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
ana.rita.fontes@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

October, 2021

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 project. 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 types 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

1.Introduction

1.1 Context and Motivation

Developed countries have an ageing population, with an increasing burden of chronic diseases. Traditional healthcare systems are not handling it in the most efficient way, since they were mostly built to deal with acute symptoms. Chronic diseases require a long-term management that aims to first stabilize the patient's condition and then prevent long term complications [1]. That is in accordance with two emerging approaches: patient-centered care and value-based healthcare. The former ensures all health-related decisions are valued by the patient and respect their personal needs and preferences [2]. The latter aims to change the way hospitals are reimbursed: instead of being rewarded for the number of services they provide, the focus should be on maximizing the value brought to the patient [3].

The ConnectedHealth Project is a project by Glintt and some key partners and its overall goal is to create a set of interoperable and personalized tools to improve the way healthcare is delivered, towards a more preventive medicine. In specific, the subproject relevant for this thesis aims to create a health information system focused on the individual and proposes the creation of a Clinical Data Repository (CDR), a structured way to organize the health data collected, and of a Clinical Decision Support System (CDSS), a system to enhance decisions and actions related to health. The project also has the overall goals of converting the information collected into knowledge, in order to provide new insights to the medical community (data mining). Additionally, it aims to: transform Portugal in an international reference for value-based healthcare and give the community pharmacy a bigger role in the care for the patients. The last step of the project is to validate the solutions proposed in in real-life contexts, with 6 study case pilots planned for exploration. This work is focused in *Study Case 1: Providing Care at Home for Patients with common chronic diseases: Diabetes Mellitus type 2 (T2DM).*

1.2. T2DM

Diabetes Mellitus is a chronic disease characterized either by lack of enough insulin (hormone that regulates glucose) production by the pancreas or by lack of efficient use of the insulin produced by the body [4]. If poorly regulated, it leads to an increase in glucose levels, a condition known as hyperglycemia. Statistically, the World Health Organization (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 [4]. There are two main types of diabetes: type 1 is an autoimmune disease with an early onset in life; type 2, on the other hand, usually appears at some point in adulthood, associated with lifestyle factors such as overweight and lack of physical activity.

Short and Long-Term Complications

The main goal of diabetes management is to control the most life-threatening outcomes of the disease: both short-term and long-term complications.

Short-term complications encompass mostly hypoglycemia, that can be lethal [5]. Long-term complications occur because over time high blood glucose can damage both small (microvascular) and large (macrovascular) vessels [6]. Microvascular complications can affect eyes (retinopathy), kidneys (nephropathy) and nerves (neuropathy).

Lifestyle Management

To control the disease in the best way possible, it is important to maintain an adequate lifestyle. Relevant aspects to manage include: food intake [7] (an adequate and healthy diet with low caloric and carbohydrates (CHO) intake has the potential to keep an adequate metabolic function and prevent further complications); physical exercise [8] (aerobic activity can improve blood glucose control by stimulating insulin production and its transport to the cells); medication management [9] (diabetic patients take anti-hyperglycemic medication to help maintain blood glucose levels as normal as possible). Additionally, patients should control their glycemia tightly. There are two approaches for blood glucose measurement in daily life: self-blood glucose monitoring (SBGM), a fingerstick to measure glycemia at certain times of the day [10] and continuous glucose monitoring (CGM), that includes wearing a small device underneath the skin, that measures glycemia continuously throughout day and night [11].

1.3 Clinical Decision Support Systems

CDSS aim to improve the way healthcare is delivered by using the adequate available clinical knowledge and patient’s information at the right time, enhancing health-related decisions and actions [12]. More specifically, CDSS can be useful for [12]: Patient’s Safety (for example, by detecting harmful drug interactions in prescriptions or something that the patient is allergic to); Clinical Management (follow up and treatment reminders, making sure there is an adherence to clinical guidelines); Diagnostic Support (providing diagnostic suggestions based on patient’s data or even by interpreting medical images and laboratory results), among other aspects.

CDSS can be subclassified into knowledge versus non-knowledge based systems [12][13]. 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 and clinical evidence. Non-knowledge-based systems also require a data source, but they apply artificial intelligence algorithms 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 clinicians have problems understanding the logic they use to produce recommendations, hence why they are not so widespread, despite often obtaining good results [14].

