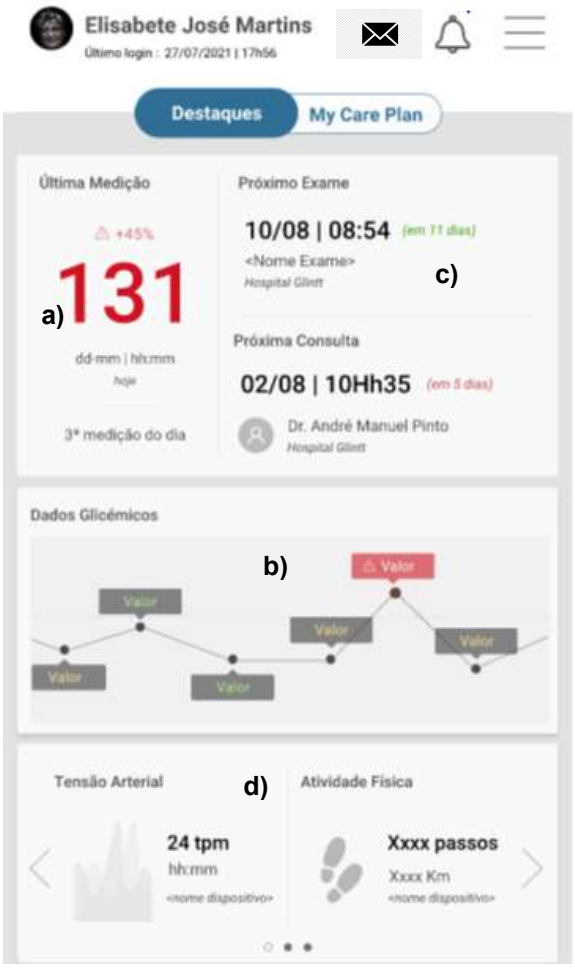


Figure 4. 2 Mockup of the App- 1

- On the bottom, it is also possible to consult the data obtained in the pharmacy and the data from the fitness bracelet (in the image: d)). Ideally, one will click there and have access to visualization charts like the ones described for glycemic values

In Figure 4.3, the patient has chosen to see “My Care Plan”. In this, he/she can consult all the “activities” planned for the day, including measurements and medications and the hours they are supposed to be performed/taken. The patient also has an exam that day and eventually the results can be consulted by clicking there. “My Care Plan” is basically a compilation of both drug and attitude prescriptions but organized in a schedule over the day. Regarding the measurements, it ideally will be automatically obtained from the sensors, but manual input has to be a possibility in case something goes wrong. In Figure 4.4, that data can be manually input. An observation or comment can be added, as suggested by a nurse and as happens in glUCModel, so that the patient has the possibility to justify certain outliers if he wishes to do so. Regarding the medication, the patient has to put checks that he has been taking them properly, otherwise there is an alert at the end of the day. Another interesting aspect to notice is that, in this case, the patient has two new notifications/alerts. There is a brief warning at the bottom and a permanent (until the patient sees the alerts) red ball on the bell symbol at the top.

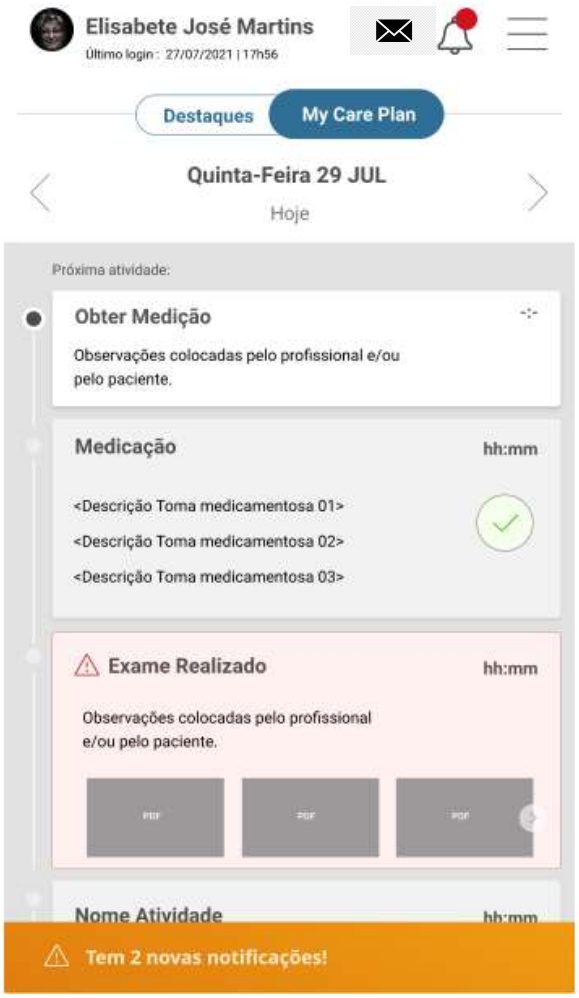


Figure 4. 3 Mockup of the App- 2

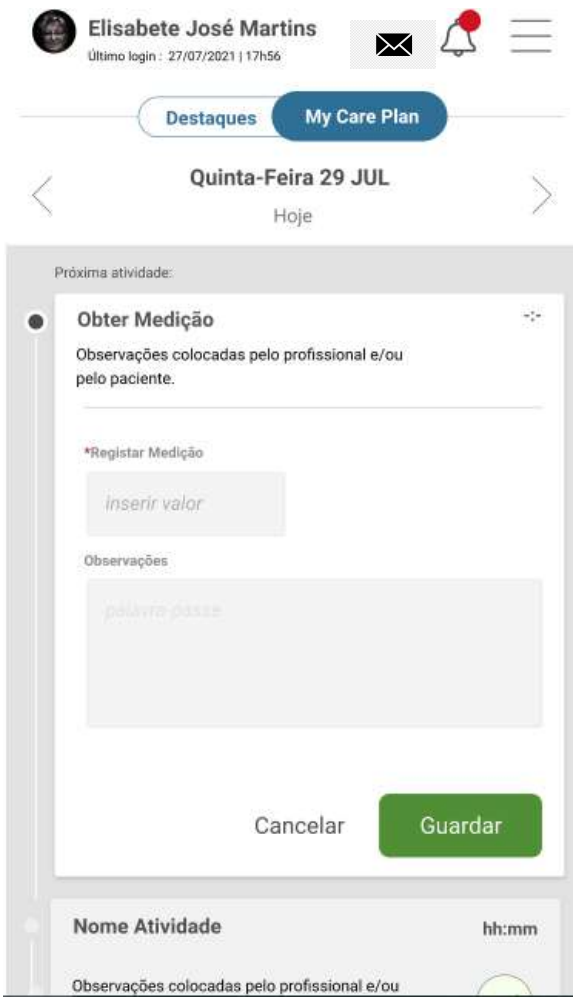


Figure 4. 4 Mockup of the App- 3

Figure 4.5 is what is shown when the patient clicks in the bell figure. The latest notifications/alarms are presented. Figure 4.6, on the other hand, is what the app presents when one clicks on the 3-bar figure on the right upper corner of the other figures. It offers the possibility of adding measurements quicker (easier for manual inputs), to consult the results of the latest exams and to consult the wearables and devices connected to the system (in principle, glucometer and fitness bracelet). The patient also has the possibility to log out of the application here.

Figure 4. 5 Mockup of the App- 4

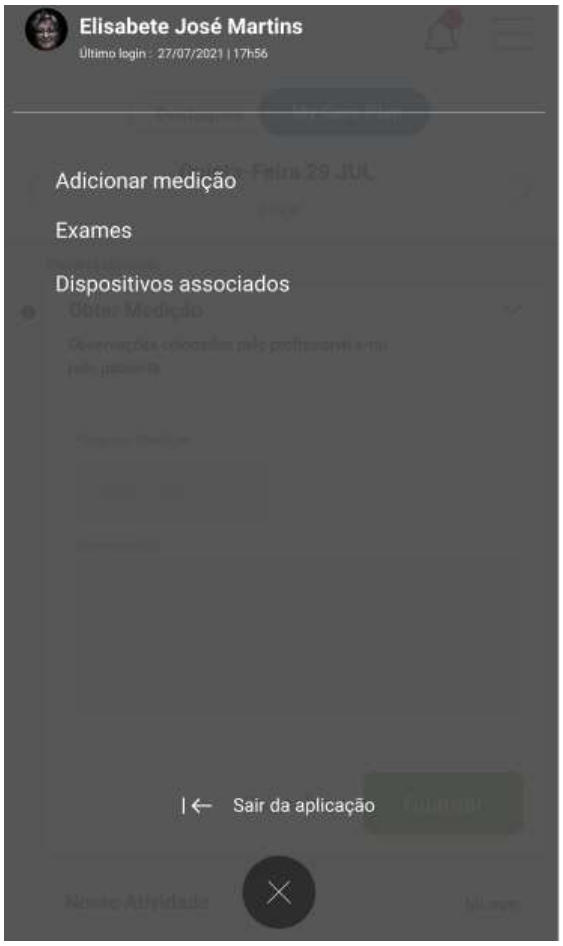


Figure 4. 6 Mockup of the App- 5

4.2 Non-Knowledge Based CDS: Data Mining as a Complement of the Solution

One of the main goals of the project is to turn the data collected into valuable information. This objective can be achieved by a second CDSS that receives data from the CDR, but this time, instead of predefined rules, performs Data Analytics. Some creativity was allowed in the context of this work, to explore and propose ways to do this. In Chapter 2, there was some exploration of how data mining is being applied in the context of diabetes mellitus. It is mostly being used for prediction of the disease, which is not the objective of this work as one is already dealing with patients after their diagnosis, at least for now.

There were also some projects with the aim of predicting the appearance of diabetes related complications, both micro and macrovascular, since, as explained, those are the most harmful consequences of diabetes. Cardiovascular disease, the biggest consequence of macrovascular disease, is, in itself, another consequence of an ageing population and its risk can mostly be controlled by checking blood pressure and cholesterol levels regularly. The underlying mechanisms of microvascular consequences, on the other hand, are more difficult to understand.

As pointed out by one of the doctors in the first validation meeting, controlled patients in terms of glycemia can develop it and patients with many hyperglycaemias may ultimately never develop anything and there does not seem to exist a general rule to predict it. If some insights were provided on its mechanisms, for example if certain variables were detected for the risk of retinopathy, there could be a tighter ophthalmologic control to prevent the worst outcomes. Therefore, predicting the appearance of microvascular complications became one of the goals of this work.

In Chapter 2 during the active research part, several algorithms for the prediction of blood glucose were also described, namely some in the projects that were considered “benchmarks” for lifestyle management, METABO and gluCModel. This makes sense, since being able to predict adverse hypo or hyperglycaemic events in advance allows the patients to take action before worse health outcomes occur. For example, to eat fast acting carbohydrates if there is a prediction of a hypoglycemia. These sorts of predictions may also allow the understanding of the underlying causes of such events and to adapt some lifestyle factors in order to prevent it. However, both METABO and gluCModel deal with insulin-dependent patients and, in the case of METABO, patients with CGM, which collects much more glycemic-related data than in the solution proposed, that uses SBGM. Therefore, it makes sense to go with a simpler model for this work in specific, where only SBGM data is collected, and aim only to predict hypoglycemic events at least in an initial phase, since they are the ones that can be life-threatening.

In summary, based on the state of the art, input variables and algorithms will be proposed to handle two aspects of the disease in the context of this work: prediction of microvascular complications and prediction of the future occurrence of hypoglycemias. Note that both these ML tasks consist in classification and supervised learning tasks. The former refers to the fact that the algorithms will put the data into labels: in the case of complications, the prediction of zero, one or more complications. In the latter case, it refers to the prediction of the future existence of a hypoglycemic event or not. Supervised learning, on the other hand, means that the algorithms will use labelled data from the patients, collected by the system, to learn (each patient delivered to the algorithm for the learning phase has the correspondent complication associated). That data would initially be used to train the algorithm and, eventually, the algorithm will be able to classify new data, based only on the input variables given.

It is also worth mentioning the importance of selecting the right set of input variables for the algorithms to make an accurate prediction. The variables should be independent enough to not cause redundancy of the algorithm, but, on the other hand, they should represent the real complexity of the problem: one does not want to oversimplify it.

4.2.1 Data Mining for the Prediction of Microvascular Complications

4.2.1.1 Input Variables

All the following proposals are based on the state of the art, described in Section 2.3.3. Taking into account the conclusions presented in Table 2.3, the following variables seem to be indispensable to be collected as input variables for the algorithms:

- **HbA1c levels**: its importance for the prediction of microvascular complications seems evident in studies 2 and 3 from the Table, which also makes sense, since the damage of blood vessels is caused by prolonged states of hyperglycaemia and that is reflected in high HbA1c levels. Study 1 captured glycemia by using as predictors daily blood glucose levels. However, HbA1c seems more accurate for this, since it captures the glucose in the blood on the average of the previous 3 months, whereas in a day there can be punctual fluctuations that do not have much meaning (for example, an occasional nutritional excess).
- **Gender**: Capturing the gender seems to be important as well. Being a female was associated with a higher risk for retinopathy and neuropathy in the first study and just neuropathy in the third study.
- **Age at Onset**: The third study pointed out that the age in which there is a diabetes onset can influence the severity of complications. An earlier onset could be associated with more genetic predisposition and a more aggressive form of disease.
- **Duration of Diagnosis**: A longer time since diagnosis is related to a higher exposition to the risk factor (hyperglycemia) and may therefore increase the appearance of microvascular complications. It seems to be associated with nephropathy in study 1 and retino and neuropathy in studies 2 and 3. In study 4, it is also associated with the appearance of complications in general. Note that duration of diagnosis can be directly calculated if one has information on the current age of the patient and the age at onset.
- **Smoking Status**: In study 3, the authors did not seem to find a direct association between smoking and these complications but pointed out that it needed further studies. However, study 4 determined that patients that ever smoked were one and a half more likely to develop any sort of complication.
- **BMI**: BMI was overall considered a risk factor. It was associated with neuropathy in study 1; with retino and neuropathy in study 2 and again with neuropathy in study 3.
- **Hypertension**: Like BMI, hypertension was associated with complications across the different studies. It was associated with retinopathy in study 1 and with retinopathy and nephropathy in studies 2 and 3. Study 4 determined that hypertensive patients were overall one and a half more likely to develop any sort of complication. Note that only study 1 defined blood pressure values in an ordinal way. The others were categorical and just defined the presence or not of hypertension in the patient.

Table 4.5 summarizes the proposed features and suggests how they can be delivered to the algorithm:

Table 4. 5 Input Variables and Suggestion of their Categorization

Input Variable	Variable delivered to the algorithms
Age at onset (save the variable in Viewer)	<25; 25-35; 36-45; 46-55; 56-65; >65
Gender	F/M
Duration of Diagnosis (Current age-age at onset)	<5 years; 5-10 years; >10-15 years; >15 years
Smoking Status	Yes (if patient smokes or has a smoking history)/no (never smoked)
BMI (most recent measure)	<25; 25-30 – overweight; >30 - obese
Hypertension	Yes/No (if the patient has or not a history of hypertension, according to the guidelines values)
HbA1c (most recent measurement)	<6.5; 6.5-7.5; 7.5-8.5; 8.5-9.5; >9.5
Classification	Retinopathy; Neuropathy; Nephropathy; none (or combinations, if it is the case)

Other aspects that are worth considering to be included as input variables are:

- **Rural/Urban Environment:** Patients with higher level of education seemed to develop less complications in the conclusions from Table 2.3, probably because they have more health literacy and look for medical care quicker, understanding better the possible outcomes of the disease and making a bigger effort to change their lifestyle. This could be an interest reality to capture, not in terms of education, but in terms of the environment of the patient: rural or urban. In Portugal, there is a big disparity in the provision of healthcare in the countryside compared to the big cities: in the countryside, people have to make larger dislocations to have access to healthcare. It would be interesting to study if that is affecting the outcomes of diabetes. Since one is only working with Lusíadas right now, present in the big cities, patients are more likely to live in an urban context. But eventually, this app could be interesting to be applied in a rural environment, bringing patients closer to their healthcare professionals, mitigating the effect of this distance. The effect of the environment in the appearance of complications could be studied then.
- **Family History of Diabetes:** this input variable was considered in studies 1 and 3 and associated with neuropathy in the latter. It is not a priority to be included in the present work, since the genetic variables associated with the disease are possibly already represented in the age of onset, as explained. But it could be studied if it has higher accuracy in the prediction or not.
- **Cholesterol:** High Cholesterol, or in some cases LDL levels, was an aspect analysed in most of the studies, but it was never considered to be a risk factor. Therefore, it does not seem to bring value for its inclusion. But, since it is a variable regularly collected, it could be interesting to test if it really does not have an influence in this case.

The suggestion is to first deliver the algorithms the first set of variables mentioned (HbA1c, gender, age at onset, duration of diagnosis, smoking status, BMI and hypertension), train them from there and evaluate their performance. Afterwards, these last variables (rural/urban environment, family history of diabetes and cholesterol) could be tested, individually or in different combinations, to analyse if the performance of the algorithms increases.

