UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DOCTORAL THESIS

ASSESSMENT OF RISK SCORES FOR THE PREDICTION AND DETECTION OF TYPE 2 DIABETES MELLITUS IN CLINICAL SETTINGS

Author:
Antonio
MARTINEZ-MILLANA

Supervisors:
Juan F. MERINO TORRES
Vicente TRAVER SALCEDO

A thesis submitted in fulfillment of the requirements for the Doctor of Philosophy on Technologies for Health and Wellbeing

by the

Department of Electronic Engineering
Universitat Politècnica de València

July 19, 2017

Declaration of Authorship

I, Antonio MARTINEZ-MILLANA, declare that this thesis titled, "ASSESSMENT OF RISK SCORES FOR THE PREDICTION AND DETECTION OF TYPE 2 DIABETES MELLITUS IN CLINICAL SETTINGS" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research doctoral degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed and cited.
- Where I have quoted from the work of others, the source is always given as footnote or in the Bibliography.
- I have acknowledged all main sources of help and financial support.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Essentially, all models are wrong, but some are useful."

George E. P. Box

Universitat Politècnica de València

Abstract

Doctor of Philosophy on Technologies for Health and Wellbeing

ASSESSMENT OF RISK SCORES FOR THE PREDICTION AND DETECTION OF TYPE 2 DIABETES MELLITUS IN CLINICAL SETTINGS

by Antonio MARTINEZ-MILLANA

Health and sociological indicators confirm that life expectancy is increasing, and so, the years that patients have to live with chronic diseases and co-morbidities. Type 2 Diabetes is one of the most common chronic diseases, specially linked to overweight and ages over sixty. As a metabolic disease, Type 2 Diabetes affects multiple organs by causing damage in blood vessels and nervous system at micro and macro scale. Mortality of subjects with diabetes is three times higher than the mortality for subjects with other chronic diseases.

On the one hand, the management of diabetes is focused on the maintenance of the blood glucose levels under a threshold by the prescription of anti-diabetic drugs and a combination of healthy food habits and moderate physical activity. Recent studies have demonstrated the effectiveness of new strategies to delay and even prevent the onset of Type 2 Diabetes by a combination of active and healthy lifestyle on cohorts of mid to high risk subjects. On the other hand, prospective research has been driven on large groups of population to build risk scores which aim to obtain a rule for the classification of patients according to the odds for developing the disease. Currently there are more than two hundred models and risk scores for doing this, but a few have been properly evaluated in external groups and, to date, none of them has been tested on a population based study.

The research study presented in this doctoral thesis strives to use externally validated risk scores for the prediction and detection of Type 2 Diabetes on a population data base in Hospital La Fe (Valencia, Spain). The study hypothesis is that the integration of existing prediction and detection risk scores on Electronic Health Records increases the early-detection of high risk cases. To evaluate this hypothesis three studies on the clinical, user and technology dimensions have been driven to evaluate the extent to which the models and the hospital is ready to exploit such models to identify high risk groups and drive efficient preventive strategies. The findings presented in this thesis suggest that Electronic Health Records are not prepared to massively feed risk models. Some of the evaluated models have shown a good classification performance, which accompanied to the well-acceptance of web-based tools and the acceptable technical performance of the information and communication technology system, suggests that after some work these models can effectively drive a new paradigm of active screening for Type 2 Diabetes.

Universitat Politècnica de València

Resumen

Doctorado en Tecnologías para la Salud y el Bienestar

EVALUACIÓN DE MODELOS DE RIESGO PARA LA PREDICCIÓN Y DETECCIÓN DE LA DIABETES MELLITUS TIPO 2 EN ESCENARIOS CLÍNICOS

por Antonio MARTINEZ-MILLANA

Los indicadores de salud y sociológicos confirman que la esperanza de vida está aumentando, y por lo tanto, los años que los pacientes tienen que vivir con enfermedades crónicas y comorbilidades. Diabetes tipo 2 es una de las enfermedades crónicas más comunes, especialmente relacionadas con el sobrepeso y edades superiores a los sesenta años. Como enfermedad metabólica, la diabetes tipo 2 afecta a múltiples órganos causando daño en los vasos sanguíneos y el sistema nervioso a escala micro y macro. La mortalidad de sujetos con diabetes es tres veces mayor que la mortalidad de sujetos con otras enfermedades crónicas.

Por un lado, la estrategia de manejo se centra en el mantenimiento de los niveles de glucosa en sangre bajo un umbral mediante la prescripción de fármacos antidiabéticos y una combinación de hábitos alimentarios saludables y actividad física moderada. Estudios recientes han demostrado la eficacia de nuevas estrategias para retrasar e incluso prevenir la aparición de la diabetes tipo 2 mediante una combinación de estilo de vida activo y saludable en cohortes de sujetos de riesgo medio a alto. Por otro lado, la investigación prospectiva se ha dirigido a grupos de la población para construir modelos de riesgo que pretenden obtener una regla para la clasificación de las personas según las probabilidades de desarrollar la enfermedad. Actualmente hay más de doscientos modelos de riesgo para hacer esta identificación, no obstante la inmensa mayoría no han sido debidamente evaluados en grupos externos y, hasta la fecha, ninguno de ellos ha sido probado en un estudio poblacional.

El estudio de investigación presentado en esta tesis doctoral pretende utilizar modelos riesgo validados externamente para la predicción y detección de la Diabetes Tipo 2 en una base de datos poblacional del Hospital La Fe de Valencia (España). La hipótesis del estudio es que la integración de los modelos de riesgo de predicción y detección existentes la práctica clínica aumenta la detección temprana de casos de alto riesgo. Para evaluar esta hipótesis, se han realizado tres estudios sobre las dimensiones clínicas, del usuario y de la tecnología para evaluar hasta qué punto los modelos y el hospital están dispuestos a explotar dichos modelos para identificar grupos de alto riesgo y conducir estrategias preventivas eficaces. Los hallazgos presentados en esta tesis sugieren que los registros de salud electrónicos no están preparados para alimentar masivamente modelos de riesgo. Algunos de los modelos evaluados han demostrado un buen desempeño de clasificación, lo que acompañó a la buena aceptación de herramientas basadas en la web y el desempeño técnico aceptable del sistema de tecnología de información y comunicación, sugiere que después de algún trabajo estos modelos pueden conducir un nuevo paradigma de la detección activa de la Diabetes Tipo 2.

Universitat Politècnica de València

Resum

Doctorat en Tecnologíes per a la Salut i el Benestar

AVALUACIÓ DE MODELS DE RISC PER A LA PREDICCIÓ Y LA DETECCIÓ DE LA DIABETIS MELLITUS DE TIPUS 2 EN ESCENARIS CLÍNICS

per Antonio MARTINEZ-MILLANA

Els indicadors sociològics i de salut confirmen un augment en l'esperança de vida, i per tant, dels anys que les persones han de viure amb malalties cròniques i comorbiditats. la diabetis de tipus 2 és una de les malalties cròniques més comunes, especialment relacionades amb l'excés de pes i edats superiors als seixanta anys. Com a malaltia metabòlica, la diabetis de tipus 2 afecta múltiples òrgans causant dany als vasos sanguinis i el sistema nerviós a escala micro i macro. La mortalitat de subjectes amb diabetis és tres vegades superior a la mortalitat de subjectes amb altres malalties cròniques.

D'una banda, l'estratègia de maneig se centra en el manteniment dels nivells de glucosa en sang sota un llindar mitjançant la prescripció de fàrmacs antidiabètics i una combinació d'hàbits alimentaris saludables i activitat física moderada. Estudis recents han demostrat l'eficàcia de noves estratègies per a retardar i fins i tot prevenir l'aparició de la diabetis de tipus 2 mitjançant una combinació d'estil de vida actiu i saludable en cohorts de subjectes de risc mitjà a alt. D'altra banda, la investigació prospectiva s'ha dirigit a grups específics de la població per construir models de risc que pretenen obtenir una regla per a la classificació de les persones segons les probabilitats de desenvolupar la malaltia. Actualment hi ha més de dos-cents models de risc per fer aquesta identificació, però la immensa majoria no han estat degudament avaluats en grups externs i, fins ara, cap d'ells ha estat provat en un estudi poblacional.