1.4 Objectives

This work is one of the kickstarts of the ConnectedHealth project, aiming to propose concrete and feasible solutions that take into account its overall goals, with a special focus on the CDSS that subproject 1 intends to incorporate. With the use case of T2DM, the concrete objectives of the work became: i) to understand how Clinical Decision Support (CDS) is improving health outcomes of these patients nowadays; ii) to propose a solution for its application in the context of this work.

2. State of the Art

This Section aims to meet the first objective of this work. It explores how CDSS are currently being used to improve the outcomes of diabetic patients and the current state of its interoperability.

2.1 Interoperability Considerations on CDSS

From their origin until recent days, CDSS can be divided into four main architectures, each one with an increased level of interoperability [15]. The focus here is on the most recent architecture: **Service Models.** Service Models aim to separate the clinical decision support and the information systems, recombining them through application programming interfaces (APIs). In order to fully accomplish interoperability, 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.

FHIR is a standard for healthcare data exchange. Overall, it allows to define healthcare business objects, relate them, implement them in a computable form and share them across well-defined interfaces [16]. For CDS in specific, the CDS Hooks API [17] specification was developed and describes how a client, usually an 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.

2.2 Current Applications of CDSS for Diabetes

2.2.1 Clinical Management of Diabetes

In this section, one will explore how CDS solutions are improving the way diabetes is managed clinically, in terms of efforts from the healthcare teams to maintain patient's health over the time.

2.2.2.1 Knowledge-Based CDS solutions

Overall, the majority of projects found in literature [18][19] regarding this aspect were CDS efforts to ensure that the following variables were being taken/assessed with the necessary frequency, according to certain 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.

2.2.2.2 Non-knowledge Based Solutions

Regarding current applications of data mining for diabetes, mostly through Machine Learning (ML) algorithms, examples can be found in literature for [20]: Disease prediction and diagnosis; Interpretation and prediction of blood glucose levels; Prediction of diabetes associated complications.

In the context of this work, prediction of microvascular complications is an interesting approach to explore, since some of its underlying causes are still not fully understood and ML algorithms can detect patterns. Additionally, a prediction of a complication based on the patient's current health status would allow reallocation of efforts for its prevention, bringing value to the patient. Some studies found in literature already tried to understand the underlying causes of retinopathy, neuropathy and nephropathy and the results are summarized in Table 1:

Table 1: Summary of studies from literature on underlying causes of microvascular complications

[21]	Influence of duration of diagnosis, smoking and hypertension on the overall appearance of complications.
[22]	HbA1c associated with all 3 complications; Early age at onset can mean a stronger form of the disease due to genetic factors -Retinopathy influenced by: **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**
[23]	-**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**

[24]	The following risk factors were identified for each microvascular complication: -Retinopathy: female and **hypertension** -Neuropathy: **female** and **BMI** -Nephropathy: Duration of Diabetes

One of the studies [23] also concluded that Support Vector Machines (SVMs) and Random Forests (RFs), two ML algorithms, achieve the best results.

2.2.2 Daily-Life Management of Diabetes

To control the disease in the best way possible, it is important to maintain an adequate lifestyle. With the digitalization of the population, apps and other technologies can provide some help in the management of important aspects of the disease. Some examples of literature also combine the data they collect with tools that make quality decisions for the patient (decision support systems), with the ultimate goal of enhancing therapeutic outcomes. A good example is METABO [25][26][27].

METABO has a professional and a patient end-user profile. The Patient Monitoring Device allows to check the prescribed treatment, in terms of medication, diet, physical activity, frequency that glucose should be measured and even goals and education that should be achieved. It also collects data related to: food (manual input of carbohydrate intake per meal), medication intake, physical activity (using a pedometer) and glycemic values (either discrete values using a fingerstick or with CGM). 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 identified as knowledge-based CDS. Additionally, it provides a non-knowledge based component: The (individualized) Metabolic Modelling System models patient's metabolism in order to predict hypo/hyperglycemic events and provides decision support in the form of short and long term alerts to the patient, taking as input the insulin and food intake, glycemic data, physical activity and other clinical data. It can provide immediate critical feedback or suggestions for modifications of diet, activity or medication in order to avoid future critical events.

3. Methodology

The methodology followed throughout this work can be inserted in the spectrum of *Action Research*. In Action Research, the researcher (Glintt and the other companies working for ConnectedHealth) and the client (diabetic patients and the healthcare team that follows them, in this case a Lusíadas Group team of 2 nurses, 1 endocrinologist and 1 physician from internal medicine) work together to "diagnose" a problem and develop a solution for it. In this case, that encompasses what aspects of diabetes management can be improved and propose solutions to do it. Action Research is a cycle of the following steps: planning, acting, observing and reflecting.