Other variables were presented in Table 2.3 but will be discarded, including:

- **Creatinine**: it was included in one study, associated with nephropathy and retinopathy. However, it seems to be a direct consequence of renal failure and not a predictor.
- **Insulin Treatment (yes/no):** As mentioned, initially the plan is to work only with patients treated with other anti-hyperglycemic drugs and not insulin.
- **Physical activity**: not considered an independent variable in the case of this work. If the patient is physically active, that will be reflected on the other factors such as HbA1c, hypertension and BMI and the focus will be on those.
- **Medical cover:** Since the initial project is directed at patients from Lusíadas Hospital, a private hospital, patients are expected to have an insurance that is not a barrier for their healthcare delivery.

4.2.1.2 Algorithms

In the third study of Table 2.3, Support Vector Machines [82][83] and Random Forests [84][85] had the highest accuracy in predicting microvascular complications, thus their proposal for the context of this work:

Support Vector Machines [82][83]

SVM is one of the most used ML algorithms for classification nowadays and it can also be used for regression. The algorithm works in the following way: all the example cases are put in a N-dimensional space (N being the number of features). Then, a hyperplane is drawn such that it maximizes the distance between data points of the different classes, called margin. This process is illustrated in Figure 4.7, where the optimal hyperplane is found in the second image. After the learning phase, every time a new case is presented, it is classified according to where it belongs in the N-dimensional space (i.e., according to which side of the hyperplane it belongs). If the number of input features is 2, then the hyperplane is just a line, but if the number increases to 3, then the hyperplane becomes a two-dimensional plane and so on. SVM usually use kernel functions that transform the original feature space into another, easier to work with. These kernel functions can be linear or non-linear, such as polynomial, radial basis functions, sigmoid, etc.

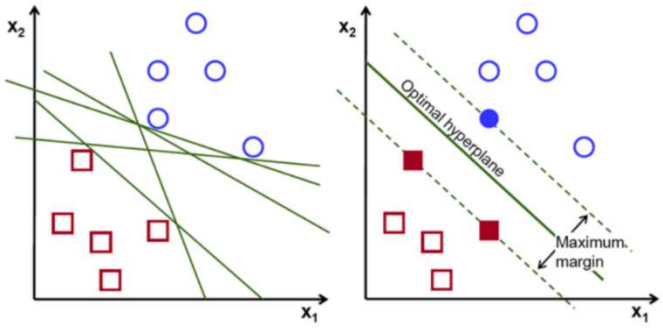


Figure 4. 7 Example of how SVM work in practice [83]

Random Forests [84][85]

Random Forests work through ensemble learning. They combine several Decision Trees, that make a classification individually. In the end, the class with more votes is the final prediction of the model. A decision tree works in a pretty intuitive way: given the dataset, it selects the features that "subdivide" it in the best way possible until it can isolate classes. Figure 4.8 shows the difference between a single decision tree and a Random Forest.

The decision trees for the Random Forests are created from the dataset through techniques called bagging and feature randomness. In bagging, random examples of the dataset, in a random number, are selected and from them a Decision Tree is built. Several Decision Trees are built this way and every time the algorithm has to make a classification, each single Decision Tree will classify individually and in the end the classification with more votes "wins". In feature randomness, when each tree is built, it can only pick from a subset of the existing features, in order to create more uncorrelated trees. This is unlike the process of building a normal decision tree, where all features are included.

By increasing the randomness and reducing the correlation between the trees when building them, it reduces the overfitting of datasets and increases precision (if a class receives many votes over uncorrelated trees, it seems logic that it is more likely to be the right class).

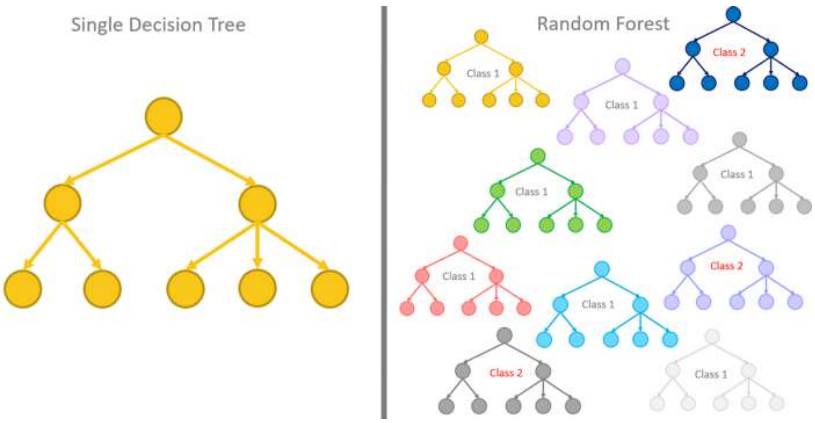


Figure 4. 8 Single Decision Tree vs Random Forest [85]

4.2.1.3 Proof of Concept

The initial goal was to leave this part of the work conceptual. However, it was decided that it could be interesting to do a proof of concept, mostly to illustrate how the data that is planned to be obtained in the future when the solution is implemented can be handled.

A synthetic dataset was created, with an extensive research in literature and taking into account some of the conclusions from Chapter 2, in order to assure the data was realistic. The process of how the data was created is explained in detail in Annex 2.

The reasons for creating the dataset were: no real data from the system would be available in time, since it will take a while to be implemented; no public dataset was found containing the features of interest. The synthetic dataset does not contain all the variables proposed, in order to simplify its creation and to guarantee that the data presented makes sense. Only the variables that were considered of extreme importance are present. Since this section is just a complement of the solution and not the main goal, doing such a simple proof of concept seemed reasonable: The main goal is to show what can be done, to eventually apply it when real data starts to be collected from the system. As explained in Annex 2, 999 patients were "created", with the variables: duration of diagnosis, HbA1c levels, BMI levels, History of Hypertension and Gender. Each of them was also attributed a classification regarding microvascular complications: neuropathy, retinopathy, both neuropathy and retinopathy or none. Nephropathy was not included for this analysis in the final version.

The next step was to test the two proposed algorithms, using a Python library, and show what sort of data analyses could be performed. Scikit-learn is a Python library for Machine Learning and provides algorithms for SVM and RF that can be used in a simple way with just a few lines of code. To run the algorithms, the data was split in 70/30%: 666 patients were randomly selected to train the algorithms and the other 333 were left out to be used in the testing phase, in order to evaluate the performance of the algorithm.

Considerations on Metrics

To evaluate the performance, some metrics can be used: accuracy, precision, recall and F1-score [86]. These will be explained using a confusion matrix, that summarizes the prediction results on a classification problem, more specifically the number of correct and incorrect predictions. Supposing one is handling a classification task with two classes, positive and negative, Table 4.6 shows a confusion matrix for it:

Table 4. 6 General Confusion Matrix for a Positive/Negative Classification Task

	Predicted Positive	**Predicted Negative**
Actual Positive	Nº of True Positives (TP)	Nº of False Negatives (FN)
Actual Negative	Nº of False Positives (FP)	Nº of True Negatives (TN)

The accuracy is simply the ratio between the total number of correct predictions and the total number of predictions:

$$\textbf{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (4.1)$$

The goal of precision is to evaluate what proportion of positive predictions was actually correct:

$$\textbf{Precision} = \frac{TP}{TP+FP} \quad (4.2)$$

The recall, on the other hand, aims to identify what proportion of actual positives was identified correctly:

$$\textbf{Recall} = \frac{TP}{TP+FN} \quad (4.3)$$

The F1-score's goal is to find a balance between precision and recall. One can just aim for a good overall F1 score, instead of focusing on the two metrics individually:

$$\textbf{F1} = 2x \frac{Precision\ x\ Recall}{Precision+Recall} \quad (4.4)$$

Hyperparameter Tuning

Another important consideration is the fact that both algorithms have parameters that should be optimized: finding the parameters that achieve the best algorithm performance is a process known as hyperparameter tuning in ML. To find the best parameters, a function called GridSeachCV [87], that performs an exhaustive search over a set of parameters and is available in scikit-learn library, was used. One needs to explicitly provide the optimizer with the parameters it must search. This is actually one of its limitations. The parameters are specified as a parameter grid, which is a list per hyperparameter, and it delivers the combination that results in the biggest accuracy of prediction.

In the case of SVMs, the hyperparameters to optimize are the following [88]:

- **Gamma:** It defines how far the influence of a single training example reaches. Low values represent "far" and higher values "close".
- **C:** The C parameter represents a tradeoff between the accuracy of the classification of training examples versus the maximization of the decision's function margin (i.e., the margin of the hyperplane), basically adding a penalty for each misclassified data point in the training set. It might be counterintuitive, but low rate of misclassification in training data does not mean the same for testing data, therefore a tradeoff should exist for the margin of the decision function. Low C values represent a bigger margin and high C values represent a lower margin.
- **Kernel:** The goal here is to find which kernel function achieves highest performance.

In the application example shown below, the initial parameter grid provided for the SVM is presented in Table 4.7:

Table 4. 7 Initial Parameter Grid provided to the GridSearchCV optimizer in the SVM application case

Parameter	Grid
C	[0.1, 1, 10, 100, 1000]

Gamma	[1, 0.1, 0.01, 0.001, 0.0001]
Kernel	['rbf', 'poly', 'sigmoid','linear'].

The function delivered the best combination and, from there, one was able to approximate the values of C and gamma and use GridSearchCV again to get closer to the best parameters.

In the case of RF, the parameters are [89]:

- **Maximum Depth:** Maximum depth of the tree
- **Maximum Features:** Maximum features at each split.

The parameter grid provided for the RF aimed to search the maximum depth from 0 to 15 and the maximum number of features from 0 to 5 (5 is the total number of input variables). Again, GridSearchCV role was to search which combination of both provided the highest accuracy.

Considerations on Overfitting

Overfitting occurs in ML when a statistical model fits exactly the training data but has low performance when tested against new data, not seen during the training stage [90]. In other words, the predictor is not able to generalize. One always wants to avoid overfitting and several techniques can be used to verify if it is happening and/or to avoid it. In the context of this work, the following additional tasks were performed:

- A **new dataset** of 450 patients was created, following the same rules, as described in Annex 2, and the previous model was used to predict this new data. In principle, if the accuracy is not considerably lower, one can guarantee overfitting is not an issue. However, this is a rather simplistic approach as the new data share the same statistical properties as the data used during the training stage.
- **Addition of noise:** the original dataset was slightly adapted to incorporate 5% of noise and then 10% of noise. To do so, in the first scenario the same input variables were kept but 50 patients were given a wrong classification to their complication. In the second scenario, that purposeful misclassification was given to 100 patients.

SVM: Results

- **Hyperparamater Tuning**

The hyperparameter optimization delivered the results presented in Table 4.8.

Table 4. 8 Hyperparameter Optimization Results- SVM

Hyperparameter	Best Value
C	0.03
Gamma	0.15
Kernel	polynomial

As explained previously, GridSearchCV does not give an automatic delivery of the best parameters and one has to do the iteration “manually”. Figure 4.9 shows differences in accuracies

for the surrounding C and gamma that were delivered has the best, with a polynomial kernel already chosen:

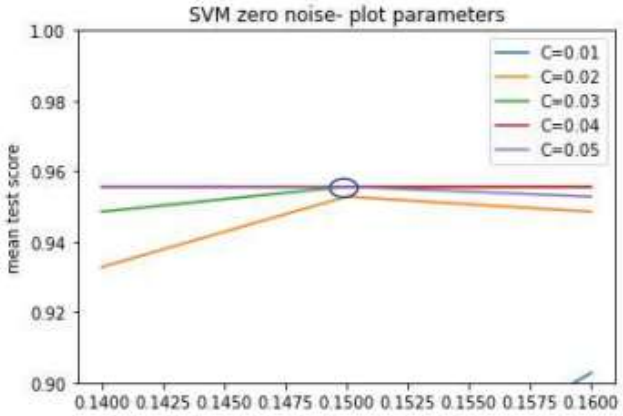


Figure 4. 9 Hyperparameter Tuning for SVM (without noise)

The function delivers the first maximum score it finds, but as observed, C=0.04 and C=0.05 have similar scores. Another aspect important to mention, is that the parameters delivered as the best ones might change every time the algorithm is run, since that means the creation of a new hyperplane and some differences might exist. However, no significant differences were observed in the metrics that will be displayed next, therefore it seems safe to keep as best parameters the ones initially indicated.

- **Metrics**

Having chosen the best hyperparameters configuration for the SVM, the results when making predictions for the 30% of the data used as test set, after training with the remaining 70%, are presented in Table 4.9.

Table 4. 9 Performance of the SVM in the test set (without noise)

Class	Precision	Recall	F1-score
No complication	1.00	1.00	1.00
Neuropathy	0.98	1.00	0.99
Retinopathy	0.99	0.81	0.89
Neuro + retinopathy	0.79	0.97	0.87
Macro average	**0.94**	**0.95**	**0.94**
Weighted average	**0.95**	**0.94**	**0.94**
Accuracy		**0.94**	

Note that the precision, recall and F1-score are obtained individually for each class and then combined in a macro average form, which is the arithmetic average, and a weighted average, that takes into account the representation of each class, giving less impact to the classes less represented. The weighted average will be used in the context of this work. Both the accuracy and the weighted average for the F1-score were of 0.94, which is very positive, meaning that the algorithm was predicting well against the data provided.

For better data visualization, the plot from Figure 4.10 was obtained:

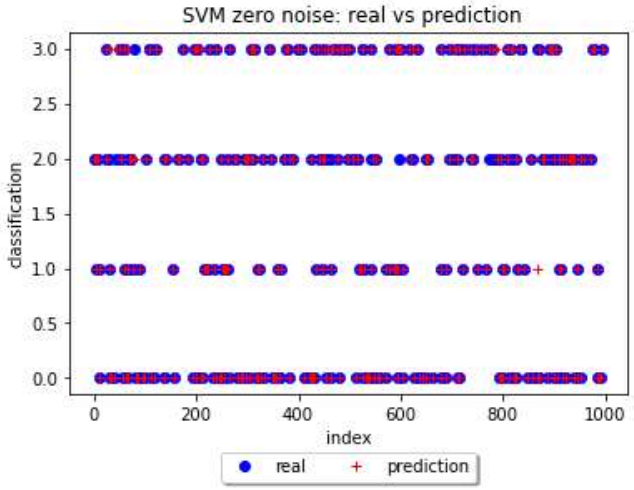


Figure 4. 10 SVM: Overlap of Real and Predicted Labels (without noise)

As mentioned, 30% of the original data, randomly selected, was used to test the algorithm, after it trained with the other 70%. In Figure 4.10, the blue circles represent the real labels and the red crosses represent the prediction of the algorithm. Note that, on the y scale, 0= no complication; 1=neuropathy; 2=retinopathy and 3=neuro+retinopathy. As one can observe, the predicted labels mostly overlap, which is in accordance with an accuracy of 0.94.

The data obtained can also be included in a confusion matrix, Figure 4.11, and the accuracy for each class can be summarized in a bar plot, Figure 4.12:

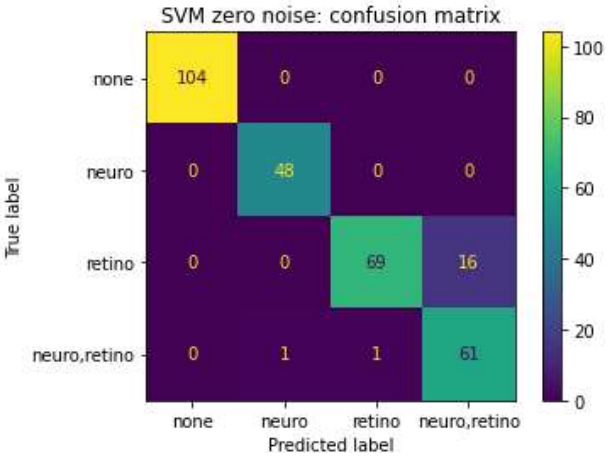


Figure 4. 11 SVM: Confusion Matrix (without noise)

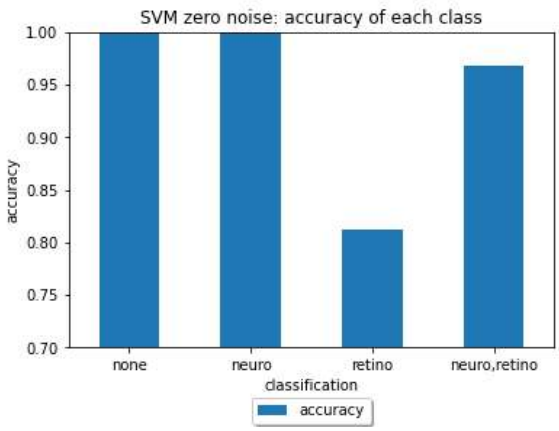


Figure 4. 12 SVM: Accuracy for each class (without noise)

- **New Testing Data Set and Addition of Noise**

The next step was to create a new set of data comprising 450 patients and to test the algorithm again, keeping the same parameters defined as optimal. The results obtained are shown in Annex 3. For now, it is just important to analyze that both the accuracy and the weighted average of the F1-score decrease from 0.94 to 0.85. A decrease of about 10% of the performance of the predictor might be indicative of some overfitting. Hence, the dataset was corrupted with some noise to avoid overfitting, as explained next.