L'estudi d'investigació presentat en aquesta tesi doctoral utilitza models de risc validats externament per a la predicció i detecció de diabetis de tipus 2 en una base de dades poblacional de l'Hospital La Fe de València (Espanya). La hipòtesi de l'estudi és que la integració dels models de risc de predicció i detecció existents la pràctica clínica augmenta la detecció de casos d'alt risc. Per avaluar aquesta hipòtesi, s'han realitzat tres estudis sobre les dimensions clíniques, de l'usuari i de la tecnologia per avaluar fins a quin punt els models i l'hospital estan disposats a explotar aquests models per identificar grups d'alt risc i conduir estratègies preventives. Les troballes presentades sugereixen que els registres de salut electrònics no estan preparats per alimentar massivament models de risc. Alguns dels models avaluats han demostrat una bona classificació, el que va acompanyar a la bona acceptació d'eines basades en el web i el rendiment tècnic acceptable del sistema de tecnologia d'informació i comunicacions implementat. La conclusió es que encara es necesari treball per que aquests models poden conduir un nou paradigma de la detecció activa de la diabetis de tipus 2.

Acknowledgements

The work presented in this Doctoral Thesis has been possible thanks to the invaluable support, implication and guidance of Dr. Vicente TRAVER SALCEDO and Dr. Juan Francisco MERINO TORRES, to whom I want to acknowledge first.

My second acknowledgement is for the researchers, developers and business people from many universities and companies who helped me during the three years adventure which comes now to an end: María ARGENTE PLA, Carlos FERNANDEZ LLATAS, Bernardo VALDIVIESO, Lucas DONAT, Raquel FAUBEL, Jose Luis BAYO MONTÓN, Giuseppe FICO, Daniele SEGAGNI, Arianna DAGLIATI, Lucia SACCHI, Riccardo BELLAZZI, Rafael GABRIEL, Francesco SAMBO and Barbara Di CAMILO. I really fell like walking on the giant's shoulder.

Moreover, I want to specially appreciate the support of the entire SABIEN family: María MARTINEZ, Maria Jesús ARNAL, Gema IBÁÑEZ, Álvaro MARTÍNEZ, Álvaro FIDES, Zoe VALERO, Manuel TRAVER, Ignacio BASAGOITI, Aroa LIZONDO, Cristian AGUILAR and Christian LÓPEZ, the contributions of whom have been key to develop the work reported herein.

My third and last acknowledgement is for the people who have primed new ways of understanding the research process: Paulo de CARVALHO and Jorge HENRIQUES from University of Coimbra and, lastly Magda ROSSENMOLLER from IESE Business School and Ursula Mühle form EIT Health Europe.

Contents

List of Figures

List of Tables

List of Abbreviations

2h-OGTT	Two hour - Oral Glucose Tolerance Test
2hPG	Two hour Plasma Glucose
ADA	American Diabetes ASsociation
AMI	Acute Myocardical Infarction
AUC	Area Under the Curve
BMI	Body Mass Index
CODE-2	Cost of Diabetes Type 2
DW	Data Warehouse
EASD	European Association for the Study of Diabetes
EHR	Electronic Health Record
ETL	Extraction Transformation Load
EU	European Union
FIPA	Foundation for Intelligent Physical Agents
FPG	Fasting Plasma Glucose
GIP	Glucose-dependent Insulinotropic Polypepide
GLD-1	Glucagon-like Peptide
GLUT-2	Glucose Transporter 2
HbA1C	Glycated Hemoglobin
HF	Heart Failure
I2B2	Informatics for Integrating Biology and the Bedside
IAPP	Islet Amyloid Polypeptide
ICD	International Classification of Diseases
ICT	Information and Communications Technology
IDF	International Diabetes Federation
LDL	Low-Density Lipoprotein
MS	Metabolic Syndrome
NA	Not Available
NPV	Negative Predictive Value
NS	Not Specified
OLAP	Online Analytical Processing
PPV	Positive Predictive Value
QALY	Quality-Adjusted Life-Year
ROC	Receiving Operators Characteristic
S	Sensitivity
SGLT-2	Sodium-Glucose Cotransporter 2
SOAP	Simple Object Access Protocol
Sp	Specificity
SUS	System Usability Scale
T2DM	Type 2 Diabetes Mellitus
TAM	Technology Acceptance Model
UC	Use Case
UML	Unified Modeling Language

WHO	World Health Organization
WSDL	Web Service Description Language
XML	Extensible Mark-up Language

Dedicated to my wife Amparo Chillarón, my daughter Alejandra and my son to come, Bruno.

Chapter 1

Introduction

1.1 Motivation and rationale

Type 2 Diabetes Mellitus (T2DM) prevalence is rapidly rising throughout all the world (Guariguata *et al.*, 2014). In 2013 there were 382 million people with T2DM and there are estimates that the proportion of undiagnosed diabetes is reaching the 30% of the population (Beagley *et al.*, 2014). The disease is linked with an increased odds of morbidity and mortality, whereas there is strong evidence that early detection improves its prognosis in the 54% of the cases (Tuomilehto *et al.*, 2001; Knowler, 2002).

A booming field in clinical research is the use of mathematical models to assess the probability of an individual for developing a disease, also known as prediction models (Steyerberg *et al.*, 2013) or risk scores. Such scores are based on equations or probability relationships among multiple variables (such as demographics, laboratory tests and explorations) which have been collected on a specific context. The implementation of a risk score involves *i)* model development in a derivation subset of data, *ii)* model internal validation in a validation subset of data and *iii)* model external validation, to assess its performance in new data.

Methodological aspects in the development and validation of risk scores are in the scientific spot, with not a few concerns about model development, internal validation, external validation and impact evaluation (Hemingway *et al.*, 2013; Riley *et al.*, 2016). Nevertheless, Electronic Health Records (EHR) and computational modeling have paved the way to develop advanced T2DM risk scores to determine precisely high risk subjects, what may enable targeting effective preventive actions (Noble *et al.*, 2011). Unfortunately, there are a few scores which have been externally validated, and moreover, T2DM risk scores are rarely used in clinical practice. Shortcomings on external validation studies are attributed to generalization issues (context of data used for development extraordinary differs from external validation) and lack of data (variables not available or missing data). Moreover, quality of hospital based EHR is of a big concern, as it contains data routinely collected that might not be as rigorous as records done under a clinical study (Riley *et al.*, 2016).

A study on the implementation of a T2DM risk score in a clinical setting is a huge challenge that involves many dimensions, stakeholders and resources. In a boundary limited context it is possible to experiment how feasible is to incorporate software tools which integrate T2DM risk scores to identify high risk subjects in a health department. This doctoral thesis reports the process of designing, developing and evaluating the utilization of best-performing state of the art risk models for T2DM in Hospital La Fe (Valencia, Spain) clinical settings.

1.2 Type 2 Diabetes Mellitus

1.2.1 Pathophysiology and diagnose

Diabetes mellitus is commonly defined as a syndrome in which several factors related to genetic predisposition, lifestyle and environment (Dam, 2003) cause an impaired insulin secretion and/or action which leads to a chronic hyperglycemia (high blood glucose levels). Untreated diabetes conducts to several micro- and macro- vascular injuries on several organs (Steven M. Haffner *et al.*, 1998).