The first step of planning consisted in the research presented and was common for both the knowledge and non-knowledge-based solutions that will be proposed in Chapter 4. However, after that first step, the cycles diverged. For the non-knowledge solution, the rest of the cycle consisted in proposing input variables and algorithms for prediction, as well as doing a "proof of concept", to show how data can be handled. For the knowledge-based solution, the rest of the first cycle consisted in building an use case describing how the different intervenients would interact and benefit from the system, which ended in its validation by the clinical team. The second cycle consisted mostly in building the functional requirements and their internal validation.

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

4.1 Use Case: Presentation and Discussion

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. Before explaining it, note that Clinical Pathways are implementations of clinical protocols and guidelines specific to each care setting. They ensure that the doctor follows all the necessary steps.

The use case is described next:

1-António goes to the pharmacy and does a routine risk assessment for diabetes. He obtains a high score and is referenced to a specialist doctor in Lusíadas.
2-He goes to a first medical appointment. The doctor follows a clinical pathway to ensure all the necessary steps are followed until arriving to a diagnosis. After

the confirmation of diagnosis, a clinical pathway is also useful to guide the doctor in making all the necessary prescriptions and referencing to other specialists, including a specialist nurse that is a central piece in the follow-up of the patient.
3-The prescription has a medication component and a non-medication one, related to attitudes: how many glycemic measurements to take daily, prescription of physical activity and nutrition. The nurse can also participate in this part of prescription.
4-The doctor advises António to start using an app to have better control over his health. António installs the app and allows all health parties involved (hospital and pharmacy) to have access to his data.
5-In the app, there is a "My Care Plan" section, where António can consult his prescriptions (medication and non-medication): he has an organized plan of all the activities he has to take daily.
6-António has a glucometer to perform SBGM in the frequency defined by the doctor. The values are directly sent to the app via APIs. He receives urgent alerts if a critical hypo or hyperglycemia is detected. The system can also detect deteriorations over time and notifies patient and nurse.
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.
8-The app has suggestions of healthy meals adequate for the disease and António is advised to follow them.
9-He also wears a fitness bracelet, that monitors his physical activity and heart rate. It sends alerts if it detects that the patient is falling behind in what was prescribed.
10-To confirm he is taking all his medication properly, he has to put a check in the app at the end of the day. Otherwise, another alert is issued.
11-The doctor receives an alert when the medical prescription ends to issue a new one or, if he considers necessary, schedule a new appointment with the patient.
12-António is reminded in advance of the appointments and exams he has to attend, which, in theory, 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.
13-In the case of laboratory analysis, the doctor can, in advance, prepare the laboratory request and it becomes available in the app.
14-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.
15-There is close proximity between nurse and patient. The nurse can consult regularly the values measured by the patient and how they are evolving. Additionally, there is a communication channel between them.
16-In this process, there is an integration with the community pharmacy. It can anticipate the need for medication and send refill alerts to the app. Every time medication is dispensed, that information is sent to the app. That is how the system is aware a prescription is over and sends an alert to the doctor (point 11).
17-When António goes to the pharmacy for refill, his blood pressure and weight are also measured and the values are sent to the app. In case of outliers, a clinical pathway is activated for new referencing to the doctor.
18- Every year, António takes a questionnaire in order to assess whether this solution is bringing value to his life.

Considerations on Clinical Management

The solution aimed to incorporate the aspects that had been perceived as important for clinical management during the research phase. Table 2 summarizes them, how they can be obtained and in which point of the use case they are mentioned:

Table 2: Summary of Clinical Management Aspects in the Proposed Use Case

Variables	How to obtain them	Point of the Use Case
Blood Pressure, BMI	Instruments from the Pharmacy	17
HbA1c, Kidney Values, Lipid Profile	Blood tests, prescribed by the specialist doctor	2 (when assessing diagnosis); 13 and 14
Foot Check	Podiatrist/Nurse (referenced by the specialist doctor)	2, 12
Eye Check	Ophtalmologist (referenced by the specialist doctor)	2,12

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 presence of neuropathy, retinopathy 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.