Adding noise might prevent overfitting, since the algorithm trains with data that does not follow what is "expected", making it closer to real life conditions and, hopefully, less biased. First, it was studied how adding 5% of noise influenced the results. In a second scenario, 10% of noise was considered. Note that, with this addition of noise, the best hyperparameters for the algorithm might change, since it involves the creation of a different hyperplane than the initial one. So, again, the best hyperparameters were optimized as described before. Table 4.10 summarizes the results:

Table 4. 10 Best Parameters and Performances of the SVM under Different Noise Conditions

	0% noise (=previous results)	**5% noise**	**10% noise**
Best parameters	C=0.03; gamma=0.15; kernel=polynomial	C=0.07; gamma=0.11; kernel=polynomial	C=10; gamma=0.15; kernel=rbf
Accuracy (30% testing set of the initial data)	0.94	0.91	0.87
F1-score (30% testing set of the initial data)	0.94	0.91	0.87
Accuracy (new data)	0.85	0.82	0.85
F1-score (new data)	0.85	0.82	0.86

The effect of the addition of noise is visible: 5% noise does not seem to be enough, but in the 10% noise addition, it is already visible that the accuracy and F1-scores for the never-before-seen data is improved. Similar plots to non-noise conditions can be obtained, but their analysis was not considered relevant in this part of the work. The full set of results for these scenarios is shown in Annex 3.

- **Learning Curve and Iterations**

Additional plots were obtained for a more detailed analysis of the learning process of the algorithm. The Learning Curve is used to show how the model performs with an increasing number of training samples, monitoring the accuracy on the training dataset and on a hold-out validation dataset, after each update during training. It is mostly used to detect issues in the learning process, including underfitting and overfitting [91].

Underfitting occurs when the model cannot learn from the training data: it is visible in a learning curve if both the training and the cross validation have low accuracies, which is not the case, since they both reach above 0.9. Overfitting, as explained, is when the model learns the training data too well and does not have a good performance on new data, since it cannot generalize its predictions. It is visible in the learning curve if the there is a big gap between the training and the validation accuracy which, again, is not the case, since the curves eventually overlap. So, from the learning curve obtained in Figure 4.13, one can assess that the model had a good learning behavior, with no overfitting (has had been previously concluded) nor underfitting. Additionally, the chosen training data size (666) also seems good, since both the training and the validation accuracies had already stabilized at that point.

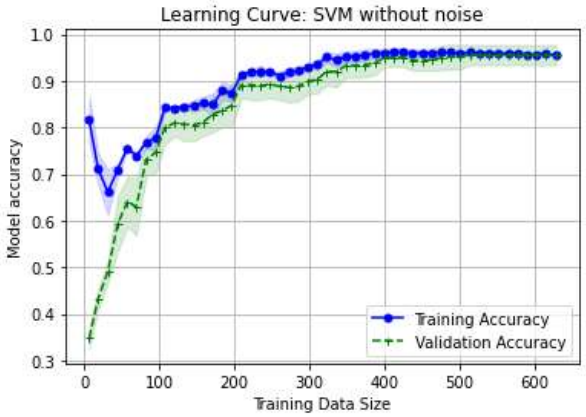


Figure 4. 13 SVM: Learning Curve (without noise)

A plot of the accuracy depending on the number of iterations was also obtained, as shown in Figure 4.14. In ML, the number of iterations refers to the number of times the parameters of the models are updated. To run the algorithm, the default value of this parameter was used, which is "no limit". But it was interesting to analyze from what value onwards it is unnecessary to keep iterating, since the accuracy stabilizes. That value in this case is of about 100 iterations. In real life context, where more data with more variables is hopefully available, making this sort of studies might be useful to put a reasonable limit on the number of iterations and save computation time.

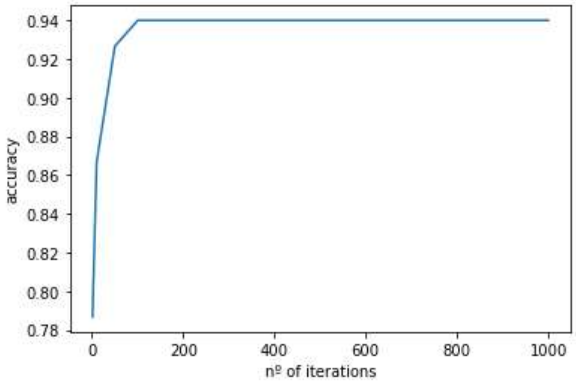


Figure 4. 14 SVM: Accuracy vs number of iterations (without noise)

RF: Results

- **Finding the Best Parameters**

GridSeachCV was also used to find the best parameters for the RF algorithm and produced the following output, presented in Table 4.11 and observed in the plot of Figure 4.15.:

Table 4. 11 Hyperparameter Optimization Results- RF

Hyperparameter	Best Value
Maximum depth	4
Maximum number of Features	2

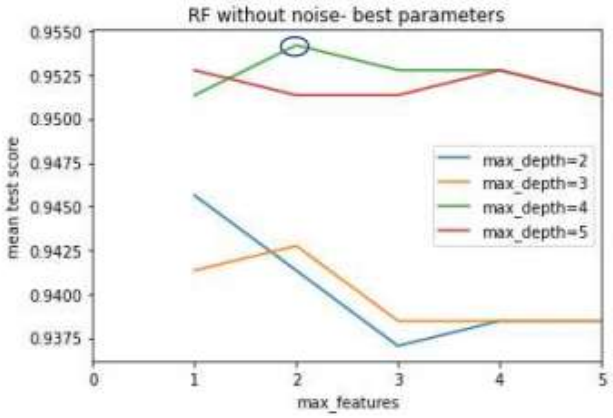


Figure 4. 15 Hyperparameter Tuning for RF (without noise)

Again, these parameters might change every time the algorithm runs, but after running several times, the metrics shown next do not have significant changes, therefore these parameters can be adopted.

- **Metrics and Effect of Adding Noise**

The plots obtained for the RF are similar to those shown in the SVM section and can be consulted in Annex 3. However, it is just relevant for the context of this work to present Table 4.12, with the different performances according to the level of noise:

Table 4. 12 Best Parameters and Performances of the RF under Different Noise Conditions

	0% noise (=previous results)	5% noise	10% noise
Best parameters	Max depth=4; max features=2	Max depth=7; max features=2	Max depth=6; max features=2
Accuracy (30% testing set of the initial data)	0.93	0.92	0.87
F1-score (30% testing set of the initial data)	0.93	0.92	0.87

Accuracy (new data)	0.84	0.80	0.96
F1-score (new data)	0.85	0.81	0.96

Similarly to what happened with SVMs, adding 5% noise did not have a positive effect, but with 10% noise the performance of the algorithm is even better than before. The accuracy and F1 score for new data are, for the first time, higher than for the original data and were, actually, the highest values obtained during the entire analysis.

4.2.1.3.6 Conclusions

It is important to keep present that this entire analysis was based on a very simplistic version of the data that is planned to be obtained and the main goal was to provide insights about the application of this type of methods in real life context, when real data with the proposed input variables starts to be collected.

This process consisted in hyperparameter tuning to make sure the algorithm was using the parameters that achieved better performance. Important metrics, including accuracy and F1-score were also analyzed to check the performance of the algorithms. These values were overall high, which means that the algorithms were learning well and making good predictions from the data created according to literature. Additionally, some techniques were used to analyze the presence of and/or avoid overfitting. It was concluded that when the initial data learns from a noisier dataset it allows better generalizations for predicting unknown data. A noisier dataset might be more similar to the data obtained in real life context that will, in principle, be more complex, as opposed to the data used here, created according to defined rules.

Regarding the differences between the SVM and RF algorithm, both achieved good performance. However, it is important to highlight the performance of the RF with the 10% noise dataset: not only it seemed to avoid overfitting, but it achieved the best overall accuracy and F1 score.

However, as mentioned, this data is a simpler version with less input variables than the data that is planned to be used, which makes it complicated to generalize RF as the best algorithm for now. In the future, if such data is obtained, both algorithms should be tested again and a similar analysis to the one performed here should be made.

4.2.2 Data Mining for the Prediction of Hypoglycemia

Both METABO and gluCModel have incorporated models that predict blood glucose values and a CDSS component associated, to provide alerts if it detects an adverse event is coming. This undoubtedly provides value to healthcare management, since it gives time to the patient to take action and prevent even a life-threatening situation before it happens. Therefore, one wanted to make sure the proposed solution had a similar feature.

There were, however, some concerns that not enough data would be collected to train a model for this. Most of the studies presented in the state of the art took as input data collected with CGM, which is constantly collecting glycemic values, generating more data than SBGM measurements, collected around twice or three times a day, which is the proposal of the present work. Besides that, METABO and gluCModel, for instance, deal mostly with insulin-dependent patients. The interactions between insulin and the amount of CHO ingested also influence the glycemic outcomes, being therefore an input of value for such a predicting algorithm. The proposed use case, however, aims to deal with non-insulin patients, only taking anti-hyperglycemic medication.

Nevertheless, the study presented in State of the Art [54], in Section 2.3.3.2, seems promising: with only SBGM data from the prior seven days, it could predict if a hypoglycemic event would occur in the eighth day. Adding medication-related data, it could predict the hour that the event would occur in the eighth day, with 90% sensitivity and specificity. The authors concluded that adding more data, including data related to physical activity and nutrition would probably improve the performance.

METABO and gluCModel make this problem a regression one and predict a numeric value for blood glucose in the future, classifying it then into hypo or hyperglycemia, according to the value, and warning the patient. It seems clear that, for the problem at hands, it would be simpler to make it a classification problem, like the one described in Section 2.3.3.2, i.e, determine if a hypoglycemia will occur or not.

In the proposed solution, SBGM data will be collected by the daily measurements taken by the patient that are sent (or manually input) to the system. The system also has information about the type and doses of medications that the patient takes and the hours they are taken, since that is already available for consultation in the My Care Plan, for example. This is already the information that the study from Section 2.3.3.2 took, obtaining good results in predicting the hour hypoglycemia would occur in the eighth day.

Additionally, from the proposed solution, even more information could be fed into the system. One can obtain information regarding physical activity in the previous seven days as well, such as number of steps and calories, for example, since that data is being collected by the sensors and sent to the system. Data related to nutrition, namely the CHO values, can also be interesting for such an algorithm, since it directly affects glycemia. It is possible to obtain those values in the proposed solution from the nutrition app: every meal planed has its nutritional values defined. However, it is important to keep present that doing this means fully trusting the patient that he takes his meals according to plan, with no deviations. The data that the proposed system could provide to an algorithm for hypoglycemia prediction is summarized in Table 4.12:

Table 4. 12 Summarization of the Information to be Used as Input for the Hypoglycemic Prediction Algorithm

Input Variables for the Algorithm	**Information to be Collected**	**Origin of the Data**
SBGM Values	Values, timestamp	Glucometer

	Difference between consecutive values	
Medication	Type, dose, timestamp	Drug Prescription
Physical Activity	Nº of steps for each day	Physical Activity Bracelet
CHO	Value in a meal, timestamp meal was eaten	Nutrition App

That study from Chapter 2 used SVMs and RFs for the classification and they can both be tested for this purpose and experimented with different time horizons for prediction, in order to see if they have a good performance. However, at first, doing the analysis with data from the previous seven days and predict the occurrence of hypoglycemia in the eight day with an hour window can be an option, since it obtained good results in the example studied. Another consideration is related to the fact that the models can be both individual or generalized. It makes sense to have an individual model for each patient, just receiving data specific from him/her, since every metabolism is different.

4.3 Legal Considerations

4.3.1 General Data Protection Regulation

Security and data privacy are a significant concern regarding applications in healthcare. Health data must be compliant to the General Data Protection Regulation (GDPR), which is Europe's framework for legality concerning data protection. According to GDRP, personal data is any information related to an individual that can be identifiable directly or indirectly and there is a set of 6 contexts in which it is legal to use that sort of data, as one can consult in article 6 of the GDPR [92].

There is also a specification for "health concerning data", defined as personal data related to the *"physical or mental health of a natural person, including the provision of health care services, which reveals information about his or her health status"* [92]. The data that will be collected in the proposed solution falls into this category. Article 9 of the GDPR [92] describes the contexts in which it is legal to use this sort of data:

- The data subject gives explicit consent
- Processing the data is a need for preventive or occupational medicine, for assessing the working abilities of the employee, diagnosis, provision of health, treatment or the management of health or social care systems and services.
- Processing the data is necessary for public interest in terms of public health: protecting against cross-border threats or to ensure high standards for quality and safety regarding health care.

Note that, in point 1, the article mentions "explicit consent" and not "consent" alone. This means the patient needs to be informed in great detail of the uses of the data when giving consent. GDPR is respected in point 4 (Section 4.1.1) of the use case. Every time a new party wants to

access the data, either the hospital or a new pharmacy, the patient must be informed of how and why his/her data will be used to give explicit consent. One expects the hospital and the pharmacies to be compliant to those uses that will be listed and to GDPR in general, only processing the data in the lawful ways predicted.

In Portugal in specific, Law number 58/2019 from Diário da República [93] ensures that, in the national legal order, the GDPR regulations are followed.

4.3.2 Medical Device Certificate

Another important remark concerns official medical devices. According to Infarmed, Portuguese drug and health products authority, a medical device is any instrument, equipment, software, implant, reagent, material or any other item intended by the manufacturer to be used, isolated or together with something else, in humans, for one or more of the following medical purposes: diagnosis, prevention, monitoring, prediction, prognosis, treatment or attenuation of a disease, injury or disability [94].

The softwares and the sensors of the proposed solution need to be classified as medical devices by Infarmed to be used in a large scale in the population for the purposes they were intended to and the process of certification can take some time. However, just for the purpose of a pilot research project, which is how it is thought to be initially implemented, they do not need that certification so, for now, everything can go ahead as planned. In the future, one expects to receive the certification from Infarmed to go large-scale.

4.4 Future Work

The next steps, obviously, will include the full implementation of the solution. For that, the technical team needs to develop the CDSS, develop or buy the CDR, acquire the sensors and wearables, implement the rules in the knowledge-based system and, finally,create all the APIs that do not exist yet for full interoperability. Once the solution is implementable, the next step is to study it with a group of 10 patients and, if approved, scale it. For its approval, it would be interesting to create a control group and an intervention group, with patients using the proposed app for disease management, and study the disease outcomes in both.

In this work, the focus is on type 2 diabetic patients who take medication and SBGM. Eventually, insulin-dependent patients could be explored (which would incorporate type 1 patients), but an additional feature for insulin management would have to be created. CGM would provide more data on the glycemic values of the patient, which would improve the ML algorithm related to the prediction of hypoglycemia. If such data was available, instead of predicting hypoglycemias alone in a classification task, one could aim to predict glycemic values in general and have warnings for critical hyperglycemias as well.

Another possible consideration for the future is related to education regarding the disease. Both METABO and gluCModel have a strong focus on that part, either through PDF files with the relevant information or quizzes. This subject should be explored in the future, with the

medical team and perhaps the first group of patients to experiment the solution, in order to understand what would make sense and bring value.

Again, regarding the value of the solution, throughout this work considerations on value-based healthcare were made. Point 21 of the use case (Section 4.1.1) mentions questionnaires on the perception of well-being and usefulness on the patient's side. It is still unclear if those questionnaires can be adapted from others that already exist, from the WHO for example, or if they should be created specifically for this, making questions on particular aspects of the solution. The proposal from ICHOM, presented in Section 2.7, that uses questionnaires to assess general well-being, depression and diabetes specific emotional distress, could be a promising start. This is another issue that should be explored in the future with the medical team and maybe the first group of patients to experiment it.

This issue was already explored in section 4.2.1.3, related to the proof of concept, but it is worth mentioning again that, even though the algorithms performed well in the dataset that was created and some conclusions could be drawn regarding the best parameters and models, the analysis should be made again if real data is collected. Note, however, that training and testing the ML algorithms for this will only be possible when there is a considerable number of patients using it, since a lot of data is required to feed the models. Additionally, models that provide information on the underlying causes that lead to a complication could be explored. Both RF and SVMs are black boxes: they give accurate predictions, but their underlying reasons are not explained. Symbolic Classification is an example of a method that could be used to explore the reasons behind the classifications

Finally, even though this entire work was focused on T2DM, the functionalities proposed for the solution aimed to be as agnostic as possible. The ConnectedHealth project considered some study cases for exploration and the management of chronic diseases, T2DM in specific, is just one of them. In theory, the app could simply serve as a "central hub of health" even for fully healthy individuals, just collecting their health data. But, even in the subject of chronic diseases, it is being equated to use the solution for chronic heart failure. Everything that was proposed for the mock-ups in Section 4.1.4 could easily be adapted: for example, instead of showing glycemic values, the main screens could focus on heart rate or other relevant values. Obviously, regarding the CDSS, some further investigation would need to take place to understand what alerts and prediction models would make sense in the context of that disease.