In a general way, the regular process of glucose metabolism starts when the human organism absorbs glucose molecules into the blood stream after digestion. The consumption of glucose molecules as energy input is mediated by an endogenous hormone produced by Langerhans islets β-cells in the pancreas. This hormone is the insulin, which is in charge of recognizing glucose molecules and activating the cell processes of glycolysis (glucose oxidation to obtain energy). Among the several types of diabetes (Thomas, 2015), T2DM is characterized by both an insulin action resistance and a progressive miss-function on the endogenous insulin release process. It is differentiated to other types of diabetes by the triggering factor, which is not related to an autoimmune (Type 1 Diabetes) or hormone-induced (Gestational Diabetes), but the long-term defect originated by ageing and obesity (Kahn *et al.*, 2006). Recent studies confirm a decrease of the onset age and T2DM has been reported in adolescents worldwide (Drake, 2002), particularly in high-prevalence populations (Alberti *et al.*, 2004).

Gold standard diagnose test to confirm T2DM is the Oral Glucose Tolerance test at 2 hours (2h-OGTT). In this test the subject intakes a 75g dose of glucose diluted in 3 dL of water (concentration <25 g/dL) through oral way in less than 5 minutes. Previously to the test, the subject has to achieve a basal metabolic performance by a specific food prescription, glucose-related drugs absentia and fasting for 8 hours.

The test is completed by two blood glucose determinations at the beginning and at the end of 2 hours. The patient has to lay down and rest during the entire 2 hour duration, in which no smoking and no drinks and food are allowed. If the first blood glucose determination (t=0) is over 126 mg/dL, the test should be re-scheduled for a second different determination. The American Diabetes Association (ADA, 2016) has recently defined the diagnose criteria for these tests (Table 1.1). In the absence of unequivocal hyperglycemia, results should be confirmed by repeating the test.

ADA DIAGNOSTIC TESTS CRITERIA FOR DIABETES DIAGNOSE
Fasting Plasma Glucose $\geq$126 mg/dL. Fasting is defined as no caloric intake for at least 8 h.
OR
2-h Plasma Glucose $\geq$200 mg/dL during an OGTT. The test should be performed using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water (WHO)
OR
A1C $\geq$6.5%. The test should be performed in a laboratory using a method that is NGSP certified (Little, 2003) and standardized to the DCCT assay (Sacks, 2012).
OR
In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose $\geq$200 mg/dL.

TABLE 1.1: American Diabetes Association criteria for diabetes diagnose

1.2.2 Prevalence and impact

T2DM accounts for the 90% of all diabetes cases and it has a prevalence in adult population that reaches 6% to 14%. Anyhow, prevalence and incidence are region and cultural specific (Talmud *et al.*, 2010).

According to two parallel independent studies, the prevalence of diabetes in adults (<18 years old) in Valencia Region (*Comunidad Valenciana*) is around 14% (Bueno H., 2008; Ruiz-Ramos *et al.*, 2006). This ratio is confirmed with the overall prevalence age and sex adjusted, which is 13.8%. It is remarkable that near the half of this (6%) had not been diagnosed yet. T2DM prevalence and glucose regulation impairments where significantly related to age ($p << 0.01$) and gender ($p << 0.01$), with an increased impact in males. T2DM prevalence is highly dependent with the age of the subjects, which goes from 2% on subjects under 20 years old to 35% for subjects over 65 years old. This means that one out of three adults over 65 years old has T2DM.

Social impact is quantified as the costs involved in the acquisition and purchase of treatments, productivity decrease as means of lost labor days, disability and mortality. Health care impact is quantified as the economic costs involved in hospital wards, out-patient services visits, primary care, pharmacological and not pharmacological treatment and tests (ADA, 2016). Moreover, cost increase as complications and multi-morbidity appear. There is a 50% increase on the average annual expense when a complication is detected and more than 360% when this complication is related to a cardiovascular disease. CODE-2 study (Massi-Benedetti, 2002) estimated the direct cost of T2DM on primary care, which was around 1305 euro/patient/year.

As population life-expectancy increases, T2DM prevalence will increase, driving to an augmented impact in social and health care environments. This prevision moves health organizations and governments to stress prevention as paramount.

1.2.3 Leading causes

T2DM appears as a consequence of insulin resistance and disturbance on the insulin secretion. This two processes are affected by each other, and depending on the leading factor, the disease will lead to a specific subtype (insulin dependent/ not insulin dependent /oral treated /not treated).

Insulin resistance is related to several metabolic changes, which are commonly labeled under the Metabolic Syndrome (MS). MS is generally characterized by having three or more biological abnormalities (increase on abdominal perimeter, increase of triglycerides, decrease of LDL and hypertension).

Insulin secretion is the response to glucose increase in blood stream as detected in B-cells in Langerhans islets in the pancreas. B-Cells catch glucose molecules by GLUT2 membrane transporter (without the mediation of insulin), and after a combination of enzyme reactions, the cell releases islet amyloid polypeptide (IAPP) and insulin hexamers, which lately and in the form of monomers mediate the glucose input into organic cells for metabolization.

The incretin effect, known as the relationship among intestinal hormones and B-cells, is the main cause to an insulin secretion impairment (Umpierrez *et al.*, 2013). This effect is leaded by two substrate molecules produced in the distal gut (GIP and GLP-1). This molecules are increased after the meal intake, even before the nutrient digestion to the blood stream, and several studies confirm their effect on the stimulation on the glucose-dependent insulin secretion. T2DM patients have been found with low levels of GLP-1, even though the impairment mechanism are still unknown.

Apart from these two factors, T2DM has been related to genetic inheritance. Specific related genes, associations in polymorphism and the complete genome have driven to different susceptibility sub-types of T2DM. The mediation of genes related to the insulin resistance, to the insulin secretion, to obesity and to the glucose metabolism have an additive effect on the clinical differences of T2DM and its treatments (Phillips *et al.*, 2014).

The weight of genetic and acquired factors is yet unclear (literature ranges from 28% to 80%). Anyway, obesity and sedentary are a key factors for the onset and evolution of T2DM, which get worse prognosis for genetic predisposed subjects.

1.2.4 Management and Treatments

T2DM is a chronic condition (it has no cure), but it can be controlled through a combination of lifestyle and pharmacological treatment. A good control on blood glucose levels is directly linked to an odds decrease of developing vascular and nervous complications. T2DM is commonly asymptomatic and is usually detected accidentally, as an abnormal blood glucose result in a routine test, and considered as a temporary event rather than a disease. But this is untrue most of the cases, as T2DM is a leading factor to cardiovascular diseases and atherosclerosis, which entails high mortality and morbidity rates (Arrieta *et al.*, 2016).

The main targets of T2DM treatment are to control blood glucose levels (and HbA1c) under normal thresholds to:

- Forewarn and delay macro-vascular complications and cardiovascular disease (AMI, HF)

- Forewarn microvascular complications (diabetic retinopathy, nephropathy among others)

Among the several strategies to keep under control blood glucose trends, the most common are:

- Changes on the food intake and nutritional habits. Adjusting the calories proportion of meals to the specific case and context of each subject and the overall strategy (weight loss or maintenance). Providing nutritional education is paramount to empower T2DM patients to design their own meal routines instead of following strict recipe compositions (Pan *et al.*, 1997).

- Regular Physical Activity which increases insulin sensitivity and improves plasmatic parameters like blood glucose, fatty acids. Intensity should be moderate and prolonged during more than 30 minutes. Moderate intensity is calculated with several parameters, but a naïve approach is on the average beats per minute, that should be around 50% and 70% of the maximum peace (beats per minute) which is 220- Age. (Tuomilehto *et al.*, 2001).

- Pharmacological treatment: There is no ideal pharmacological treatment for T2DM to help to control blood glucose levels (6%< HbA1c >7%) because all of them have side effects such as hypoglycemia, damage on B-cells and weight gain (Knowler, 2002):

  - Oral drugs to increase Insulin sensitivity (Metformin), which may also be prescribed when there is a suspect of T2DM.

  - Insulin secretion stimulants (Sulfonylurea), which also has secondary effects on weight gain and hypoglycemic events.

  - Incretin therapies. Some incretin-based effect therapies are being used due to their potential protection of β-cell mass and suppression of glucagon release.