Note that the knowledge-based approaches from the state of the art (Section 2.2.2.1) defined that, for example, the patient had to go to two ophthalmology appointments per year and had a tight control over that, common for all patients. However, the approach here is not the same, mostly because the medical team pointed out that each patient is different. What the solution proposes is that the doctor schedules directly all the necessary appointments and that becomes associated with a notification system.

Considerations on Daily-Life Management

Regarding this, METABO was considered a "benchmark" for the solution. Table 3 summarizes the aspects that are contemplated in METABO and also integrated in the use case, with the indication of which point:

Table 3: Summary of how different lifestyle aspects are managed in METABO and in the proposed solution

	METABO	Proposed Use Case
Nutrition	Recommended daily calories and CHO intake per meal can be consulted.	8
Glycemic Values Control	SBGM and CGM data (not specified how it can be incorporated/ consulted).	6
Physical Activity	Duration, intensity and frequency collected by a pedometer and a metabolic holter.	9
Medication	Current administrations (ATC code of the drug); strength; time and dosage.	5;10
CDSS	-Alerts, reminders and motivation messages (knowledge-based) - Metabolic Model to forecast adverse glycemic events (non-knowledge based)	(specified in the following sections)

The Role of the Pharmacist

Overall, literature reports positive outcomes when the community pharmacist is involved in the care for the chronic patient [28]. One of the goals of ConnectedHealth was exactly to have a bigger integration of the community pharmacy in the journey of healthcare delivery. This was achieved in points 16 and 17.

4.2 Functional Requirements

After validating the use case, the next step was to define the functional requirements, describing what the different components of the system must do to make the solution described in the use case happen. Those include:

Glucometer

To fulfill point 6 of the use case, an option was found that can send data via Bluetooth [29]. However, the app has to consider manual input, in case there is an error or if the patient decides to use other options.

Fitness Bracelet

To fulfill point 9 of the use case, a good option was also found. It monitors steps, distances, calories and heart rate [30] and data is retrievable via APIs. However, manual input of data should, again, be possible.

Nutrition Control

A healthy and adequate nutrition, described in point 8, is important to maintain good health outcomes. However, some questions were raised on how to do it. Fraunhofer, one of the partners in the project proposed a nutrition app developed by them as a solution [31]. This app takes user information such eating habits; food preferences and restrictions and a professional makes 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.

The App

To bring everything that is proposed in the use case to life, the proposed app should include the following features:

-Measurements: to consult the data collected by the glucometer, fitness bracelet and measured in the pharmacy.

-Meal Plan: the simplest solution is to have a link directly to the nutrition app.

-Drug Prescription and Attitude Prescription (My Care Plan)

-Laboratory Requests and Laboratory Results

-Communication Channel

Some existing Glintt solutions can be incorporated here: The app can be an adaption from the Viewer [32]; Globalcare [33] can be used to orient the activity of the healthcare professionals in the hospital and Sifarma [34] in the pharmacy.

Clinical Data Repository

The CDR is a structured way to aggregate all the data necessary for the functioning of the app. The data should follow the FHIR specification, since it is the standard Viewer uses and the standard healthcare is transitioning to. Table 4 specifies the FHIR resources and the elements of the resources of the data necessary for each of the features predicted:

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

Macrofunctionality	FHIR Resource	Elements of the Resource
Measurements	Observations	Code; Value
Meal Plan	Nutrition Order (if a healthcare	foodPreferenceModifier;ex

	professional makes a recommendation of the meal plan)	cludeFoodModifier; NutritionOrder.oralDiettype;oralDiet.nutrition.amount
Drug Prescription	Medication Request	medication[x]; dosageInstruction
Prescription of Attitudes	Service Request	Intent; category; orderDetail; quantity[x]; occurrence[x]
Defining Next Appointments	Appointment	serviceCategory; appointmentType; start; participant.status
Laboratory Results	Diagnostic Report	Category; code; result

Clinical Decision Support System

The knowledge-based CDSS must receive information from the CDR via APIs and should be responsible for the rules and alerts summarized in Table 5:

Table 5: Triggers and Suggestions of Alerts Proposed for the Knowledge-Based CDSS