5. Conclusions

Living in the era of digitalization and big data, new opportunities are arising to make a use of the data generated every day that can ultimately benefit society. Ever since EHRs appear, digitalizing health-related data, that opportunity also appeared in the healthcare sector. CDRs are a structured way to store the data and, if associated with a CDSS, health related decisions and actions can be enhanced. That can be achieved either from a CDSS built with predefined rules that detect or warn for harmful events or a CDSS incorporating ML algorithms that detect patterns that humans are not able to see, making powerful predictions. Those predictions might warn a healthcare professional to pay closer care and attention to a patient at risk.

The objectives of this work were first to understand how CDS is currently being used for T2DM and then propose how to apply it in the context of the ConnectedHealth project, proposed by Glintt. The first objective was accomplished in the State of the Art: there was an exploration of how CDS, both knowledge and non-knowledge based, is being used to improve disease outcomes for those patients, either when applied in a more clinical context or in daily-life management. The second objective was taken to another level, since the project was still at a very early stage. Considering the goals of the entire project and the existing Glintt solutions, not only the CDS was proposed, but the other aspects of the desired solution that would have an impact in the way T2DM is managed, including all the wearables necessary and all the features the app would need to have. It was important to specify those aspects since they provide the data and information that feeds the CDSS. To accomplish all this, an use case was built, together with the functional requirements, and a dataflow diagram was specified later on.

As mentioned, the biggest focus was to improve and bring value to the way this disease is managed nowadays. One considers this was achieved in several aspects of the solution proposed, as will be specified in the following paragraphs:

Traditional healthcare systems are focused on handling acute systems, which is not the best model for a chronic disease, that presents some aspects that need to be managed daily. This work proposed a solution that aims to manage these patients outside the physical boundaries of the hospital, while involving different healthcare professionals in a close follow-up of the patient.

Healthcare data from different sources will be gathered in a single platform and is owned by patient, who authorizes the respective professionals to access it when needed. Eventually, this also aims to target existing inefficiencies in the Portuguese healthcare system, where exams are duplicated, for example, because different healthcare institutions have different systems that have difficulties interoperating.

Another interesting part of the solution is the involvement of the community pharmacy: their potential shall be explored, since they are closer with the population and can have a more active role in the overall health of the patients they attend to. Therefore, it made sense to include pharmacists in a solution that aims to bring different parties together to ultimately improve population health outcomes.

Also, as explained, currently the involvement of healthcare professionals in the daily life of the patients is very sparse. In the best practices for DM, the patient registers his/her glycemic values in an external platform to the healthcare institution and it is not associated with any alerts to warn that something harmful happened. The proposed solution aims to involve healthcare professionals, especially the nurse responsible for the follow-up of the diabetic patient, in more aspects of the lifestyle. First of all, there is an official and digitalized “prescription of attitudes”, that goes beyond the basic education on these aspects that is discussed in appointments with the patient. Being a disease so influenced by lifestyle aspects, including nutrition and physical activity, it makes sense to treat these the same way as medication is treated. The healthcare professionals can prescribe how often measurements of glycemia should be taken; the physical activity patients should have daily and also nutrition recommendations. That information can later be consulted by the patient in the “My Care Plan”, together with their drug prescription. “My Care Plan” is a feature of the proposed application that provides guidance throughout the day, facilitating disease management and easily consulted by the patient.

One of the biggest innovations for diabetes in the Portuguese healthcare, and the main focus of this work, is the incorporation of Clinical Decision Support for the daily management of the disease. The alerts suggested aim to prevent and warn for adverse situations; improve adherence to the prescribed treatment (both in terms of drugs and physical activity, for example) and also to have a control on the patient’s attendance to the necessary appointments and exams. The appointments should be scheduled in advance, by the main doctor following the patient, and the system warns a few days before about it and also sends an alert to the nurse if there was no attendance, who can later contact the patient. This is also an improvement of what happens nowadays: for instance, in an appointment with an endocrinologist, he can advise the patient to consult an ophthalmologist and to return to another endocrinology appointment in 6 months. However, there is no control if the patient will do that. The idea with the solution is that the main physician that follows the patient would automatically schedule those appointments in the system and then the proposed controls would take place.

This solution also aims to change the current healthcare spectrum by providing a more personalized medicine. The targets that generate alerts can eventually be personalized; each patient has a nurse regularly checking for his/her progress individually and the data mining aims to understand individual risk of having an hypoglycemic event, according to the latest occurrences in the patient’s daily life, and also to predict the appearance of complications by analyzing individual risk factors.

The proposed solution will aggregate a big amount of healthcare data from different patients. This data is planned to be used to provide valuable insights and, eventually, generate knowledge (data mining). While in the knowledge-based solution, the methodology included a validation and feedback component together with the medical team, the non-knowledge based was just briefly discussed, since the real proposal should be made when real data to work with starts to be collected at a large scale. However, the doctors offered some resistance when ML algorithms were mentioned. Overall, Artificial Intelligence is still a sensitive subject in the medical

community: some undoubtedly can see its potential benefits, some still fear the role of a computer in providing medical insights. But one has to keep in mind that Artificial Intelligence solutions are becoming more and more efficient, and the goal of Clinical Decision Support is never to replace the doctor, but to help him make decisions. AI is the future across many areas and its potential benefits will start to be better understood over time. Therefore, data mining was still included in this work. One example of what happened was, when talking about using algorithms to predict the appearance of microvascular complications, one of the doctors argued that sometimes their appearance can be very random and, therefore, he did not consider it would be useful. But that is exactly why it is probably useful: these algorithms can detect patterns that are not easily detected by humans. Some literacy regarding these technologies still needs to occur, but the data mining proposal is still regarded as a value offer for the future.

It is also important to make some considerations on interoperability. The problem with previous CDS project was their lack of integration between different systems. Every time a new institution wished to adopt it, an effort probably needed to exist to develop it from scratch, adapted to their own health information system. This was a barrier for their development and, as a consequence, a barrier to the provision of the best possible healthcare for patients. Throughout the proposed solution, one ensures the use of the latest and most promising standard for interoperability and communication between health information systems: FHIR. However, not many practical examples of current CDSS built on FHIR were found in the market, probably because it is still a recent standard. In the future, if there is a wish to scale the CDSS to be developed even outside Glintt that should, in theory, be possible.

Finally, some considerations should be made regarding value-based healthcare. The ultimate goal of the solution is to create value for the patient. It should be assessed whether the patient perceives that this solution is improving his/her health outcomes. Patient's feedback should be collected and, if necessary, taken into account to adapt the solution in order to provide the biggest value possible for those dealing with chronic diseases.

References

[1] Nugent, R. (2008). Chronic diseases in developing countries: Health and economic burdens. In *Annals of the New York Academy of Sciences* (Vol. 1136). https://doi.org/10.1196/annals.1425.027

[2] World Health Organization (2021). *Diabetes.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/diabetes

[3] Thimbleby, H. (2013). Technology and the future of healthcare. *Journal of Public Health Research*, *2*(3). https://doi.org/10.4081/jphr.2013.e28

[4] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. In *npj Digital Medicine* (Vol. 3, Issue 1). https://doi.org/10.1038/s41746-020-0221-y

[5] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. In *Computational and Structural Biotechnology Journal* (Vol. 15). https://doi.org/10.1016/j.csbj.2016.12.005

[6] Kharroubi, A. T. (2015). Diabetes mellitus: The epidemic of the century. *World Journal of Diabetes*, *6*(6), 850. https://doi.org/10.4239/wjd.v6.i6.850

[7] Harvard Health Publishing (2018). *Type 2 Diabetes Mellitus.* Retrieved from https://www.health.harvard.edu/a_to_z/type-2-diabetes-mellitus-a-to-z

[8] DeFronzo, R. A., Ferrannini, E., Groop, L., Henry, R. R., Herman, W. H., Holst, J. J., Hu, F. B., Kahn, C. R., Raz, I., Shulman, G. I., Simonson, D. C., Testa, M. A., & Weiss, R. (2015). Type 2 diabetes mellitus. *Nature Reviews Disease Primers*, *1*. https://doi.org/10.1038/nrdp.2015.19

[9] Raposo, J.F. (2020). Diabetes: Factos e Números 2016,2017 e 2018. *Revista Portuguesa de Diabetes. 2020; 15 (1): 19-27.* Retrieved from http://www.revportdiabetes.com/wp-content/uploads/2020/05/RPD-Mar%C3%A7o-2020-Revista-Nacional-p%C3%A1gs-19-27.pdf

[10] Grilo, A. M., dos Santos, M. C., Gomes, A. I., & Rita, J. S. (2017). Promoting Patient-Centered Care in Chronic Disease. In *Patient Centered Medicine*. https://doi.org/10.5772/67380

[11] Fico, G., Arredondo, M. T., Protopappas, V., Georgia, E., & Fotiadis, D. (2015). Mining data when technology is applied to support patients and professional on the control of chronic diseases: The experience of the METABO platform for diabetes management. *Methods in Molecular Biology*, *1246*, 191–216. https://doi.org/10.1007/978-1-4939-1985-7_13

[12] Epstein, R. M., & Street, R. L., Jr (2011). The values and value of patient-centered care. *Annals of family medicine*, *9*(2), 100–103. https://doi.org/10.1370/afm.1239

[13] Peart, A., Lewis, V., Barton, C., Brown, T., White, J., Gascard, D., & Russell, G. (2019). Providing person-centred care for people with multiple chronic conditions: Protocol for a qualitative study incorporating client and staff perspectives. *BMJ Open*, *9*(10). https://doi.org/10.1136/bmjopen-2019-030581

[14] Associação Portuguesa de Administradores Hospitalares (2019). BUSINESS INTELIGENCE NO SNS: Principais Desafios, https://apah.pt/ebook/business-intelligence-do-sns-principais-desafios/

[15]Teisberg, E., Wallace, S., & O'Hara, S. (2020). Defining and Implementing Value-Based Health Care: A Strategic Framework. *Academic medicine : journal of the Association of American Medical Colleges*, *95*(5), 682–685. https://doi.org/10.1097/ACM.0000000000003122

[16] Baumhauer, J. F., & Bozic, K. J. (2016). Value-based Healthcare: Patient-reported Outcomes in Clinical Decision Making. *Clinical orthopaedics and related research*, *474*(6), 1375–1378. https://doi.org/10.1007/s11999-016-4813-4

[17] ICHOM. *Our Mission*. Retrieved from: https://www.ichom.org/mission/

[18] Middleton, B., Sittig, D. F., & Wright, A. (2016). Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. *Yearbook of Medical Informatics*. https://doi.org/10.15265/IYS-2016-s034

[19]Wasylewicz, A. T. M., & Scheepers-Hoeks, A. M. J. W. (2018). Clinical decision support systems. In *Fundamentals of Clinical Data Science* (pp. 153–169). Springer International Publishing. https://doi.org/10.1007/978-3-319-99713-1_11

[20] Peiffer-Smadja, N., Rawson, T. M., Ahmad, R., Buchard, A., Pantelis, G., Lescure, F. X., Birgand, G., & Holmes, A. H. (2020). Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. In *Clinical Microbiology and Infection* (Vol. 26, Issue 5). https://doi.org/10.1016/j.cmi.2019.09.009

[21] Khairat, S., Marc, D., Crosby, W., & al Sanousi, A. (2018). Reasons for physicians not adopting clinical decision support systems: Critical analysis. In *JMIR Medical Informatics* (Vol. 20, Issue 4). https://doi.org/10.2196/medinform.8912

[22] Marcial, L. H., Richardson, J. E., Lasater, B., Middleton, B., Osheroff, J. A., Kawamoto, K., Ancker, J. S., van Leeuwen, D., Lomotan, E. A., Al-Showk, S., & Blumenfeld, B. H. (2018). The Imperative for Patient-Centered Clinical Decision Support. *EGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, *6*(1). https://doi.org/10.5334/egems.259

[23]Wright, A., & Sittig, D. F. (2008). A four-phase model of the evolution of clinical decision support architectures. In *International Journal of Medical Informatics* (Vol. 77, Issue 10). https://doi.org/10.1016/j.ijmedinf.2008.01.004

[24] HL7 FHIR (2019). *FHIR Overview*. Retrieved from: https://www.hl7.org/fhir/overview.html

[25] HL7 FHIR (2021). *Introducing HL7 FHIR*. Retrieved from: http://hl7.org/fhir/2021May/summary.html

[26] HL7 FHIR (2019). *FHIR Overview-Architects*. Retrieved from: https://www.hl7.org/fhir/overview-arch.html

[27] CDS Hooks (2018). *Overview*. Retrieved from: https://cds-hooks.org/

[28] Semenov, I., Osenev, R., Gerasimov, S., Kopanitsa, G., Denisov, D., & Andreychuk, Y. (2019). Experience in Developing an FHIR Medical Data Management Platform to Provide Clinical Decision Support. *International journal of environmental research and public health*, *17*(1), 73. https://doi.org/10.3390/ijerph17010073

[29] Nadkarni, P. (2016). *Clinical Data Repositories: Warehouses, Registries, and the Use of Standards. Clinical Research Computing, 173–185.* doi:10.1016/b978-0-12-803130-8.00

[30] Smile CDR (2021). *About Smile.* Retrieved from: https://www.smilecdr.com/about-smile

[31] Smile CDR (2021). *The Data Liberation Platform.* Retrieved from: https://www.smilecdr.com/smilecdr

[32] National Health Service (2020). *Low Blood Sugar (hypoglycaemia).* Retrieved from: https://www.nhs.uk/conditions/low-blood-sugar-hypoglycaemia/

[33] Zimmerman RS. (2016). Diabetes mellitus: management of microvascular and macrovascular complications. *J Cleveland Clinic: Centers for Continuing Education.*

[34] Endocrine Web (2021). Type 2 Diabetes Complications. Retrieved from: https://www.endocrineweb.com/conditions/type-2-diabetes/type-2-diabetes-complications