  - Sodium–Glucose Co-transporter Inhibitors, which performs an inhibition of the type two co-transporter sodium-glucose (SGLT-2), by blocking glucose reabsorption in the proximal renal tubule. This lowers blood glucose and increases urine glucose concentration.

  - Subcutaneous insulin therapy to achieve and maintain adequate blood glucose levels. Normally the strategy is based on low-action insulin (basal) and fast-action (bolus) mixing.

- Bariatric surgery for cases in which lifestyle and therapy was not successful to control blood glucose values. Beneficial effect is for patients with BMI $\geq$35Kg/m^2 (Gloy *et al.*, 2013).

- Dyslipidemia and hypertension control, based on lipid-lowering and anti-hypertensive drugs (Bernstein *et al.*, 2015).

1.2.5 Screening strategies

Notwithstanding the advance on molecular and epidemiological science, underlying mechanism sand factors for T2DM onset and evolution are still unknown (Hippisley-Cox *et al.*, 2009). Increase of available data and knowledge from epigenetic studies may improve diagnose and prognosis, towards a most efficient and effective management of this disease (Meigs *et al.*, 2008). Several reasons support the importance of screening for T2DM:

- The growing prevalence of diabetes in the world
- The long asymptomatic period before it can be diagnosed
- The large number of people with unknown or undiagnosed TD2M
- Newly diagnosed patients have already evidence of microvascular complications

Although evidence for the screening of population groups is weak, cohort studies prove that early detection improves prognosis (Simmons *et al.*, 2010). There are four types of approaches for screenings:

- Population screening: it was never actually recommended by health and scientific organizations, all the proposals have been selective.
- Selective or targeted screening to be done on subgroups of a population in which it has been identified a certain risk in relation with some specific factors (e.g. age, weight, ethnicity etc.)
- Opportunistic screening: it is performed by health care professionals on subjects experiencing other health conditions (e.g. cardiovascular disease).
- Haphazard screening: this happens in countries without a consolidated screening policy as a result of a lack of coordinated actions that could be coherent with the real risk for the individual's condition.

Nowadays the best risk assessment strategy is to select the target population to be screened with any of the validated risk scores (selective screening): an early detection of new cases will make it feasible to start early intervention to treat the condition and reduce the risk of complications or reduce the risk of T2DM development (Noble *et al.*, 2011).

1.3 Risk Scores for Type 2 Diabetes Mellitus

A risk score aims to quantify the interaction among several factors and their relationship to classify a subject in a binary distribution (e.eg.: healthy/ill). Such factors may be subject, population or context specific, which increases the complexity of its validation and generalization.

Risk scores first emerged for predictive cardiovascular risk. The most widespread used is the Framingham 10 years cardiovascular risk event score (Agostino *et al.*, 2008). In the United Kingdom, electronic medical systems calculate Framingham score based on electronic health records as a decision support for therapy, and general practitioners are rewarded for using it (McElduff, 2004).

Ideally, a risk score should be developed under a large and defined group of population with shared characteristics (cohort), whom should not have the endpoint disease. During a sufficient period, subjects should be monitored for a baseline and follow up evolution of factors. Upon coherent time frames, recorded measurements should be analysed to see how this factors evolved and interacted among the people who developed the disease and the people who did not develop that disease (Altman *et al.*, 2009).

Another less expensive strategy is to perform cross-sectional studies or longitudinal studies based on retrospective data, in which the factors of a population of people with and without the disease are compared in a determined time point or during a period of time. Both approaches are potentially susceptible to bias (Noble *et al.*, 2011) because such methodologies mix factors among people with and without the disease, making impossible to elucidate disease-driven causal mechanism or flags.

A risk score is usually assessed by its statistical performance. According to the TRIPOD Statement (Collins *et al.*, 2015), a risk score goes beyond relative risks, odds ratio and hazard ratio and its performance should be reported with the following quantitative indicators (Noble *et al.*, 2011; Collins *et al.*, 2015):

- Discrimination: How the risk score is able to predict the proportion of subjects that are going to develop the disease with respect to the observed proportion. Quantitative indicators such as Sensitivity (S), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV) and the Area under the Curve (AUC) of the Receiver Operating Characteristics (ROC) are commonly used to compare the discrimination power of a model (Moons *et al.*, 2012).

- Calibration: Which examines the agreement between predicted and observed proportions by modifying the threshold (Dorresteijn *et al.*, 2011) (E/O statistic and Hosmer-Lemeshow *GoF* test)

- Generalisability: Capability of achieving good calibration and discrimination metrics in a fully independent (but similar) population from it was developed and internally validated (Altman *et al.*, 2009; Noble *et al.*, 2011).

Calibration and Discrimination indicators are defined and described in subsequent chapters.

Generalisability is a major issue for the applicability of risk models (Reilly, 2006). A risk model cannot be modified neither morphed when externally validated due to data scarcity or factors categorization (e.eg.: classification ranges, new categories) as it may have been developed for a specific set of factors and under specific circumstances.To this extent, some external validation studies increased discrimination metrics by adding new predictors, which corrupts the original aim of a proper external validation (Riley *et al.*, 2016).

The generalization and integration of risk models for cardiovascular diseases have reduced morbidity and mortality since 1970 (Agostino *et al.*, 2008). With respect to T2DM, even though much work has been done to assess the statistical strength of prediction and detection risk scores (Collins *et al.*, 2011b; Noble *et al.*, 2011), it is still missing its evaluation under a usability and impact framework to determine its power to reduce morbidity and mortality. Research is now focusing on how to best use existing risk scores on population datasets to identify high risk subsets for targeted public health interventions.

Chapter 2

State of the Art

2.1 Detection of TD2M and pre-diabetic stages

2.1.1 Standard diagnose

Standard T2DM diagnose is based on cut-off values for Fasting Plasma Glucose (FPG) and 2-Hour Plasma Glucose (2hPG) concentrations and Glycated Hemoglobin (HbA1C) (Table 2.1). A revision of the guidelines for TD2M diagnose defined by the WHO (World Health Organization) and ADA (American Diabetes Association) has been released in 2016 (Table 2.1).

The clinical criteria for the diagnosis of diabetes still include diagnostic cut-off values for FPG, 2hPG and HbA1C, but the guidelines differ as to which tests are recommended: IDF (International Diabetes Federation) , WHO (World Health Organization), and EASD (European Association for the Study of Diabetes) favor 2h-OGTT while the ADA promotes the use of HbA1c.

Diagnose	**Measurement**	**IDF/WHO/EASD (EASD, 2013)**	**ADA (ADA, 2016)**
Type 2 Diabetes	HbA1c	≥6.5% (48 mmol/mol)	≥6.5% (48 mmol/mol) **AND**
	FPG	≥126 mg/dL (L≥7.0 mmol/L) OR	≥126 mg/dL (≥7.0 mmol/L) **AND**
	2hPG	≥200 mg/dL (≥11.1 mmol/L)	≥200 mg/dL (≥11.1 mmol/L)
Impaired Glucose Tolerance (IGT)	FPG	≥126 mg/dL (≥7.0 mmol/L) AND	The A1C, FPG, and 2hPG under cut-off points.
	2hPG	140 - 199 mg/dL (7.8 - 11.1 mol/L)	
Impaired Fasting Glucose (IFG)	FPG	110 -125 mg/dL (6.1 -7.0 mmol/L)	
	2hPG	<140mg/dL (<7.8mmol/L)	
Not specific hyperglycaemia	HbA1C		5.7-6.4% (5.6-6.9 mmol/L)

TABLE 2.1: Standard Cut-off values for T2DM Diagnose

2.1.2 Early detection for delaying or preventing T2DM onset

There are multiple environmental and subject specific factor leading T2DM onset (Chapter 1). According to a review on the European guidelines for the prevention of T2DM, dietary habits and lifestyle are strong factors affecting on its development (Paulweber *et al.*, 2010). There is strong evidence (from meta-analysis and systematic reviews) on Randomized Clinical Trials which supports that subjects at high risk of developing T2DM whom were assigned with an intervention for weight loss (based both on healthy dietary habits and moderated physical activity) reverted the development of the disease in approximately 20 cases per 100 person-years (Gillies *et al.*, 2007).