Trigger	Alert/Notification
Glycemia < x or Glycemia > x	Alert to the patient to take the adequate SOS plan; Notification to the nurse
Nº of hyperglycemias in the past 7 days > x	Alert to the patient that his condition is deteriorating and reminder of the importance of following his care plan; Notification to the nurse.
Nº of steps or steps in the last 3 days < what was defined by the professional	Motivation message to the patient
Medication check was not inserted	Reminder to the patient
Arrythmia detected through the fitness bracelet	Alert for the patient to go to the hospital
5 days prior to a scheduled event in the calendar (exam/appointment)	Notification to remind the patient
If the date in the calendar passed and the patient did not attend the event	Message to the patient, asking to reschedule; Notification to the nurse
Alert to refill 5 days before medication ends (Glintt's pharmacy software already does this via text/e-mail)	Alert to the patient
Pharmacy dispensed amount of medication that was in the prescription	Notification to the doctor (that first prescribed it) to prescribe a new one or schedule another appointment
Outliers of laboratory results	Alert to the doctor that prescribed it and the nurse (might want to initiate a SOS plan)

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. These alerts were overall approved by the healthcare professionals, who will be responsible for providing those initial values.

By the time this work was finished, the architecture for the CDSS and its integration with the other components was still under development, together with the CDR. However, it was left clear that it will be a service model in terms of architecture, using the FHIR CDS Hooks specification, since the goal is to achieve as much interoperability as possible across different systems.

5. Non-knowledge Based CDSS: 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. That can be achieved by a second CDSS that receives data from the CDR, but this time, instead of predefined rules, performs ML. Some creativity was allowed in the context of this work, to explore and propose ways to do reach this objective. After the research presented in the state of the art, it was considered interesting to propose data mining for the **prediction of microvascular complications.**

5.1 Input Variables

The basic goal of such a ML algorithm is to take a certain set of input variables and, from there, make predictions on the appearance of microvascular complications. Table 6 summarizes the variables proposed for this work, considering the conclusions of Table 1 from Chapter 2.

Table 6: 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 measure)	<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)

The algorithms should, from those variables, predict if the patient will have no complications, or, if he will, inform which ones.

Algorithms

In one of the studies from Table 1, SVMs and RFs had the highest accuracy, thus their proposal for this work:

-Support Vector Machines [35]: All the example cases are put in a N-dimensional space (N being the number of features) and it draws a hyperplane that maximizes the distance between data points of the different classes, called margin. 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). SVMs 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 and sigmoid.

-Random Forests [36]: Combine several Decision Trees, that make a classification individually. In the end, the class with more votes is the final prediction of the model. The decision trees for the random forests are created from the dataset through techniques called bagging and feature randomness, that basically add randomness to the number of examples in the dataset used for learning and the number of features considered to create trees, to guarantee that they are as uncorrelated as possible.

Proof of Concept

The goal of this section is to illustrate how the data that is planned to be obtained in the future when the solution is implemented can be handled. In order to do so, a synthetic dataset was created, in accordance with information from literature, to make sure it was accurate. To facilitate its creation, only neuro and retinopathy were included (nephropathy was left out) and only HbA1c levels, duration of diagnosis, BMI, Hypertension and Gender were included as input variables. Afterwards, a library from Python directed to ML, scikit-learn, that provides both SVM and RF, was used and both were tested on the data.

To study the performance of the algorithms, some metrics were used [37]: accuracy (ratio between the number of correct predictions and the total number of predictions) and the F1-score (measures the combination of precision: what proportion of positive predictions was actually correct and recall: what proportion of actual positives was identified correctly on the prediction task). The algorithms trained with 70% of the data and made predictions on the other 30% and those metrics described were used to assess the performance of that prediction.

Both SVM and RF have their own parameters that should be optimized. In SVM [37] those are the kernel, C (tradeoff between the accuracy of the classification of training examples versus the maximization of the decision's function margin) and gamma (how far the influence of a single training example reaches). For RF [39], one is dealing with the maximum depth of the tree and with the maximum number of features at each split. GridSearchCV [39], a function provided by scikit-learn, was used to search for the best combination of parameters for each.

When performing ML, one should be careful to prevent 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: it is not able to generalize [41]. Some techniques were used to prevent it and/or analyze if it was occurring: a new dataset was created to test the algorithm against it and, later on, add was noised to the dataset to see how the performance varied. Table 7 presents the overall results of the proof of concept.