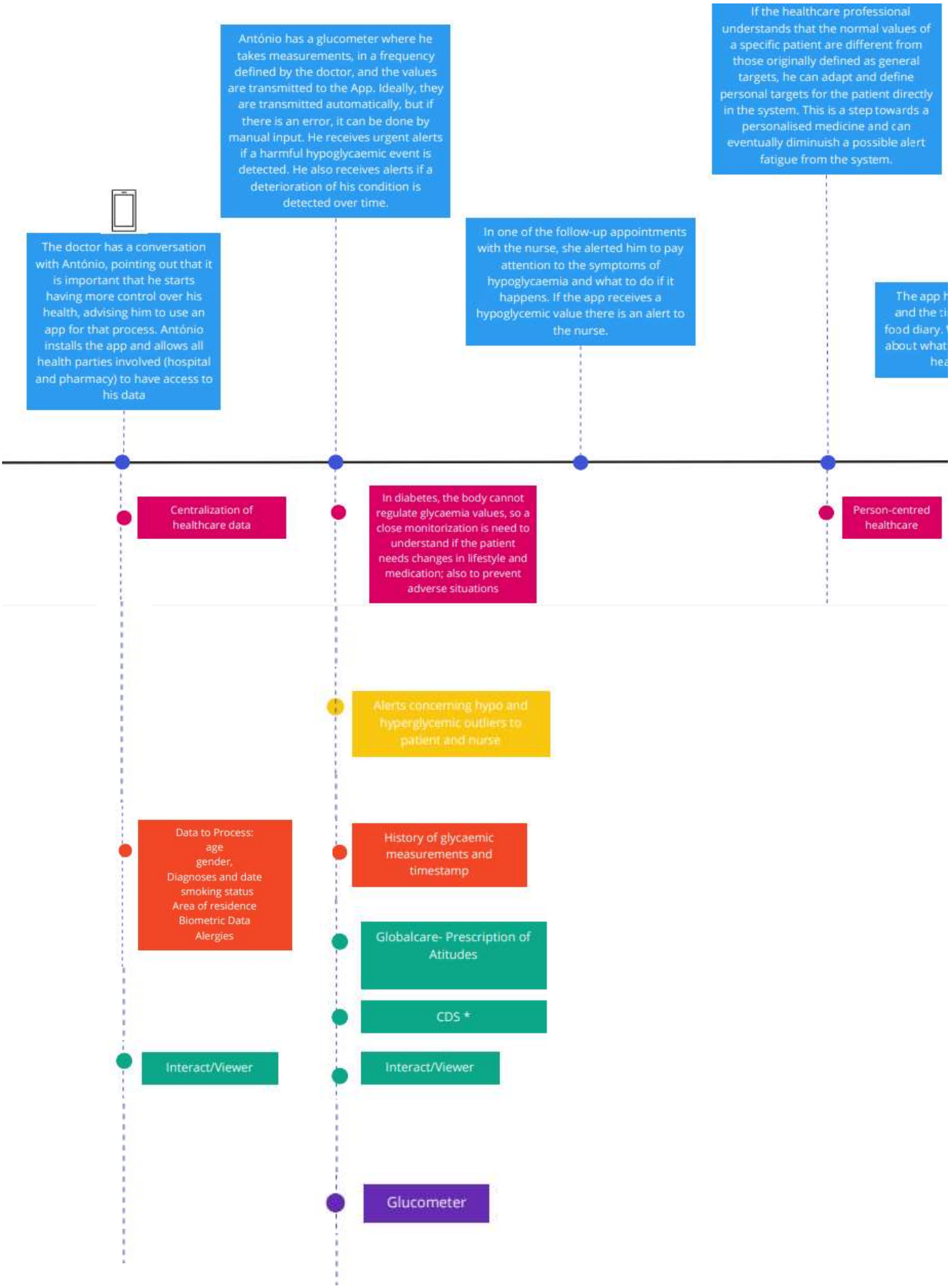
[35] Tracey, M. L., McHugh, S. M., Fitzgerald, A. P., Buckley, C. M., Canavan, R. J., & Kearney, P. M. (2016). Risk Factors for Macro-and Microvascular Complications among Older Adults with Diagnosed Type 2 Diabetes: Findings from the Irish Longitudinal Study on Ageing. *Journal of Diabetes Research*, *2016*. https://doi.org/10.1155/2016/5975903
[36] diaTribeLearn (2021).Get to know your Lab Tests: A1c, eGFR, UACR and more. Retrieved from: https://diatribe.org/get-know-your-lab-tests-a1c-egfr-uacr-and-more
[37] Kawamoto, K., & Lobach, D. F. (2005). Design, implementation, use, and preliminary evaluation of SEBASTIAN, a standards-based Web service for clinical decision support. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2005*, 380–384.
[38] Holbrook, A., Thabane, L., Keshavjee, K., Dolovich, L., Bernstein, B., Chan, D., Troyan, S., Foster, G., & Gerstein, H. (2009). Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. *CMAJ*, *181*(1–2). https://doi.org/10.1503/cmaj.081272
[39] O'Connor, Patrick J., Sperl-Hillen, J. A. M., Rush, W. A., Johnson, P. E., Amundson, G. H., Asche, S. E., Ekstrom, H. L., & Gilmer, T. P. (2011). Impact of electronic health record clinical decision support on diabetes care: A randomized trial. *Annals of Family Medicine*, *9*(1). https://doi.org/10.1370/afm.1196
[40] Ellahham, S. (2020). Artificial Intelligence: The Future for Diabetes Care. In *American Journal of Medicine* (Vol. 133, Issue 8). https://doi.org/10.1016/j.amjmed.2020.03.033
[41] Contreras, I., & Vehi, J. (2018). Artificial intelligence for diabetes management and decision support: Literature review. In *Journal of Medical Internet Research* (Vol. 20, Issue 5). https://doi.org/10.2196/10775
[42] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. In *Computational and Structural Biotechnology Journal* (Vol. 15). https://doi.org/10.1016/j.csbj.2016.12.005
[43] Marinov, M., Mosa, A. S. M., Yoo, I., & Boren, S. A. (2011). Data-mining technologies for diabetes: A systematic review. In *Journal of Diabetes Science and Technology* (Vol. 5, Issue 6). https://doi.org/10.1177/193229681100500631
[44] Khanam, P. A., Hoque, S., Begum, T., Habib, S. H., & Latif, Z. A. (2017). Microvascular complications and their associated risk factors in type 2 diabetes mellitus. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *11*, S577–S581. https://doi.org/10.1016/j.dsx.2017.04.007
[45] Cheema, S., Maisonneuve, P., Zirie, M., Jayyousi, A., Alrouh, H., Abraham, A., Al-Samraye, S., Mahfoud, Z., Al-Janahi, I. M., Ibrahim, B., Lowenfels, A. B., & Mamtani, R. (2018). Risk factors for microvascular complications of diabetes in a high-risk middle east population. *Journal of Diabetes Research*, *2018*. https://doi.org/10.1155/2018/8964027
[46] Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., de Cata, P., Chiovato, L., & Bellazzi, R. (2018). Machine Learning Methods to Predict Diabetes Complications. *Journal of Diabetes Science and Technology*, *12*(2). https://doi.org/10.1177/1932296817706375
[47]Fiarni, C., Sipayung, E. M., & Maemunah, S. (2019). Analysis and prediction of diabetes complication disease using data mining algorithm. *Procedia Computer Science*, *161*. https://doi.org/10.1016/j.procs.2019.11.144
[48] MOSAIC Project (2013). MOSAIC Project. Retrieved from: http://www.mosaicproject.eu/index.html
[49] Felizardo, V., Garcia, N. M., Pombo, N., & Megdiche, I. (2021). Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction – A systematic literature review. In *Artificial Intelligence in Medicine* (Vol. 118). https://doi.org/10.1016/j.artmed.2021.102120
[50] Mujahid, O., Contreras, I., & Vehi, J. (2021). Machine learning techniques for hypoglycemia prediction: Trends and challenges. In *Sensors (Switzerland)* (Vol. 21, Issue 2). https://doi.org/10.3390/s21020546
[51] Kodama, S., Fujihara, K., Shiozaki, H., Horikawa, C., Yamada, M. H., Sato, T., Yaguchi, Y., Yamamoto, M., Kitazawa, M., Iwanaga, M., Matsubayashi, Y., & Sone, H. (2021). Ability of current machine learning algorithms to predict and detect hypoglycemia in patients with diabetes mellitus: meta-analysis. *JMIR Diabetes*, *6*(1). https://doi.org/10.2196/22458
[52] Diabetes UK. Continuous Glucose Monitoring. Retrieved from: https://www.diabetes.org.uk/guide-to-diabetes/managing-your-diabetes/testing/continuous-glucose-monitoring-cgm
[53] Benjamin, E. M. (2002). Self-Monitoring of Blood Glucose: The Basics. *Clinical Diabetes*, *20*(1). https://doi.org/10.2337/diaclin.20.1.45
[54] Sudharsan, B., Peeples, M., & Shomali, M. (2015). Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *Journal of Diabetes Science and Technology*, *9*(1). https://doi.org/10.1177/1932296814554260
[55] Hidalgo, J. I., Maqueda, E., Risco-Martín, J. L., Cuesta-Infante, A., Colmenar, J. M., & Nobel, J. (2014). GlUCModel: A monitoring and modeling system for chronic diseases applied to diabetes. *Journal of Biomedical Informatics*, *48*. https://doi.org/10.1016/j.jbi.2013.12.015
[56] Andrès, Emmanuel, Meyer, L., Zulfiqar, A. A., Hajjam, M., Talha, S., Bahougne, T., Ervé, S., Hajjam, J., Doucet, J., Jeandidier, N., & Hajjam El Hassani, A. (2019). Telemonitoring in diabetes: evolution of concepts and technologies, with a focus on results of the more recent studies. In *Journal of medicine and life* (Vol. 12, Issue 3). https://doi.org/10.25122/jml-2019-0006
[57] Nutrition Principles and Recommendations in Diabetes. (2004). In *Diabetes Care* (Vol. 27, Issue SUPPL. 1).
[58] Colberg, S. R., Sigal, R. J., Yardley, J. E., Riddell, M. C., Dunstan, D. W., Dempsey, P. C., Horton, E. S., Castorino, K., & Tate, D. F. (2016). Physical activity/exercise and diabetes: A position statement of the American Diabetes Association. In *Diabetes Care* (Vol. 39, Issue 11). https://doi.org/10.2337/dc16-1728
[59] National Health Service (2020). Understanding Medicine: Type 2. Retrieved from: https://www.nhs.uk/conditions/type-2-diabetes/understanding-medication/
[60] Fioravanti, A., Fico, G., Arredondo, M. T., & Leuteritz, J. P. (2011). A mobile feedback system for integrated E-health platforms to improve self-care and compliance of diabetes mellitus patients. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. https://doi.org/10.1109/IEMBS.2011.6090591
[61] Georga, E., Protopappas, V., Guillen, A., Fico, G., Andigo, D., Arredondo, M. T., Exarchos, T. P., Polyzos, D., & Fotiadis, D. I. (2009). Data mining for blood glucose prediction and knowledge discovery in diabetic patients: The METABO diabetes modeling and management system. *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*. https://doi.org/10.1109/IEMBS.2009.5333635
[62] Hidalgo, J. I., Maqueda, E., Risco-Martín, J. L., Cuesta-Infante, A., Colmenar, J. M., & Nobel, J. (2014). GlUCModel: A monitoring and modeling system for chronic diseases applied to diabetes. *Journal of Biomedical Informatics*, *48*. https://doi.org/10.1016/j.jbi.2013.12.015
[63] Glintt (2016). Globalcare. Retrieved from: https://www.glintt.com/en/WHAT-WE-DO/Product-Services/SoftwareSolutions/Pages/GlobalCare.aspx

[64] Glintt (2021). Viewer. Retrieved from: https://viewer.glintt.com/
[65] Glintt (2016). Sifarma. Retrieved from: https://www.glintt.com/en/WHAT-WE-DO/Product-Services/SoftwareSolutions/Pages/Sifarma.aspx
[66] Healthy Systems. HS.Helios. Retrieved from: https://hltsys.pt/hs-helios-en/
[67] Pousinho, S., Morgado, M., Plácido, A. I., Roque, F., Falcão, A., & Alves, G. (2020). Clinical pharmacists´ interventions in the management of type 2 diabetes mellitus: A systematic review. *Pharmacy Practice*, *18*(3). https://doi.org/10.18549/PharmPract.2020.3.2000
[68] Félix, J., Ferreira, D., Afonso-Silva, M., Gomes, M. V., Ferreira, C., Vandewalle, B., Marques, S., Mota, M., Costa, S., Cary, M., Teixeira, I., Paulino, E., Macedo, B., & Barbosa, C. M. (2017). Social and economic value of Portuguese community pharmacies in health care. *BMC Health Services Research*, *17*(1). https://doi.org/10.1186/s12913-017-2525-4
[69] Nano, J., Carinci, F., Okunade, O., Whittaker, S., Walbaum, M., Barnard-Kelly, K., Barthelmes, D., Benson, T., Calderon-Margalit, R., Dennaoui, J., Fraser, S., Haig, R., Hernández-Jimenéz, S., Levitt, N., Mbanya, J. C., Naqvi, S., Peters, A. L., Peyrot, M., Prabhaharan, M., … Massi-Benedetti, M. (2020). A standard set of person-centred outcomes for diabetes mellitus: results of an international and unified approach. *Diabetic Medicine*, *37*(12). https://doi.org/10.1111/dme.14286
[70] Bryman, A. & Bell, E. (2011) "Business Research Methods" 3rd edition, Oxford University Press
[71] Business Research Methodology. Action Research. Retrieved from: https://research-methodology.net/research-methods/action-research/
[72] Medisana (2021). MediTouch 2 mg/dL. Retrieved from: https://www.medisana.com/en/Health-control/Blood-glucose-monitor/MediTouch-2-mg-dL-Blood-glucose-monitor-incl-starter-set.html
[73] Fibit(2021). Inspire 2. Retrieved from: https://www.fitbit.com/global/us/products/trackers/inspire2.
[74] Fraunhofer Portugal. Lifana. Retrieved from: https://www.aicos.fraunhofer.pt/en/our_work/projects/lifana.html
[75] HL7 FHIR (2019). Resource Observation - Detailed Descriptions. Retrieved from: https://www.hl7.org/fhir/observation-definitions.html
[76] HL7 FHIR (2019). Resource NutritionOrder - Detailed Descriptions. Retrieved from: https://www.hl7.org/fhir/nutritionorder-definitions.html
[77] HL7 FHIR (2019). Resource NutritionIntake - Detailed Descriptions. Retrieved from: http://hl7.org/fhir/2020Feb/nutritionintake-definitions.html
[78] HL7 FHIR (2019). Resource MedicationRequest - Detailed Descriptions. Retrieved from: http://www.hl7.org/fhir/medicationrequest-definitions.html
[79] HL7 FHIR (2019). Resource ServiceRequest- Detailed Descriptions. Retrieved from: https://www.hl7.org/fhir/servicerequest-definitions.html):
[80] HL7 FHIR (2019). Resource Appointment- Detailed Descriptions. Retrieved from: https://www.hl7.org/fhir/appointment-definitions.html
[81] HL7 FHIR (2019). Resource Diagnostic Report- Detailed Descriptions. Retrieved from: https://www.hl7.org/fhir/diagnosticreport-definitions.html
[82] Christmann A., Steinwart I. (2008). *Support Vector Machines*. Springer
[83] Towards Data Science (2018). *Support Vector Machine- Introduction to Machine Learning Algorithms.* Retrieved from: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47
[84] Breiman, L. (2001). Random Forests - Springer. *Machine Learning*, *45*(1).
[85] Towards Data Science (2019). *From a Single Decision Tree to a Random Forest.* Retrieved from: https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147
[86] Towards Data Science (2018). Accuracy, Precision, Recall or F1? Retrieved from: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9
[87] Towards Data Science (2020). GridSearchCV for Beginners. Retrieved from: https://towardsdatascience.com/gridsearchcv-for-beginners-db48a90114ee
[88] Scikit Learn. *RBF SVM Parameters*. Retrieved from: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html
[89] Scikit Learn. *Sklearn.ensemble.RandomForestClassifier*. Retrieved from: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[90] IBM (2021). *What is Overfitting?* Retrieved from: https://www.ibm.com/cloud/learn/overfitting
[91] Data Analytics (2020). *Learning Curves Explained with Python SkLearn Example*. Retrieved from: Learning Curves Explained with Python Sklearn Example - Data Analytics (vitalflux.com)
[92] Official Journal of the European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of27 April 2016. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679[
[93] Diário da República n.º 151/2019, Série I de 2019-08-08 (2019). Lei n.º 58/2019. Retrieved from: https://dre.pt/pesquisa/-/search/123815982/details/maximized
[94] Infarmed (2016). O que são dispositivos médicos? Retrieved from: https://www.infarmed.pt/web/infarmed/perguntas-frequentes-area-transversal/dm
[95] Kosiborod, M., Gomes, M. B., Nicolucci, A., Pocock, S., Rathmann, W., Shestakova, M. v., Watada, H., Shimomura, I., Chen, H., Cid-Ruzafa, J., Fenici, P., Hammar, N., Surmont, F., Tang, F., & Khunti, K. (2018). Vascular complications in patients with type 2 diabetes: Prevalence and associated factors in 38 countries (the DISCOVER study program). *Cardiovascular Diabetology*, *17*(1). https://doi.org/10.1186/s12933-018-0787-8
[96] Nabais C., Pereira J., Pereira P., Capote R., Morbeck S., Raposo J. (2011). Retinopatoa Diabética e Condições Associadas Que Relação? Um estudo em doentes com diabetes mellitus tipo 2. *Acta Med Port 2011; 24(S2): 71-78.* Retrieved from: https://research.unl.pt/ws/portalfiles/portal/3192443/1525_2175_1_PB.pd
[97] Pfannkuche, A., Alhajjar, A., Ming, A., Walter, I., Piehler, C., & Mertens, P. R. (2020). Prevalence and risk factors of diabetic peripheral neuropathy in a diabetics cohort: Register initiative "diabetes and nerves." *Endocrine and Metabolic Science*, *1*(1–2). https://doi.org/10.1016/j.endmts.2020.100053
[98] Sociedade Portuguesa de Diabetologia (2019). Diabetes: Factos e Números – O Ano de 2016, 2017 e 2018− Relatório Anual do Observatório Nacional da Diabetes
12/2019. Retrieved from: https://www.spd.pt/images/uploads/20210304-200808/DF&N-2019_Final.pdf