Habit based interventions have shown also benefits With respect to the mortality, as concluded in the Malmö Feasibility Study (Eriksson *et al.*, 1998). This study compared dietary and lifestyle interventions with regular care (pharmacological treatment) during a 12-year follow-up. The study revealed that all-cause mortality among men in the lifestyle intervention group was lower than that among men who had received routine treatment (6.5 vs. 14.0 per 1000 person years).

A Swedish study calculated the distribution of the annual direct and indirect costs of T2DM (Norlund *et al.*, 2001). From the total cost per person, 28% of the costs were for health care expenses, 41% for lost productivity and 31% fell on the municipality and relatives. Direct and indirect costs of T2DM is high, as it is stated by the CODE-2 study (Massi-Benedetti, 2002; González *et al.*, 2006).

Emphasizing the importance of identifying cost-effective strategies for prevention, the Diabetes Prevention Program estimated that compared with placebo, the lifestyle intervention delayed the onset of T2DM by 11 years and decreased the absolute incidence by 20%. This translated into a cost per QALY of approximately 1100$ (having a cost per person in the United States is estimated on 2046$, this suggests a 53.7% of savings).

However, even though health benefits are clear and confirmed, most of the available evidence for the cost-effectiveness of prevention programs came from research settings and despite encouraging perspectives, so far the economic case for a widespread dietary, lifestyle or drug intervention to delay or prevent T2DM onset has not been made.

2.1.3 Risk scores to predict or detect T2DM

As mentioned in Chapter 1, a risk score is based on the relationships among subject variables. Many predictors (variables) have been proposed during the last decades, but less than one quarter have been externally validated (Buijsse *et al.*, 2011). Current ADA guidelines recommend screening for all overweight subjects with BMI $\geq$ 25 kg/m^2 of any age who have one or more TD2M risk factors (hypertension, family history etc) (ADA, 2015), whereas European Association for the Study of Diabetes and International Diabetes Federation recommend the use of a risk score questionnaire (Ryden *et al.*, 2013).

During the last 25 years many risk scores have been proposed and they have been compared in systematic reviews (Noble *et al.*, 2011; Collins *et al.*, 2012), with a no clear consensus on which is the best risk score.

On their review, Noble *et al.* (2011) analyzed 94 T2DM risk scores tested on 6.88 million participants. From these, authors judged 7 risk scores to be the most promising for use in public health practice. Table 2.2 and Table 2.3 reports these 7 studies which have been externally validated. Table 2.4 reports performance indicators of these risk scores without adding new predictors and without doing a re-training, just as the original models. Table 2.2 also includes the PREDIMED study (Guasch-Ferré *et al.*, 2012), not included in the review done by Noble *et al.* (2011) but relevant for the analysis of this sate of the art. These tables have been self elaborated taking the metrics reported in the original publications. Where a metric was not specified nor available, the table reports NS.

The validated Finnish Diabetes Risk Score (FINDRISC) has been successfully implemented as a practical screening tool to assess the diabetes risk and to detect undiagnosed T2DM. The Pizarra study (Soriguer *et al.*, 2012) had good results with the FINDRISC both to detect undiagnosed T2DM (AUC 0.74) and to predict incident T2DM (AUC 0.75). The best prediction of the risk of incident T2DM was found in those subjects with fasting glucose >100mg/dL and a FINDRISC score $\geq$ 9 (Odds Ratio: 19.37; 95% IC: 8,86-42,34; P«.001).

A recurrent problem when developing and validating is missing data. To reduce the biases that can occur in a complete-cases analysis, multiple imputation is frequently used to replace missing values for key risk factors (Collins *et al.*, 2011a). Multiple imputation (Rubin, 1996), is a statistical technique for analyzing incomplete data sets, that is, data sets for which some entries are missing. Missing data problem is not often reported though, and moreover several studies only report discrimination statistics and not calibration. A recent research project (MOSAIC, 2014) introduces an algorithm to explore the probabilistic relations between the set of variables comprising T2DM risk factors (Sambo *et al.*, 2015). In this work, based on a large longitudinal clinical study (Pyykkonen *et al.*, 2011), it was possible to identify and impute accurately missing values to build an assess T2DM risk scores.

The effectiveness of these models has recently been assessed in clinical practice systematic review (Barry *et al.*, 2017). Both standard diagnose cut-off points and prediction scores are specific but no sensitive, which leads to a high proportion of subjects with false negative results (i.e.: large number of undiagnosed and untreated population). Table 2.4 shows studies which were focused on providing high C statistic (Area Under the Curve of Receiver Operating Characteristic), instead of finding the most effective cut-off point that will maximize effectiveness of a particular intervention.

Screen and treat policies are limited because the effect of the intervention is limited to the duration of the clinical trial (ICER, 2016; Tuomilehto, 2014), with no real evidence of significant reduction of risk on the long term. On the basis of this state of the art on risk scores for T2DM, further research should undertake feasibility studies on the translational gap between clinical trials and real world limitations: data availability, acceptability of risk scores in clinical practice, implementation of decision support software and effectiveness of preventive interventions at population level.

Risk Score Name and Validation Study	Population Characteristics	Mathematical Model	T2DM Diagnose Criteria
Findrisc Internal (Lindstrom *et al.*, 2003)	NS Ages: 35-64 Follow up: 5 years	Weighted (Logistic regression)	WHO criteria (FPG or 2hPG)
Findrisc External (Alssema *et al.*, 2010)	Northen Europe, Dutch, Australian African. Ages: 35.2-71 Follow up: 5 years		WHO criteria (FPG or 2hPG)
ARIC Internal (Schmidt *et al.*, 2005)	United States Comunities Ages: 45–64 Follow up: 10 years	Loggistic regression	WHO criteria OR clinical diagnosis OR diabetic treatment
ARIC External (Mann *et al.*, 2010)	United States Comunities Ages: 45–84 Follow up: 4.75 years		FG>126 mg/dL or diabetic treatment
San Antonio Internal (Stern, 2002)	Mexican-Americans and Random Sample Ages: NA Follow up: 7.5 years	Linear regression	ADA criteria (FPG or 2hPG only)
San Antonio External (Abdul-Ghani *et al.*, 2011)	Finland and Sweden. Ages: 44-55 Follow up: 7-8 years		WHO criteria (FPG or 2hPG)
QDScore Internal (Hippisley-Cox *et al.*, 2009)	Caucasian Ages: 25-79 Retrospective (15 years) Qresearch Data Base	Proportional hazards model. Multiple imputation for missing variables	Diagnosis Read code for diabetes in EHR
QDScore External (Collins *et al.*, 2011a)	Caucasian (93% and other ethnic groups) Ages:,25-79 Retrospective (15 years) THIN DataBase		Diagnosis Read code for diabetes in EHR
Cambridge Internal (Rahman *et al.*, 2008)	UK population Ages: 40–79 Follow up: 5 years	Logistic regression	Diagnostic Code OR Diabetic Medication
Cambridge External (Talmud *et al.*, 2010)	UK population Ages: 35-55 Retrospective data base (11.7 years)		Diagnostic Code OR Diabetic Medication
PREDIMED Internal (Guasch-Ferré *et al.*, 2012)	Spanish Caucasian individuals Ages: 55-80 Follow up: 3.8 years	Multivariate Cox regression	ADA criteria (FPG or 2hPG only)
PREDIMED External (Guasch-Ferré *et al.*, 2012)	Spanish Caucasian High Risk individuals Ages: 45–75 Follow up: 4.2 years		ADA criteria (FPG or 2hPG only)

TABLE 2.2: Description of risk scores with potential adaptation in clinical use -1