Table 7: Results of the Proof of Concept for SVMs and RFs

		0% noise	5% noise	10% noise
	Best parameters (C; gamma; kernel)	0.03;0.15; polynomial	0.07;0.11; polynomial	10;0.15; rbf

SVM	Accuracy (30% testing set)	0.94	0.91	0.87
	F1-score (30% testing set)	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
RF	Best parameters (max_depth ;max_features)	4; 2	7; 2	6; 2
	Accuracy (30% testing set)	0.93	0.92	0.87
	F1-score (30% testing set)	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

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. That effect is visible when 10% of noise is added: the performance of the algorithm is equally good (or even better, for RFs) in the never before seen data. Between the algorithms, due to that aspect, one would say RFs are performing better. However, 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 show what can be done in real life context., If real data is obtained one day, both algorithms should be tested again and a similar analysis should be made.

6. Conclusions

Living in the era of digitalization and big data, new opportunities are arising to use it in benefit of society. CDSS are a way to do this: they enhance health-related actions and decisions either through detection and warning of harmful events or through ML algorithms that make valuable predictions.

The objectives of this work were first to understand how CDS are currently being used for T2DM and then propose how to apply it in the context of the ConnectedHealth project. The first objective was accomplished in the State of the Art: there was an exploration of how CDS, both knowledge and non-knowledge based, are 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 a proposal of how the entire solution surrounding it could be put into practice, with specification of an use case and its functional requirements.

As mentioned, the biggest focus was to improve and bring value to the way this disease is managed nowadays. That passed through the suggestion of a system that is centered on the patient; that involves all healthcare professionals in the follow-up process; that takes prescriptions to a new level, with an attitude component that also contributes to a personalized care and, finally, that has concerns on whether the patient perceives the solution as useful, bringing value to his life.

Finally, it is important to mention that, even though they have a great potential, the issue with previous CDS solutions was their lack of integration with different systems. In the present proposal, one aims to use the latest architectures and standards, in order to achieve maximum interoperability.

References

[1] 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

[2] 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

[3] 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

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

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

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

[7] Nutrition Principles and Recommendations in Diabetes. (2004). In *Diabetes Care* (Vol. 27, Issue SUPPL. 1).

[8] 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

[9] National Health Service (2020). Understanding Medicine: Type 2. Retrieved from: https://www.nhs.uk/conditions/type-2-diabetes/understanding-medication/

[10] Benjamin, E. M. (2002). Self-Monitoring of Blood Glucose: The Basics. *Clinical Diabetes*, *20*(1). https://doi.org/10.2337/diaclin.20.1.45

[11] Diabetes UK. Continuous Glucose Monitoring. Retrieved from: https://www.diabetes.org.uk/guide-to-diabetes/managing-your-diabetes/testing/continuous-glucose-monitoring-cgm

[12] 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
[13] 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
[14] 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
[15] 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
[16] HL7 FHIR (2019). *FHIR Overview-Architects*. Retrieved from: https://www.hl7.org/fhir/overview-arch.html
[17] CDS Hooks (2018). *Overview*. Retrieved from: https://cds-hooks.org/
[18]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.
[19] 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
[20] 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
[21] 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
[22] 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
[23] 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
[24] 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
[25] 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
[26] 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
[27] 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
[28] 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
[29] 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
[30] Fibit(2021). Inspire 2. Retrieved from: https://www.fitbit.com/global/us/products/trackers/inspire2.
[31] Fraunhofer Portugal. Lifana. Retrieved from: https://www.aicos.fraunhofer.pt/en/our_work/projects/lifana.html
[32] Glintt (2021). Viewer. Retrieved from: https://viewer.glintt.com/
[33] Glintt (2016). Globalcare. Retrieved from: https://www.glintt.com/en/WHAT-WE-DO/Product-Services/SoftwareSolutions/Pages/GlobalCare.aspx
[34] Glintt (2016). Sifarma. Retrieved from: https://www.glintt.com/en/WHAT-WE-DO/Product-Services/SoftwareSolutions/Pages/Sifarma.aspx
[35] Christmann A., Steinwart I. (2008). *Support Vector Machines*. Springer
[36] Breiman, L. (2001). Random Forests - Springer. *Machine Learning*, *45*(1).
[37] Towards Data Science (2018). Accuracy, Precision, Recall or F1? Retrieved from: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9
[38] Scikit Learn. *RBF SVM Parameters*. Retrieved from: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html
[39] Scikit Learn. *Sklearn.ensemble.RandomForestClassifier*. Retrieved from: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[40] Towards Data Science (2020). GridSearchCV for Beginners. Retrieved from: https://towardsdatascience.com/gridsearchcv-for-beginners-db48a90114ee
[41] IBM (2021). *What is Overfitting?* Retrieved from: https://www.ibm.com/cloud/learn/overfitting