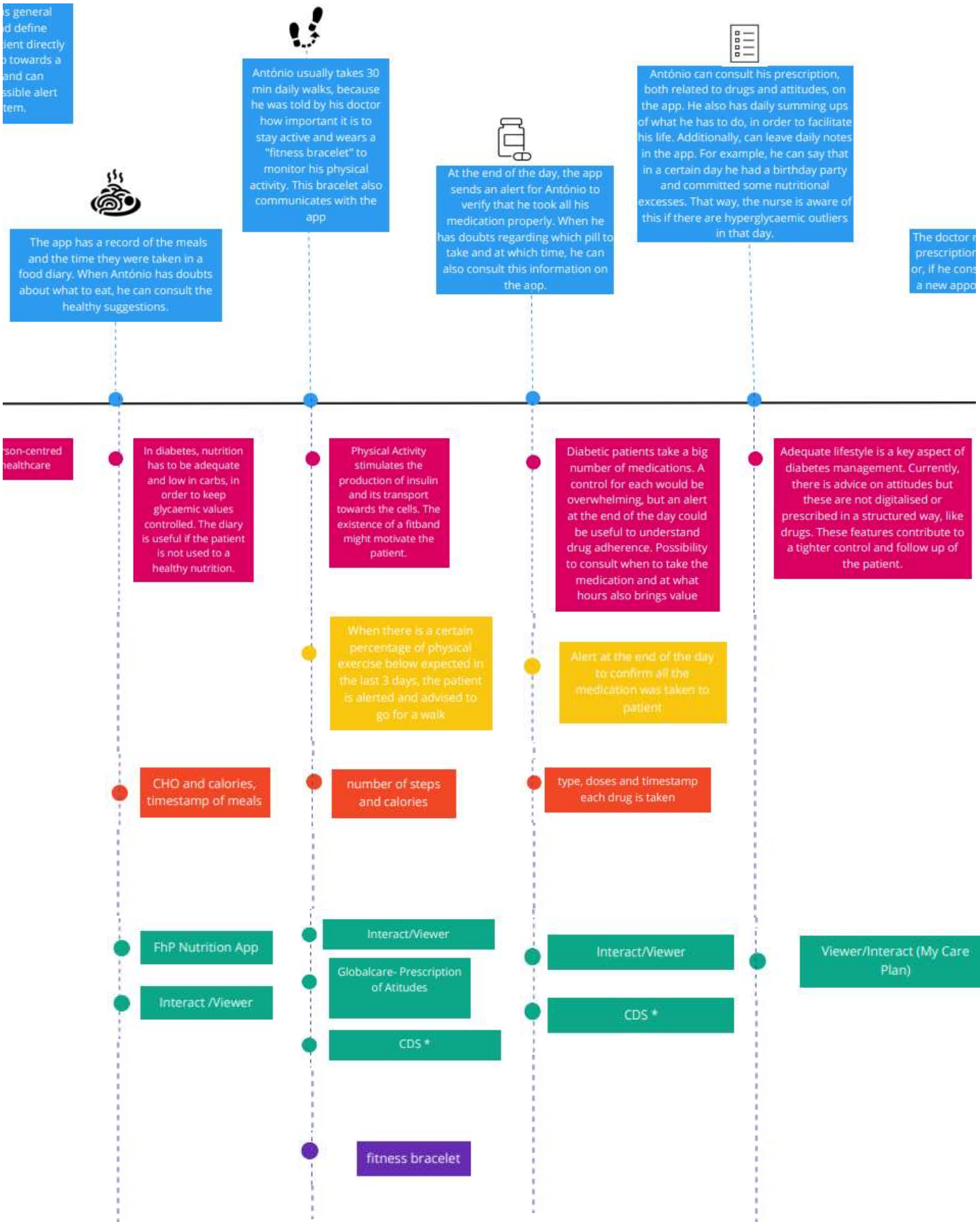
[99] Cheema, S., Maisonneuve, P., Zirie, M., Jayyousi, A., Alrouh, H., Abraham, A., Al-Samraye, S., Mahfoud, Z., Al-Janahi, I. M., Ibrahim, B., Lowenfels, A. B., & Mamtani, R. (2018). Risk factors for microvascular complications of diabetes in a high-risk middle east population. *Journal of Diabetes Research*, *2018*. https://doi.org/10.1155/2018/8964027
[100] Szumilas M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent*, *19*(3), 227–229.
[101] Sociedade Portuguesa de Diabetologia (2016). Diabetes: Factos e Números – O Ano de 2015− Relatório Anual do Observatório Nacional da Diabetes 12/2016. Retrieved from: https://www.sns.gov.pt/wp-content/uploads/2017/03/OND-2017_Anexo2.pdf

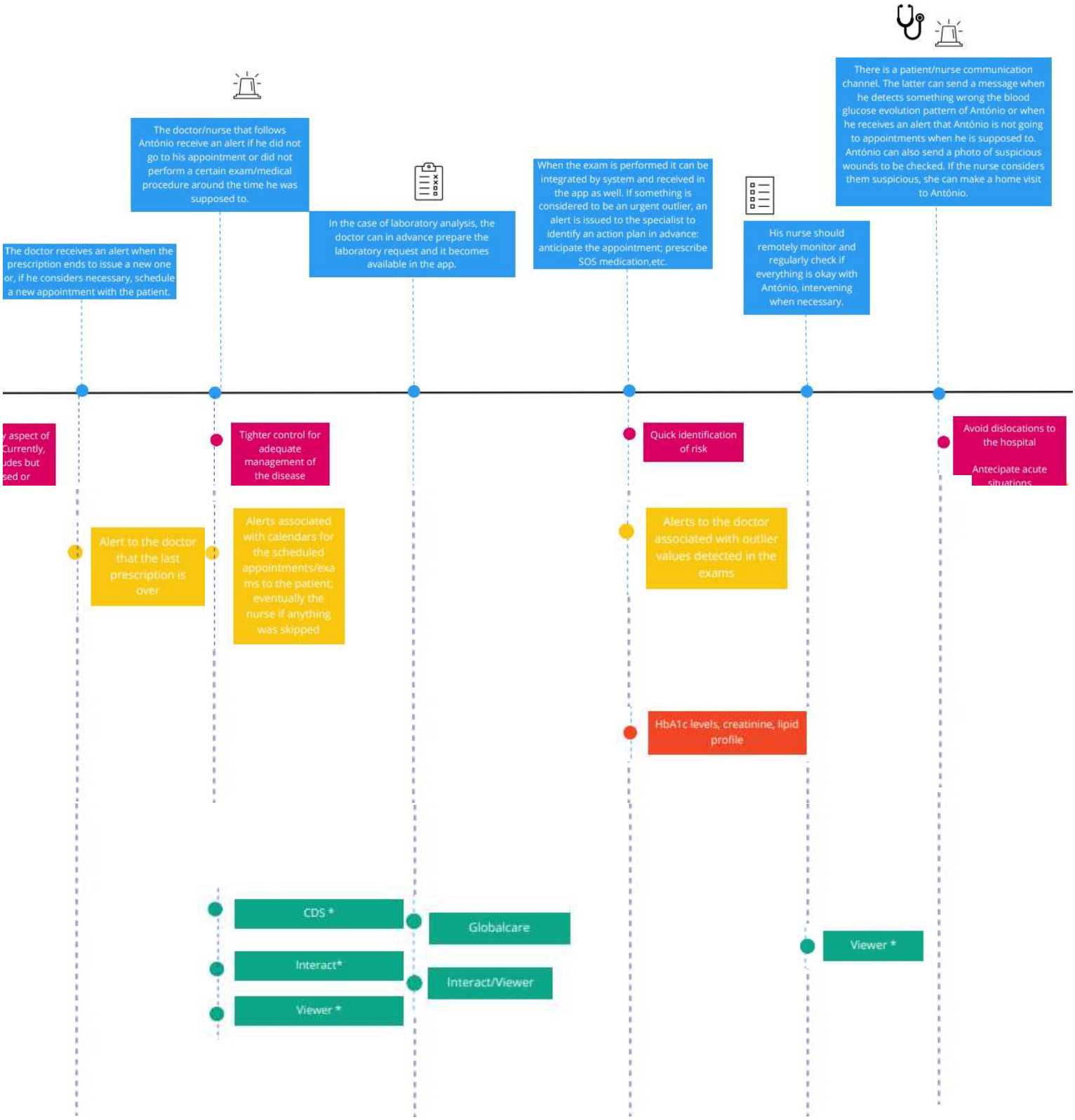
Annex 1

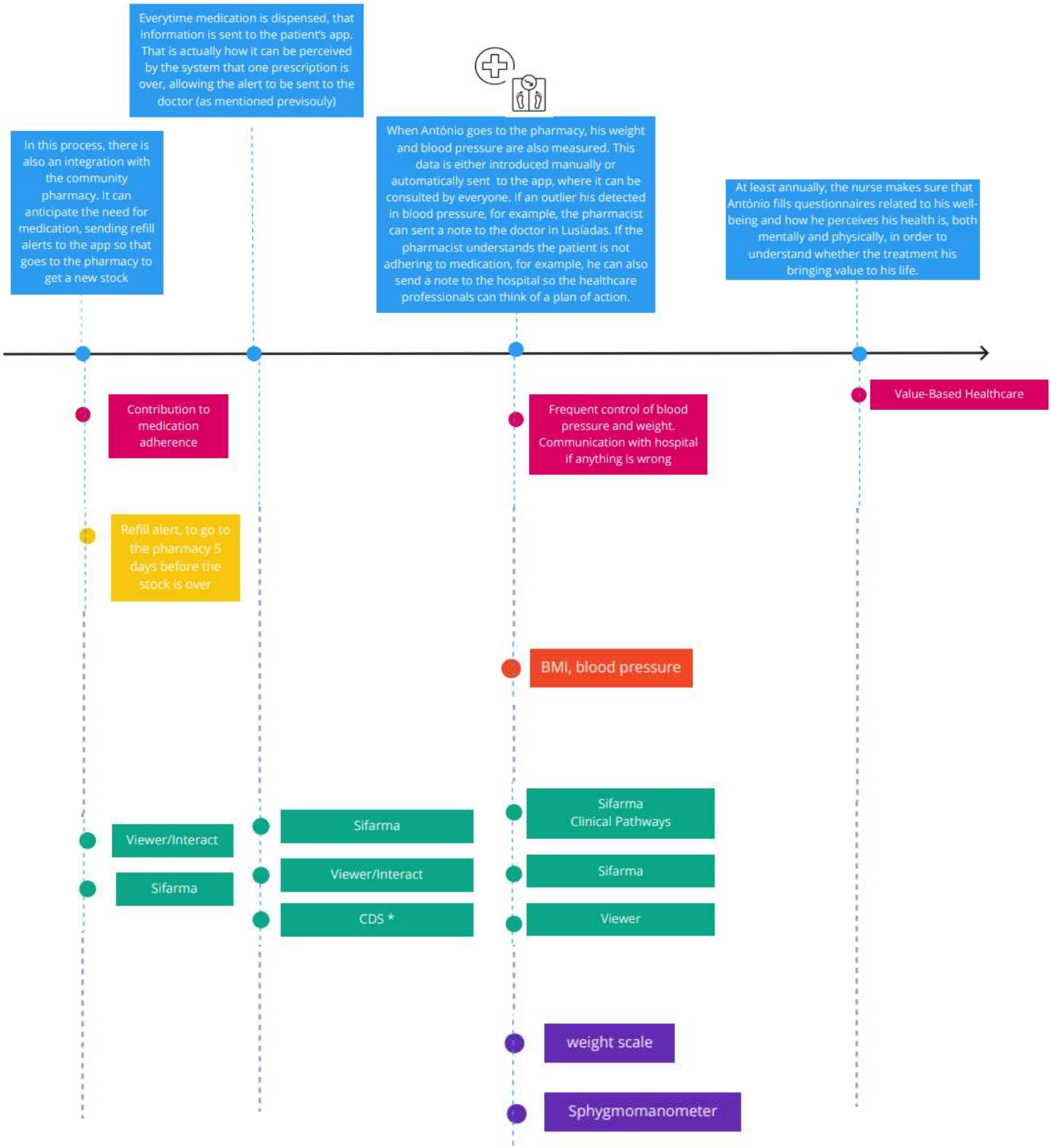
Use Case: Patient's Journey


The doctor has a conversation with António, pointing out that it is important that he starts having more control over his health, advising him to use an app for that process. António installs the app and allows all health parties involved (hospital and pharmacy) to have access to his data
António has a glucometer where he takes measurements, in a frequency defined by the doctor, and the values are transmitted to the App. Ideally, they are transmitted automatically, but if there is an error, it can be done by manual input. He receives urgent alerts if a harmful hypoglycaemic event is detected. He also receives alerts if a deterioration of his condition is detected over time.
In one of the follow-up appointments with the nurse, she alerted him to pay attention to the symptoms of hypoglycaemia and what to do if it happens. If the app receives a hypoglycemic value there is an alert to the nurse.
If the healthcare professional understands that the normal values of a specific patient are different from those originally defined as general targets, he can adapt and define personal targets for the patient directly in the system. This is a step towards a personalised medicine and can eventually diminuish a possible alert fatigue from the system.
Centralization of healthcare data
In diabetes, the body cannot regulate glycaemia values, so a close monitorization is need to understand if the patient needs changes in lifestyle and medication; also to prevent adverse situations
Person-centred healthcare
Alerts concerning hypo and hyperglycemic outliers to patient and nurse
Data to Process:
age
gender,
Diagnoses and date
smoking status
Area of residence
Biometric Data
Alergies
History of glycaemic measurements and timestamp
Globalcare- Prescription of Atitudes
CDS *
Interact/Viewer
Interact/Viewer
Glucometer

A1.2


s general
d define
ient directly
towards a
and can
ssible alert
tem.
The app has a record of the meals and the time they were taken in a food diary. When António has doubts about what to eat, he can consult the healthy suggestions.
António usually takes 30 min daily walks, because he was told by his doctor how important it is to stay active and wears a "fitness bracelet" to monitor his physical activity. This bracelet also communicates with the app
At the end of the day, the app sends an alert for António to verify that he took all his medication properly. When he has doubts regarding which pill to take and at which time, he can also consult this information on the app.
António can consult his prescription, both related to drugs and attitudes, on the app. He also has daily summing ups of what he has to do, in order to facilitate his life. Additionally, can leave daily notes in the app. For example, he can say that in a certain day he had a birthday party and committed some nutritional excesses. That way, the nurse is aware of this if there are hyperglycaemic outliers in that day.
The doctor
prescriptio
or, if he cons
a new appo
son-centred
healthcare
In diabetes, nutrition has to be adequate and low in carbs, in order to keep glycaemic values controlled. The diary is useful if the patient is not used to a healthy nutrition.
Physical Activity stimulates the production of insulin and its transport towards the cells. The existence of a fitband might motivate the patient.
Diabetic patients take a big number of medications. A control for each would be overwhelming, but an alert at the end of the day could be useful to understand drug adherence. Possibility to consult when to take the medication and at what hours also brings value
Adequate lifestyle is a key aspect of diabetes management. Currently, there is advice on attitudes but these are not digitalised or prescribed in a structured way, like drugs. These features contribute to a tighter control and follow up of the patient.
When there is a certain percentage of physical exercise below expected in the last 3 days, the patient is alerted and advised to go for a walk
Alert at the end of the day to confirm all the medication was taken to patient
CHO and calories, timestamp of meals
number of steps and calories
type, doses and timestamp each drug is taken
FhP Nutrition App
Interact /Viewer
Interact/Viewer
Globalcare- Prescription of Atitudes
CDS *
Interact/Viewer
CDS *
Viewer/Interact (My Care Plan)
fitness bracelet


The doctor receives an alert when the prescription ends to issue a new one or, if he considers necessary, schedule a new appointment with the patient.
The doctor/nurse that follows António receive an alert if he did not go to his appointment or did not perform a certain exam/medical procedure around the time he was supposed to.
In the case of laboratory analysis, the doctor can in advance prepare the laboratory request and it becomes available in the app.
When the exam is performed it can be integrated by system and received in the app as well. If something is considered to be an urgent outlier, an alert is issued to the specialist to identify an action plan in advance: anticipate the appointment; prescribe SOS medication,etc.
His nurse should remotely monitor and regularly check if everything is okay with António, intervening when necessary.
There is a patient/nurse communication channel. The latter can send a message when he detects something wrong the blood glucose evolution pattern of António or when he receives an alert that António is not going to appointments when he is supposed to. António can also send a photo of suspicious wounds to be checked. If the nurse considers them suspicious, she can make a home visit to António.
Alert to the doctor that the last prescription is over
Tighter control for adequate management of the disease
Alerts associated with calendars for the scheduled appointments/exams to the patient; eventually the nurse if anything was skipped
Quick identification of risk
Alerts to the doctor associated with outlier values detected in the exams
HbA1c levels, creatinine, lipid profile
Avoid dislocations to the hospital
Antecipate acute
CDS *
Interact*
Viewer *
Globalcare
Interact/Viewer
Viewer *


Everytime medication is dispensed, that information is sent to the patient's app. That is actually how it can be perceived by the system that one prescription is over, allowing the alert to be sent to the doctor (as mentioned previsouly)
In this process, there is also an integration with the community pharmacy. It can anticipate the need for medication, sending refill alerts to the app so that goes to the pharmacy to get a new stock
When António goes to the pharmacy, his weight and blood pressure are also measured. This data is either introduced manually or automatically sent to the app, where it can be consulted by everyone. If an outlier his detected in blood pressure, for example, the pharmacist can sent a note to the doctor in Lusíadas. If the pharmacist understands the patient is not adhering to medication, for example, he can also send a note to the hospital so the healthcare professionals can think of a plan of action.
At least annually, the nurse makes sure that António fills questionnaires related to his well-being and how he perceives his health is, both mentally and physically, in order to understand whether the treatment his bringing value to his life.
Value-Based Healthcare
Contribution to medication adherence
Frequent control of blood pressure and weight. Communication with hospital if anything is wrong
Refill alert, to go to the pharmacy 5 days before the stock is over
BMI, blood pressure
Viewer/Interact
Sifarma
Sifarma
Viewer/Interact
CDS *
Sifarma Clinical Pathways
Sifarma
Viewer
weight scale
Sphygmomanometer

Annex 2

Creation of the Synthetic Dataset

First of all, to facilitate the creation of the dataset, the input variables will be a simpler version of what was proposed in Section 4.2.1.1. The goal was to only include the "indispensable" variables, defined here as variables that were considered underlying causes of a particular microvascular complication in at least two of the four studies from the state of the art, in Section 2.3.3. These complications are highlighted in Table 2.3 from that Section and include:

-HbA1c: studies 2 and 3 revealed high values of HbA1c are the only risk factor associated with all 3 complications.