Risk Score Name and Validation Study	Population Characteristics	Mathematical Model	T2DM Diagnose Criteria
Framingham Internal (Wilson, 2007)	NS	Logistic regression	ADA criteria (FPG or 2hPG only)
	Ages: 44.2-63.9		
	Follow up: 7 years		
Framingham External (Mann *et al.*, 2010)	Caucasian, African-American, Hispanic and Chinese-American		WHO criteria (FPG or 2hPG) OR Clinical Diagnose
	Ages: 45–84		
	Follow up: 4.75 years		

TABLE 2.3: Description of risk scores with potential adaptation in clinical use - 2

Risk Score Name and Validation Study	Sample size	Indicent Cases of T2DM	Cut-off point	S	Sp	PPV	NPV	AUC
FINDRISC Internal	4586	182	≥9	0.78	0.77	0.13	0.99	0.85
FINDRISC External	18301	844	≥7	0.76	0.63	0.11	NR	0.76
ARIC Internal	7915	1292	≥0.18	0.67	0.77	0.36	0.92	0.80
ARIC external	5329	446	NS	NS	NS	NS	NS	0.84*
San Antonio Internal	2903	275	NA	NA	NA	NA	NA	0.84
San Antonio External	2395	124	>0.065	0.75	0.72	0.119	NS	0.83*
QDScore Internal	3773585	115616	NS	NS	NS	NS	NS	0.83 men 0.85 women
QDScore External	2396392	72986	NS	NS	NS	NS	NS	0.80 men 0.81 women
Cambridge Internal	24 495	323	>0.37	0.55	0.80	NS	NS	0.75
Cambridge External	5135	302	>0.37	NS	NS	NS	NS	0.72
PREDIMED Internal	1381	155	≥ 6	0.72	0.72	0.25	0.95	0.78
PREDIMED External	552	124	≥ 6	0.85	0.26	0.25	0.86	0.66
Framingham External	3140	160	≥ NS	NS	NS	NS	NS	0.84
Framingham External	5329	446	≥ NS	NS	NS	NS	NS	0.83*

TABLE 2.4: Self-elaborated table containing discrimination performance of state of the art risk scores. (*) scores which needed a recalibration in the cut-off point, as initial the Hosmer-Lemeshow P value was not significant (<0.01).

2.2 Data Sharing Platforms and Services

Most of the previous risk scores were assessed over vasts amounts of data. For the purpose of this research is paramount to investigate which are the current trends on the technologies available to store large amounts of heterogeneous data. This chapter performs a review on which are the data engines capable of sharing and securely storing clinical records for the purpose of executing risk scores.

In 1993 Dr Edgar Frank Codd proposed a set of twelve rules to define the Online Analytical Processing (OLAP) for the design and implementation of Data Warehouses (E. Codd *et al.*, 1993). A Data Warehouse is an entity with specific features and built upon a specific architecture that provides sharing data services to other entities. The service distribution may be centralized, federated or distributed (Mazouzi *et al.*, 2002).

A Data Ware House is composed by one or more data bases (Figure 2.2), or Data Marts, that may have heterogeneous data models and structures, which makes very difficult to develop efficient use functions for Data Ware Houses (Herman, 2011). Use of knowledge domain descriptors and semantic references, through the definition of an Ontology, is key to formalise and map the type of data hosted in a Data Ware House (R. Roset *et al.*, 2008).

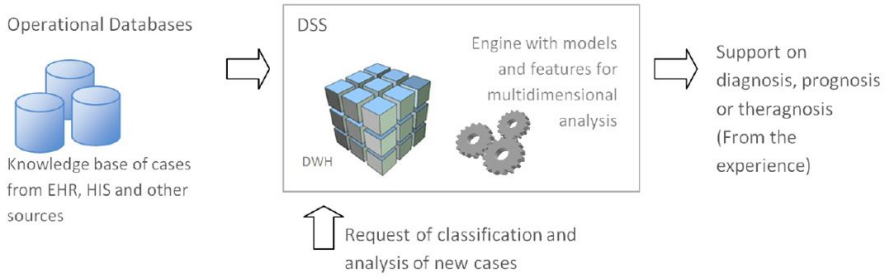


FIGURE 2.1: Data Ware House conceptual System for Decision Support Systems (Lluch-Ariet, 2016)

Even though classic SQL engines are still hard to beat, there are several commercial and non-commercial database engines with top featured options on volume, variety, speed and reliability such as MongoDB and NoSQL systems. But, regardless the engine performance, interoperability is a key factor to design a proper Data Warehousing (DW) system.

To this end, Informatics for Integrating Biology and the Bedside[1] is one of seven centers funded by the NIH Roadmap for Biomedical Computing [2]. The mission of I2B2 is to provide clinical investigators with a software infrastructure able to integrate clinical records and research data. The I2B2 tools have been exploited to support projects across the Partners Health Care system and approximately 17 sites outside of Partners Health Care are engaged in setting up their own I2B2 software

[1] I2B2 Web Site. http://www.i2b2.org/software. (Last Access 01/03/2017)
[2] NIH Web Site. http://www.ncbcs.org. (Last Access 01/03/2017)

systems to support enterprise discovery based on medical record data. The key feature of this tool is that it "fits together" medical record data and clinical trial data at a person-level so that diseases, genes, and outcomes can be related to each other.

There are several reasons why I2B2 stands as a powerful tool among the variety of available DW development tools for data sharing: first, I2B2 is the only software that is patient-specific and supports the use of ontologies for querying the DW. For this reason, there's no need to use DW dedicated languages to perform queries. Second, it provides a flexible model-data views to define and update data models based on ontologies. Third, it provides security features as means of restriction views to allow specific access, modify and replace operations only to authenticated users. Fourth, it can be implemented in a wide variety of data engines (MySQL, ORACLE, SQL Server, etc..) and operative systems(Windows, Linux), being portable trough virtual machines systems (HyperV, VMWare).

The I2B2 DW meets the requirements to reach the goals of an efficient data gathering strategy as makes possible to: *i)* collect multidimensional data *ii)* integrate different sources of information and *iii)* aggregate data and export them in a format suitable for temporal dimension analysis.

However, this new paradigm of clinical data organization still stands as a novel technology for European researchers. There is a huge opportunity for EU research institutions to adopt and engage their staff to exploit I2B2 paradigm to collect and share clinical data.

I2B2 architecture is made up of three layers: a Presentation Layer, a Service Layer and a Data Layer. The user accesses I2B2 at the Presentation Layer, which exposes a user interface (UI) either through a Web Client or a local application.

Data are stored into the Data Layer, which contains the I2B2 DW. The only way the UI can access data is through the Service Layer. This layer is a collection of web services, each one denoted as a "cell". The collection of these cells makes up the "I2B2 hive". The main cells in the hive are: the Project Management (PM) cell, the Clinical Research Chart (CRC) cell and the Ontology Management (ONT) cell.

The PM cell accesses a set of data structures in the DW that associate users with passwords, preferences and projects. When a user logs on to the I2B2 web client, the PM cell manages the authentication process. Every time another part of the hive tries to perform an action on behalf of the user, it goes to the project management cell to gather the proper authorizations. Once authenticated, the user (through the Web Client) performs queries through the CRC cell, also known as the data repository cell. To facilitate the query process for the user, data are mapped to concepts organized in an ontology-like structure, which is managed and accessed by the ONT cell.

The I2B2 data model is based on a "star schema". The star schema has a central "fact" table where each row represents a single fact. In I2B2, a fact is an observation about a patient. Observations about a patient are recorded by a specific observer in a specific time range (defined by start and end dates) and are related to a specific concept, such as a lab test or diagnosis, in the context of an encounter or visit. The concept can be any coded attribute about the patient, such as a code for a disease,

a medication or a specific test result. This way of representing concepts is based on prior work known as the entity-attribute-value (EAV) model (Murphy *et al.*, 2010). The reason why the I2B2 developers decided to implement this model is that querying data modeled with a star schema represented in an EAV format is extremely efficient.Figure 2.2 shows the I2B2 star schema as presented in the I2B2 CRC Design Document (Murphy *et al.*, 2009).