-Hypertension: risk factor for retinopathy in studies 2, 3 and 4;

-BMI: risk factor for neuropathy in studies 2 and 3

-Gender: being a female is a risk factor for neuropathy in studies 2 and 4

-Years of Diagnosis: interestingly, even though it was considered an underlying cause for some complications, it was never the same complication between two different articles. However, when thinking of designing a dataset, it seems like an important variable to have, since the incidence of each complication increases over the course of the disease. By the time of diagnosis some patients might have some complications, but they can appear with time if it is poorly managed.

To build the dataset, the first step was to find the incidence of the distinct microvascular complications in the diabetic population according to different durations of the disease. Three similar time periods were defined: <5 years (which also makes sense because some patients already present complications at the time of diagnosis); 5-10 years; 10-15 years. More than 15 years was discarded both to facilitate the creation of the dataset and because, in principle, limiting this to 15 years is enough to gain insights into the underlying causes of the complications and taking action to prevent or attenuate them.

Prevalences

A difficulty was the fact that these prevalences seem to differ a lot among distinct countries. For example, in a study that counted with the participation of 38 countries [95], the prevalence ranged from 4.1% in South-East Asia to 18,8% in Europe for a duration of the disease of 4.1 years. Having understood this reality, the next step was to discover the specific prevalences in Portugal. Such information was only found in a study regarding retinopathy, that counted with the participation of 874 patients [96]. This study also validated lack of glycemic control and high blood pressure as risk factors for this complication The values were obtained the following way:

The study counted with 437 diabetic patients with retinopathy and 437 diabetic patients without it (control). It presented the following distribution through years of diagnosis in each group, presented in Table A2.1.

Table A2. 1 Percentage of patients with and without retinopathy in each years since diagnosis group- a Portuguese study [96]

	Without retinopathy (control)	With retinopathy
<5 years of diagnosis	42.8%	18.3%
5-10 years of diagnosis	30.4%	30.9%
10-15 years of diagnosis	18.5%	33.4%
15-20 years of diagnosis	8.2%	17.4%

As explained, over 15 years was discarded. But knowing that each group, control and with retinopathy, had 437 patients, it was possible to figure how many patients were exactly in each, which is presented in Table A2.2.

Table A2. 2 Number of patients with and without retinopathy in each years since diagnosis group- a Portuguese study [96]

	Without retinopathy (control)	With retinopathy
<5 years of diagnosis	187 patients	80 patients
5-10 years of diagnosis	133 patients	135 patients
10-15 years of diagnosis	81 patients	146 patients

With that information, it was possible to calculate the prevalence of retinopathy for each group of years since diagnosis. For example, for <5 years of diagnosis:

$$Prevalence\ of\ retinopathy\ for < 5\ years\ of\ diagnosis = \frac{80}{187+80} x100 = 30.0\% \quad \textbf{(1)}$$

The values of Table A2.3 were obtained:

Table A2. 3 Prevalences of retinopathy

	<5 years of diagnosis	**5-10 years of diagnosis**	**10-15 years of diagnosis**
Retinopathy	30.0%	50.4%	64.3%

For the prevalence of neuropathy, the results obtained by a german cohort study with 844 T2DM patients was considered [97]. The values were taken from the plot of Figure A2.1:

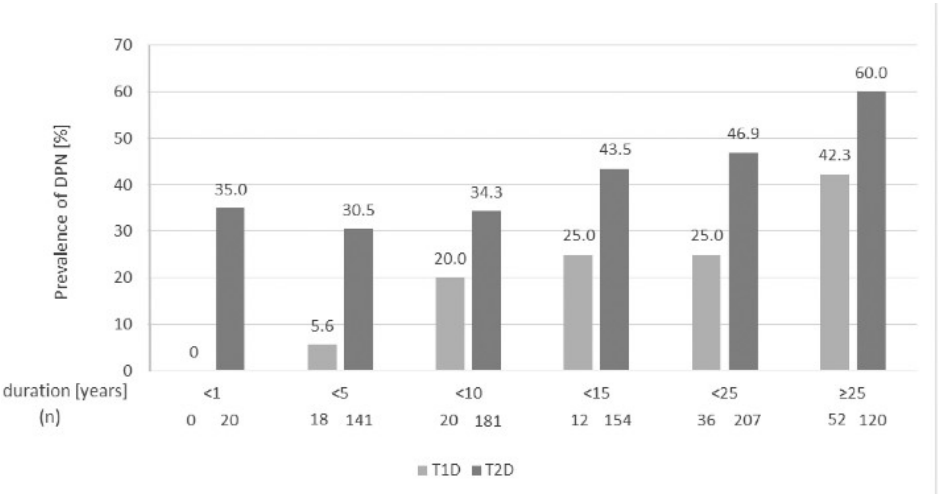


Figure A2. 1 Prevalence of Neuropathy in T1DM and T2DM patients [97]

For the <5 years period, the values were directly obtained, but for the 5-10 year period, one considered the values presented for <10 years, and the same for the 10-15 year period. The values are organized in Table A2.4.

Table A2. 4 Prevalences of Neuropathy [97]

	<5 years of diagnosis	**5-10 years of diagnosis**	**10-15 years of diagnosis**
Neuropathy	30.5%	34.3%	43.5%

Initially, the idea was to include nephropathy as well, but its prevalence is overall lower than the others. One feared that it would not have enough representativity for the algorithms to make valuable predictions. Additionally, working with two complications is easier than working with three, so the focus was on retino and neuropathy.

999 patients were created: 333 for each time period, divided among the different pathologies. Table A2.5 also presents the number of patients with each complication, considering the prevalences obtained from literature.

Table A2. 5 Prevalences of the Complications- How many patients they represent in the dataset

	<5 years of diagnosis (333 patients)	**5-10 years of diagnosis (333 patients)**	**10-15 years of diagnosis (333 patients)**
Retinopathy	30.0% (100 patients)	50.4% (168 patients)	64.3% (214 patients)
Neuropathy	30.5% (101 patients)	34.3% (114 patients)	43.5% (145 patients)

Some patients were attributed the two complications, but in a random way. The goal was to assure the total of each complication is the value in Table A2.5.

HbA1c levels

It was a common conclusion in the different studies from Table 2.3 in Chapter 2 that HbA1c levels are a risk factor for the development of all complications. However, one tried to make sure the values provided to the synthetic dataset were realistic. According to a report from *Observatório Nacional da Diabetes* from 2019 [98], in 2018, diabetic patients from the National Health Service had the following distribution of HbA1c values:

- <6.5%: 47.3% (on target)
- <7%: 67.8%
- >8%: 19,9%

From this data, one decided to divide the HbA1c levels category in 4 groups: <6.5% (on target); 6.5%-7%;7%-8% and >8%, that had the percentages presented in Table A2.6. Those percentages were then applied to a group of 333 patients, in order to know how many patients from each year of diagnosis group should be in each category of HbA1c levels.

Table A2. 6 Perecentages of the different HbA1c groups and how many patients they represent in the current dataset

	<6.5%	**6.5%-7%**	**7%-8%**	**>8%**
% of patients	47.3%	20.5% (67.8%-47.3%)	12.3% (100-47.3-19.9-20.5)	19.9%
% applied to a group of 333 patients (each year of diagnosis group)	158	68	41	66

The distribution of HbA1c in the synthetic dataset is in full accordance with the table. To build it, one started by distributing the highest HbA1c levels in the patients presenting two complications and then moved to the patients with only one complication and so on. The idea was to be in accordance with the fact that uncontrolled glycemia leads to the development of complications, whereas controlled patients can achieve better health outcomes. The distribution of HbA1c in the synthetic dataset is in full accordance with the table.

Gender

Another input variable to be considered is the gender. Being a female was defined to be an influence factor for neuropathy. To work with specific values, one took as basis one of the studies from the state of the art [99] that defined that the odds ratio of being a female for neuropathy is of 2.07 [99]. Odds ratio represent the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring without that exposure [100]:

$$OR = \frac{\frac{nº of\ exposed\ cases}{nº\ of\ unexposed\ cases}}{\frac{nº\ of\ exposed\ non-cases}{nº of\ unexposed\ non-cases}} \quad \textbf{(2)}$$

OR>1 mean that the exposure is associated with the outcome.

To build the dataset, one worked again with each group of years since diagnosis individually, each with 333 patients. There was an assumption there would be 166 women and 167 men in each, to keep it balanced. The formula for the odds ratio became:

- <5 years of diagnosis (101 patients with neuropathy)

$$OR = \frac{\frac{nº of\ females\ with\ neuropathy}{nº\ of\ males\ with\ neuropathy}}{\frac{nº\ of\ females\ that\ do\ not\ have\ neuropathy}{nº of\ males\ without\ neuropathy}} = \frac{\frac{x}{101-x}}{\frac{166-1}{167-1}} = 2.07 \qquad \textbf{(3)}$$

The goal was to find the value of x, which was the number of females with neuropathy in that group. After manual calculation, a value of x=68 was obtained. This means that, from the 101 patients with neuropathy, 68 are female and 33 are male. The gender was distributed according to those values. For the other patients, without neuropathy, the gender was then attributed to guarantee than in the end 166 were female and 167 were male.

The same logic was applied to the other groups of years since diagnosis and the following values were obtained and are presented in Table A2.7.

Table A2. 7 Distribution of Gender in the Neuropathic Population

	<5 years of diagnosis	**5-10 years of diagnosis**	**10-15 years of diagnosis**
Females (with neuropathy)	68	77	89
Males (with neuropathy)	33	37	45
Total of patients with neuropathy	101	114	134

BMI

According to *Observatório Nacional da Diabetes* from year 2016 [101], in 2015, 49.2% of diabetic patients were overweight (BMI between 25 and 30), 39.6% were obese (BMI above 30), which means that only 11.2% were “normal”.

Again, those percentages were applied to each group of 333 patients and are presented in Table A2.8.

Table A2. 8 Number of patients in each BMI group

	Normal	**Overweight**	**Obese**
Nº of patients from year of diagnosis group (333 patients)	38 patients	163 patients	132 patients

For this distribution of values, it is important to keep present that BMI was defined as a risk factor for neuropathy. Therefore, the distribution of obesity started with the patients that presented that complication.

Hypertension: To facilitate the creation of the dataset, and as observed in the examples from the state of the art, hypertension was divided in two categories: yes (the patient has a clinical

history of hypertension) or no. Since hypertension was defined as a risk factor for retinopathy, all the patients that presented this complication were given a yes, and the others given a no.

Considerations on the Creation of the New Dataset

As explained in section 4.1.2.3, a new dataset was later created to make considerations on overfitting. For that, 450 patients were constructed, 150 for each year of diagnosis group. For the rest of the variables, they were all built the same way, in accordance with all the percentages presented previously. Considerations on the distribution of patients for prevalence of diasese, HbA1c levels, Gender and BMI are presented in Tables A2.9, A2.10, A2.11 and A2.12, respectively.

Prevalences

Table A2. 9 Prevalences in the New Dataset

	<5 years of diagnosis (150 patients)	**5-10 years of diagnosis (150 patients)**	**10-15 years of diagnosis (150 patients)**
Retinopathy	30.0% (45 patients)	50.4% (76 patients)	64.3% (96 patients)
Neuropathy	30.5% (46 patients)	34.3% (51 patients)	43.5% (65 patients)

HbA1c levels

Table A2. 10 HbA1c levels in the new dataset

	<6.5%	**6.5%-7%**	**7%-8%**	**>8%**
% of patients	47.3%	20.5% (67.8%-47.3%)	12.3% (100-47.3-19.9-20.5)	19.9%
% applied to a group of 150 patients (each year of diagnosis group)	71	31	18	30

Gender

Table A2. 11 Gender in the New Dataset

	<5 years of diagnosis	**5-10 years of diagnosis**	**10-15 years of diagnosis**
Females (with neuropathy)	31	34	44
Males (with neuropathy)	15	17	21
Total of patients with neuropathy	46	51	65

BMI

Table A2. 12 BMI in the new dataset

	Normal	Overweight	Obese
Nº of patients from year of diagnosis group (150 patients)	17 patients	74 patients	59 patients

Considerations on the Addition of Noise

Again, with the purpose of avoiding overfitting, noise was added to the original dataset, first 5% noise and then 10% noise.

5% of 999 patients represents approximately 50 patients. 999 divided by 50 is approximately 20. So, starting from patient 1 and to every other 20 patients, the value of classification (the complication) was altered to one that did not make sense according to the other input variables. If a patient had hypertension, was obese and had a high level of HbA1c, he had probably both retinopathy and neuropathy. Adding noise meant changing that classification to having no complications, for example.

In the case of 10%, adding noise meant doing that to every other 10 patients.

As a final note, ML algorithms require all input and output variables to be numeric. Therefore, all values from the input variables were mapped to integers. The mapping is described in Table A2.13:

Table A2. 13 Mapping between Input Variables Values and Integers

Input Variable	Mapping
Years since Diagnosis	<5 year = 0 ; 5-10 years = 1; 10-15 years = 2
HbA1c Levels	<6.5% = 0; 6.5%-7% = 1; 7%-8% = 2; >8% = 2
Hypertension	no = 0 ; yes = 1
Gender	Female = 0; Male = 1
BMI	normal = 0; overweight = 1; obese = 2
Complication	no complication = 0; neuropathy = 1; retinopathy = 2; neuropathy + retinopathy = 3

Annex 3

Results from the Proof of Concept

1) SVM

1.1) SVM: zero noise

-New Data

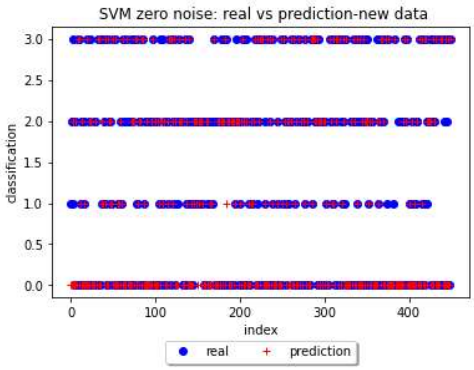


Figure A3. 1 Overlap of real and predicted labels-new data (without noise)

1.2) SVM: 5% noise

-Original Dataset

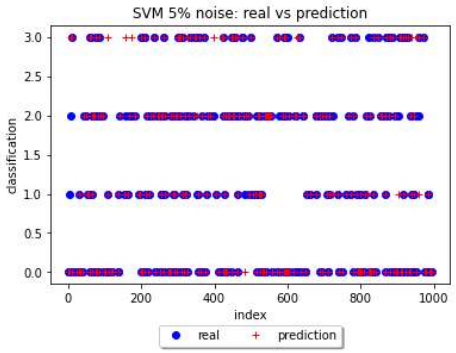


Figure A3. 2 SVM: Overlap of real and predicted labels (5% noise)

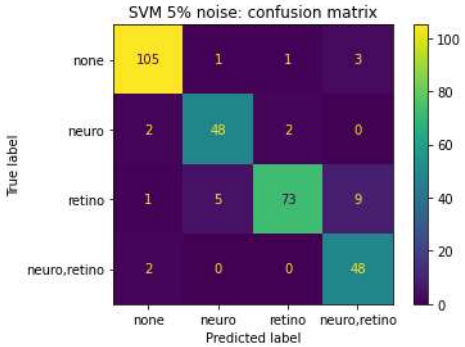


Figure A3. 3 SVM: Confusion Matrix (5% noise)

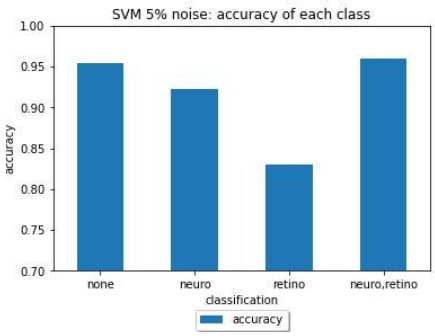


Figure A3. 4 SVM: Accuracy for each class (5% noise)

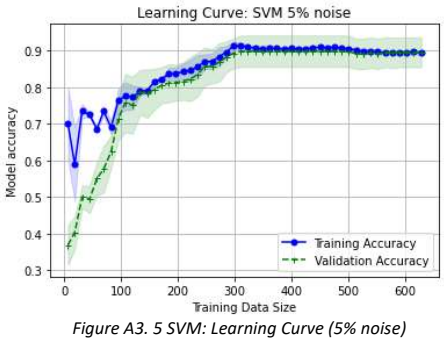


Figure A3. 5 SVM: Learning Curve (5% noise)

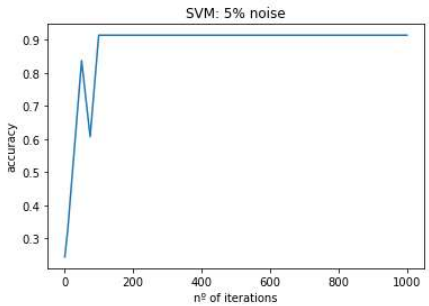


Figure A3. 6 Nº of iterations vs Accuracy (5% noise)

-New Data

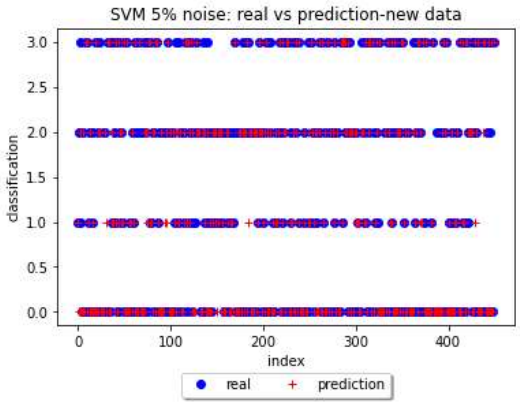


Figure A3. 7 SVM: Overlap of real and predicted labels-new data (5% noise)

1.3) SVM: 10% noise

-Original Data

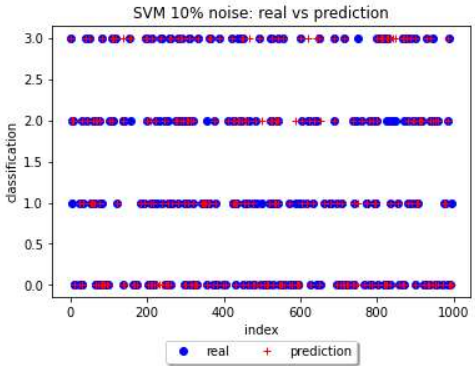


Figure A3. 8 SVM: Overlap of real and predicted labels (10% noise)

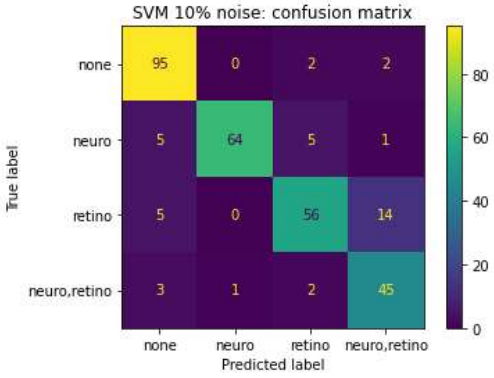


Figure A3. 9 SVM: Confusion Matrix (10% noise)

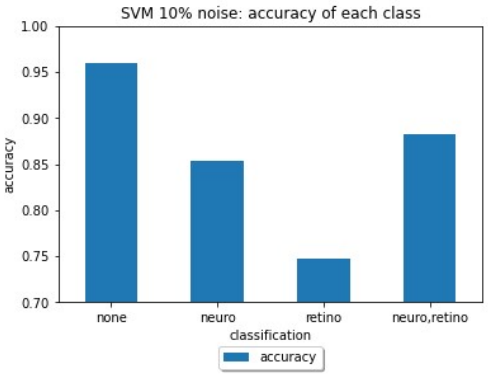


Figure A3. 10 SVM: Accuracy for each class

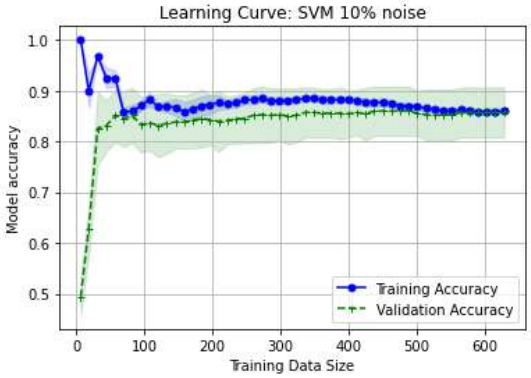


Figure A3. 11 SVM: Learning Curve (10% noise)

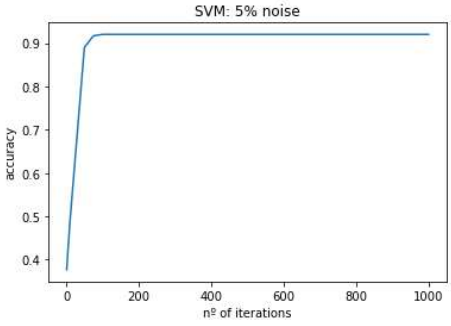


Figure A3. 12 Nº of iterations vs Accuracy (10% noise)

-New Data

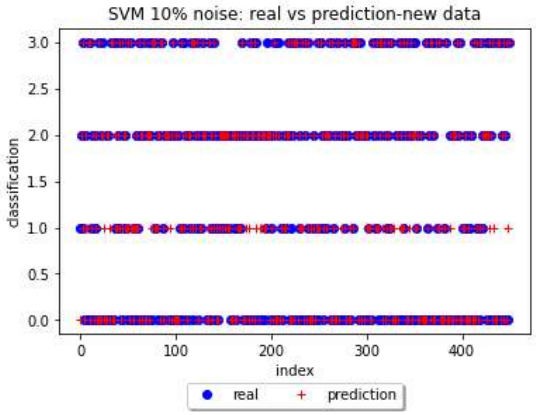


Figure A3. 13 SVM: Overlap of real and predicted labels-new data (10% noise)

2) RF

2.1) RF Zero Noise

-Original Data

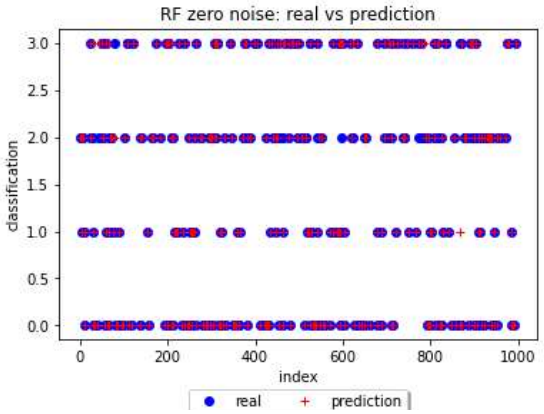


Figure A3. 14 RF: Overlap of real and predicted labels (without noise)

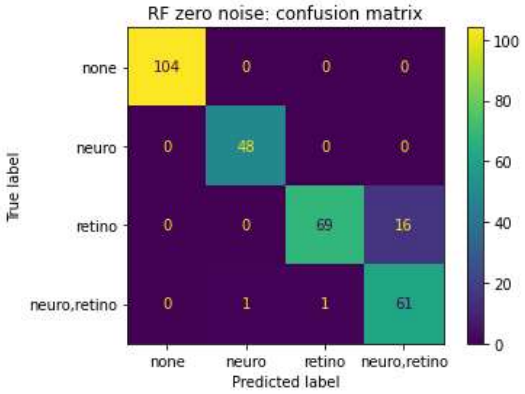


Figure A3. 15 RF: Confusion Matrix (without noise)

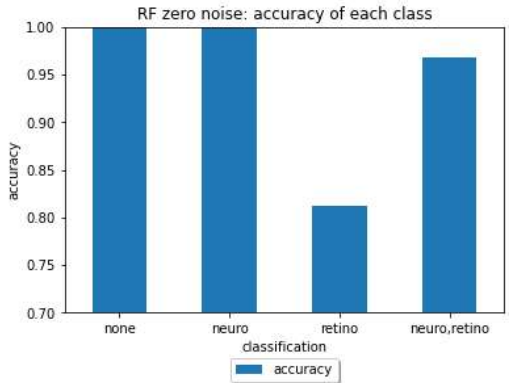


Figure A3. 16 RF: Accuracy of each class (without noise)

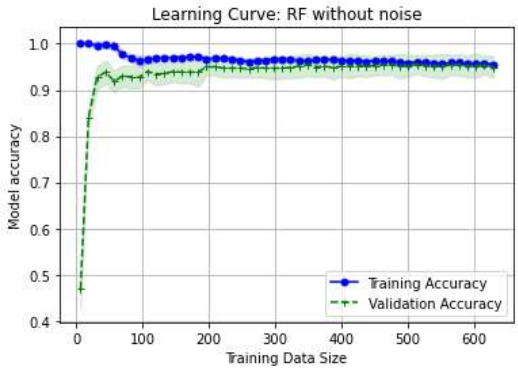


Figure A3. 17 RF: Accuracy of each class (without noise)

-New Data

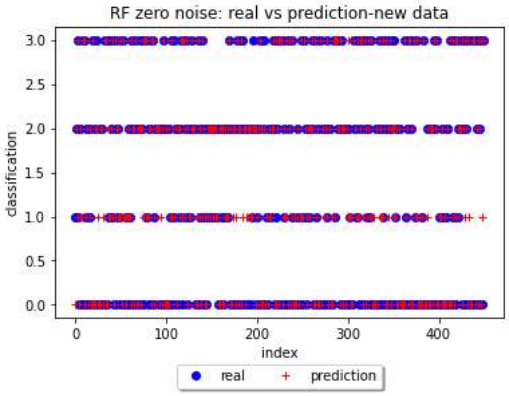


Figure A3. 18 RF: Overlap of real and predicted labels-new data (without noise)

2.2) RF 5% noise

-Original Dataset

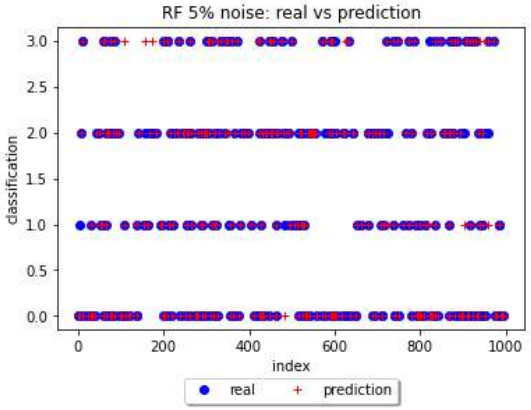


Figure A3. 19 RF: Overlap of real and predicted labels (5% noise)

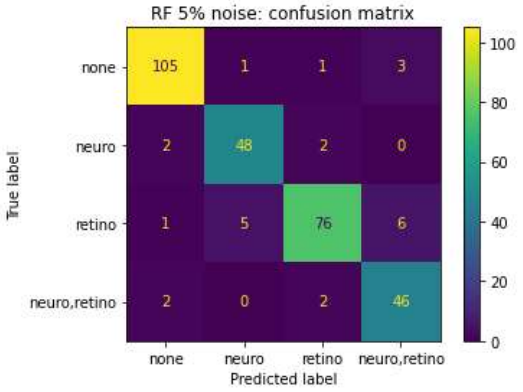


Figure A3. 20 RF: Confusion Matrix (5% noise)

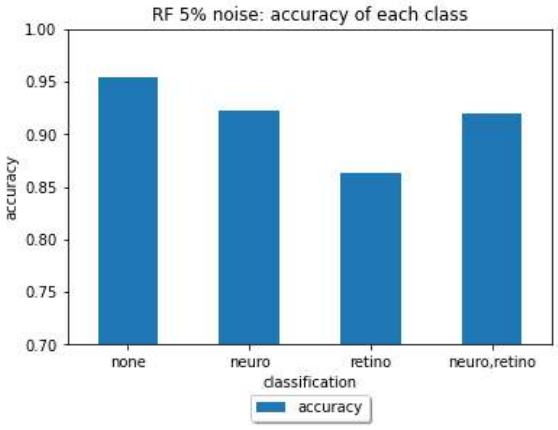


Figure A3. 21 RF: Accuracy of each class (5% noise)

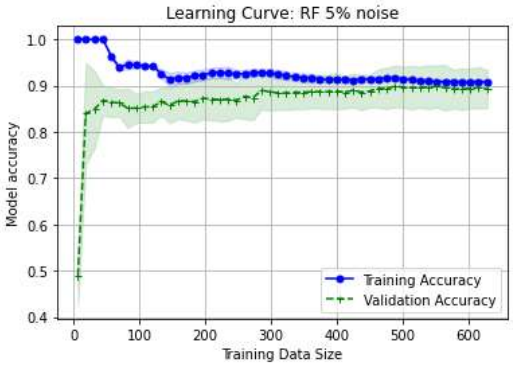


Figure A3. 22 RF: Learning Curve (5% noise)

-New Data

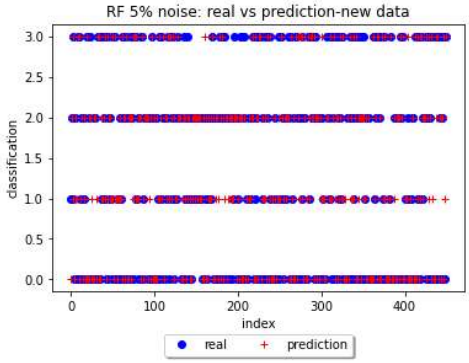


Figure A3. 23 RF: Overlap of real and predicted labels-new data (5% noise)

2.3) RF 10% noise

-Original Dataset

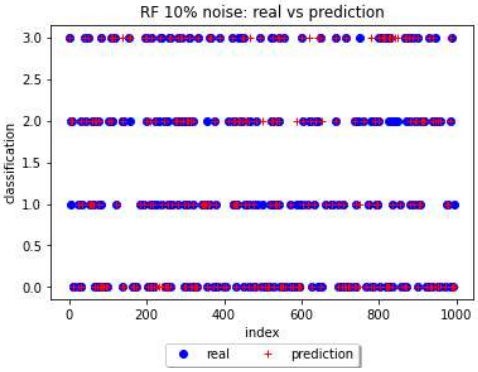


Figure A3. 24 RF: Overlap of real and predicted labels (10% noise)

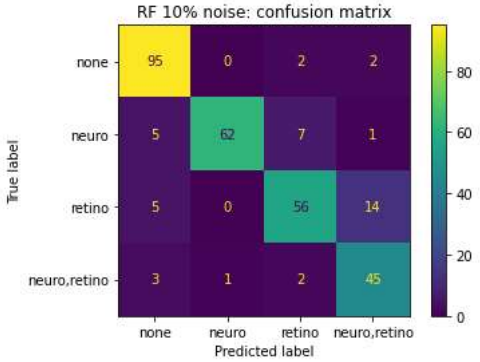


Figure A3. 25 RF: Confusion Matrix (10% noise)

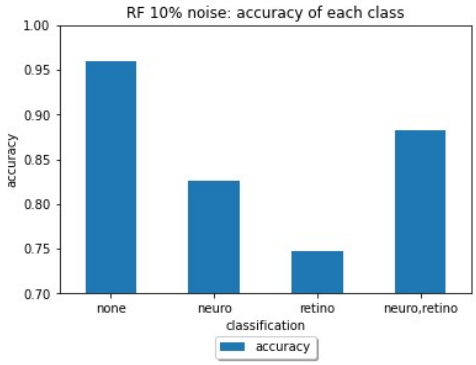


Figure A3. 26 RF: Accuracy of each class (10% noise)

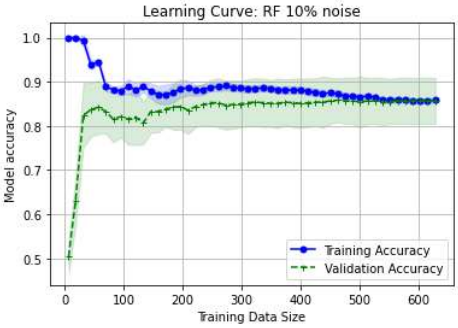


Figure A3. 27 RF: Learning Curve (10% noise)

-New Data

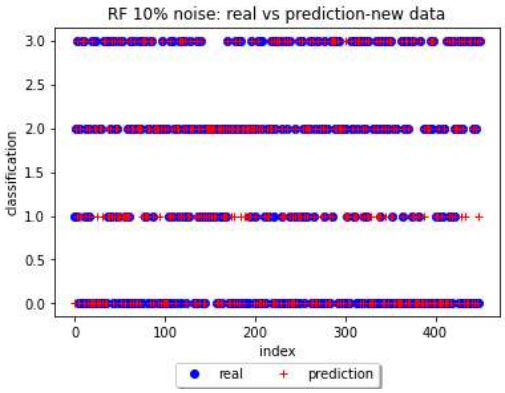


Figure A3. 28 RF: Overlap of real and predicted labels-new data (10% noise)