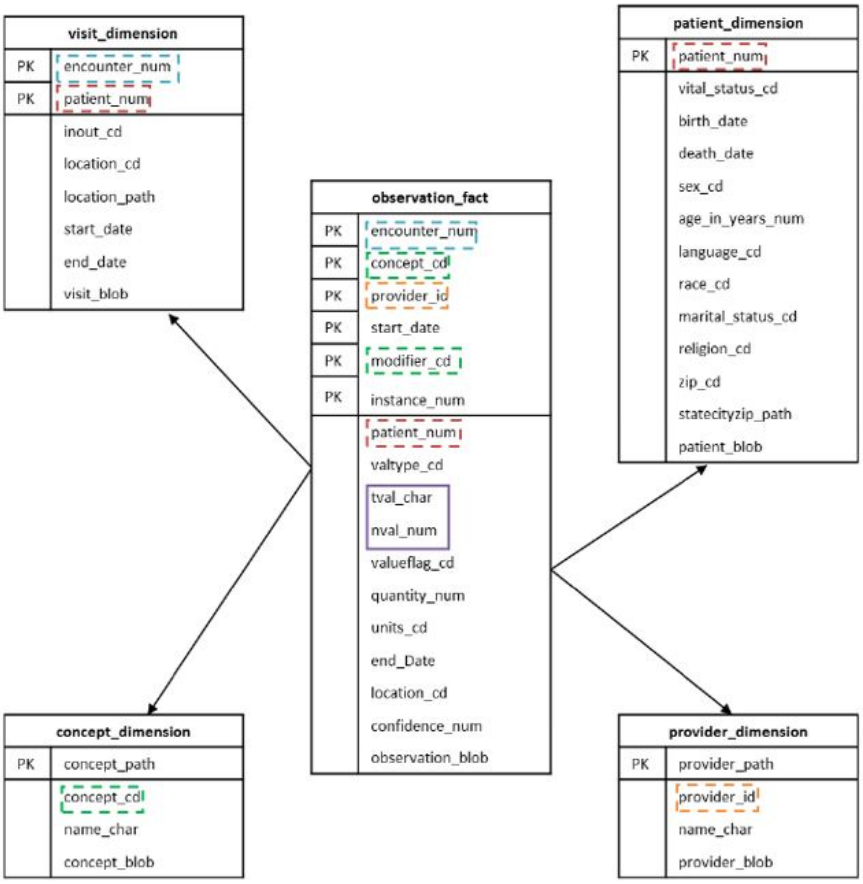


FIGURE 2.2: I2B2 star-schema. All the tables have the same basic structure. The structure of the meta-data traces the visualization of concepts in the I2B2 data visualization layer and is the underlying structure for querying the data

The central table of the I2B2 Star Schema is the *observation_fact* table, referred also to as the "fact table". It contains all the quantitative or factual data coming from observations about each visit related to each patient, and it is the table where all the values of each observation are stored.

In the health care context, a fact is an observation on a patient. It is important to note that such observation may not represent the onset or date of the condition or event being described, but instead is simply a recording or a notation of something. For example, the observation of diabetes recorded in the database as a fact at a particular time does not mean that the condition of diabetes began exactly at that time, only that a diagnosis was recorded at that time (there may be many diagnoses of diabetes for this patient over time).

Because most queries in the I2B2 database require joining the observation_fact table with one or more dimension tables together, the observation_fact table, as the central fact table of the schema, is the intersection of the dimension tables (visits, patients, concepts and providers). In the observation_fact table, facts are defined using concept codes (concept_cd). Concepts are organized in a hierarchical structure: the I2B2 ontology (also called meta-data). Each concept in the ontology is represented by a meta-data table (Murphy *et al.*, 2009), which is stands as a sufficient solution to overcome the interoperability issue among different data models and data types.

2.3 Integration of statistical engines

A risk score is based on mathematic and probabilistic computing over a specific set of variables. Even though the mathematical rules to execute a logistic regression model or a Cox Hazards model are relatively simple, the operations needed to develop, validate, re-train and assess a classification model need powerful tools and engines. This chapter overviews the state of the art statistical engines.

For its calculations, a risk score requires a proper mathematical framework and, for some cases, specific libraries (Collins *et al.*, 2016). Data scientists and researchers usually work in a Integrated Development Environment in a computer which has access to the datasets (locally or remotely) in the form of files or data bases (Figure 2.3).

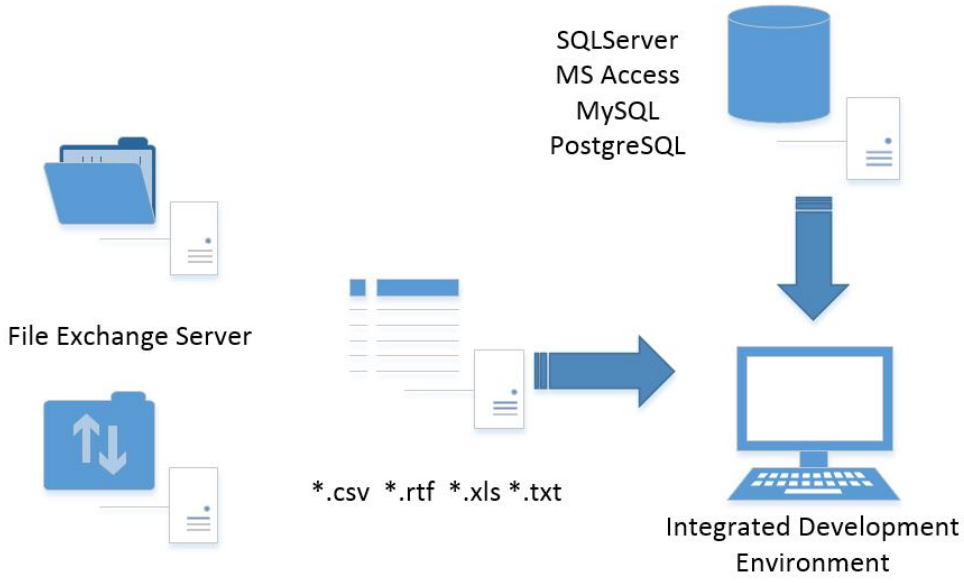


FIGURE 2.3: Development and execution of Risk Scores under a monolithic architecture

There are several commercial and non-commercial mathematical development environments and engines (Python [3], Octave [4]), but among the most extended in the academic domain, there are R and Matlab.

[3] Python.https://www.python.org/
[4] Octave. http://www.gnu.org/software/octave/about.html

2.3.1 R software

R is currently one of the most powerful free statistical environments (Hornik, 2012). It is available in source code under the terms of Free Software Foundation's GNU General Public License. R is constantly expanding via new functions (packages) and contains a complete set of mathematic and statistical tools for general purpose. University of Pavia has created a novel software bench that aims to integrate the R Engine (Segagni *et al.*, 2011) that allows the communication between an I2B2 architecture and R software. On their approach, *Segagni and colleagues* have implemented a web-client plug-in which allows to execute a Kaplan-Meier analysis (Survival models) on a selected subset of data stored in a I2B2 data warehouse.

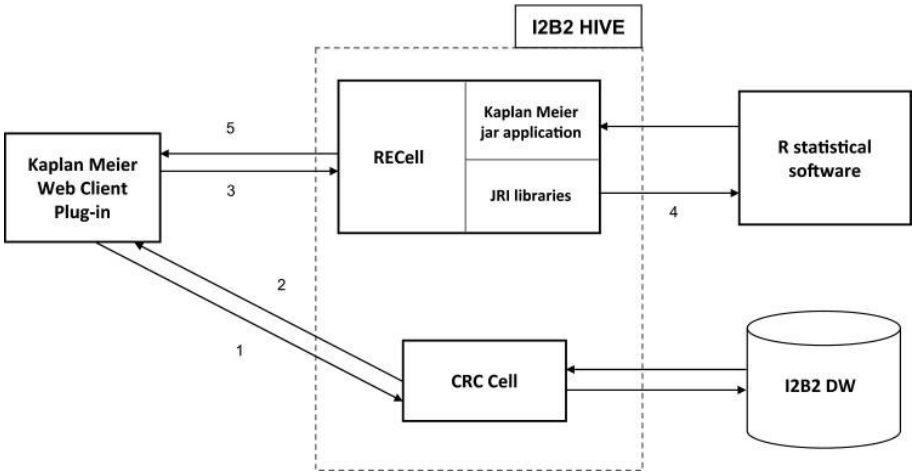


FIGURE 2.4: Rcell implementation with a mathematical filter

R Cell communicates with the I2B2 through *RESTful* web services (Richardson *et al.*, 2008) using compliant markup language to run predefined R statements(Figure 2.4). To execute the model, the user has to retrieve observations from the data warehouse using the web client plugin (1). Once data is retrieved a custom XML compatible with the i2b2 data structure is sent to the R Engine Cell (RECell) (2) and, then the packages that integrates the mathematical model is executed on the selected dataset using the R engine (4). Once executed, the RECell returns the outcomes to the web client plugin (5), enabling the user to download them to further analysis.

2.3.2 Matlab software

Another popular resource for developing and implementing a wide range of applications in research and academics is Matlab (Guide, 2014). Due to the different toolboxes (like Real Time Control, Image Processing and Machine Learning) it has, Matlab with Simulink provides a good environment for risk modeling development. Matlab is based on a proprietary engine and a friendly Graphical User Interface which allows to develop scripts in common programming languages, object oriented programming and a featured debugger. There is some criticism though with respect to the execution of heavy operations, when the efficiency seems to be compromised (Help, 2017)

The underlying reason for this is that Matlab engine has to convert its own language instructions into machine code at the same time it execute the processes. Nonetheless, there are tools to increase efficiency by building pipelines between Matlab and C/Fortran compilers[5].

Matlab supports several possibilities for data exchange with other applications through COM objects, developed initially by Microsoft. COM defines a language independent binary standard for component interoperability. It is used to enable inter-processes communication and dynamic object creation in any programming language that supports the technology.

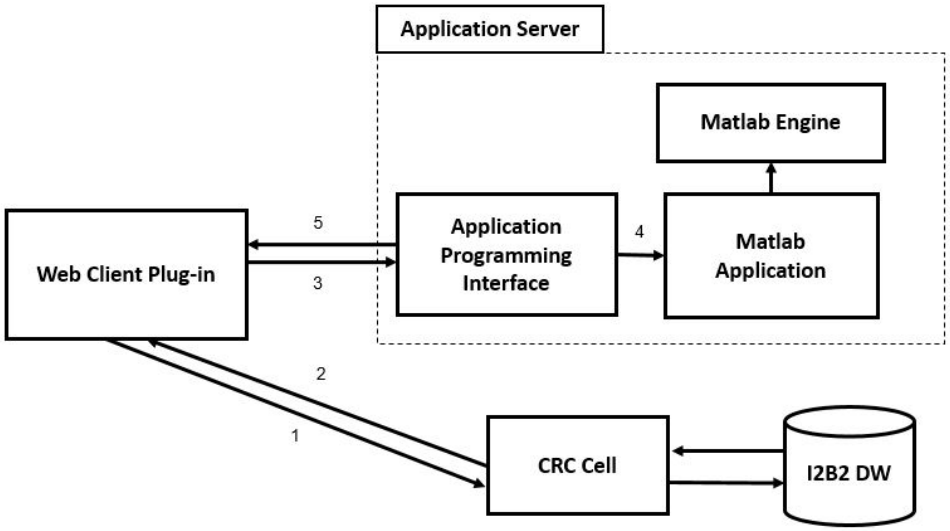


FIGURE 2.5: Matlab integration in a distributed architecture

Using COM, developers and end users can select application specific components produced by different vendors and integrate them into a complete solution. The essence of COM is a language neutral way of implementing objects such that they can be used in environments different from the one they were created in, even across machine boundaries. Using COM technology Matlab can control another component or be controlled by another component.

Therefore, to integrate Matlab software it is possible to combine client or server applications application (Java, .Net,...), with the Matlab engine through COM. It is possible to build open solutions by implementing platform independent programming languages in the server side connected directly to the Matlab engine, which is capable of executing Risk Models over a selected subset of population.

The client-server approach aims to distribute the execution of Matlab application from the data base management and client application. The execution work flow is similar to the R case, but without using the I2B2 workbench.

[5]Low-Level Coding adaptation in Matlab. https://www.mathworks.com/help/matlab/ref/mex.htm

2.4 Software architectures to support risk modeling

Software architectures are based on Agent Technology or Multi-Agent Systems, which is a growing discipline focused on the communication between independent distributed entities. According to a recent research on networked clinical systems (Lluch-Ariet, 2016), an Intelligent Agent is an entity capable of:

1. Acting according to a knowledge base
2. Reasoning to determine internal behavior according to rules and restrictions
3. Communicating with software and human entities
4. Having a unique identity.

Multi-Agent architecture has been extensively applied with the health care and health management (Moreno *et al.*, 2006; Martinez-Millana *et al.*, 2015).

Agent architectures, as means of communication, management and applications, were initially promoted by the Foundation for Intelligent Physical Agents [6]. In 2005 IEEE standards committee created the IEEE-FIPA standard specifications classifying the different working lines into five major categories:

1. Agent Communication Language (ACL) specifications, for the messaging structure and formats, interaction protocols and content language.
2. Agent Management and Control specifications, which has to do with the rules and states of agents within a multi-agent environment.
3. Agent Message Transport specifications, for the definition of the possible network level protocols for message dispatching (e.eg: HTTP).
4. Abstract Architecture specifications, which perform a high-level description on the abstract entities needed to bould a multi-agent environment. (This concept is further developed in the IEEE 42010 standard)
5. Application Specifications, as an example of the possible specific context to deploy multi-agents.

Still though, FIPA implementation is generally used as a reference framework to build context-specific applications which do not meet all the standard specifications (Komatsoulis *et al.*, 2008).

To this extent, when the problem is on how to describe a complex environment as a prior step to the definition of functional requirements for a posterior real deployment, we should reference ISO/IEEE 42010 standard for Systems and Software Architecture Description.

The standard was superseded in 2011 and literally "*addresses the creation, analysis and sustainment of architectures of systems through the use of architecture descriptions*". The main strength of this standard is the separation between knowledge domain and technology, leaving the former for subsequent implementation stages and non-compromising the faithful understanding of all the possible variations in a context. The standard defines conceptual foundations and models, and stands the description of a software architecture on 4 different views:

[6] FIPA Web Site. http://www.pa.org/. *(Last Access 25 Feb 2017)*

- architecture description (AD), the overall description of artifacts, interactions and rationale.

- architecture viewpoint, as the description of the entire architecture from the viewpoint of each component or stakeholder.

- architecture framework, which establishes a common practice for using, creating, interpreting, and analyzing architecture descriptions

- architecture description language (ADL), which states that the description needs to be based on existing languages like Unified Modeling Language [7], BPMN, Architecture Analysis & Design Language, Acme, ArchiMate.

Besides the standards, there is a European initiative named FI-STAR [8] which is currently proposing an open stack cloud based framework for the integration of ICT systems: the FI-WARE infrastructure. FI-WARE is currently a project that will provide an open source platform, based upon a series of elements (called Generic Enablers) which offer reusable and commonly shared functions serving multiple areas of use across various sectors (Figure 2.6). In FI-WARE, the Cloud Hosting is the fundamental layer to provide computation, storage and network resources services to be provisioned and managed. The abstraction of software and hardware resources is managed by Generic Enablers (GE) and the Service Management (SM) modules, which interact across them to provide a dynamic environment for system hosting and development.

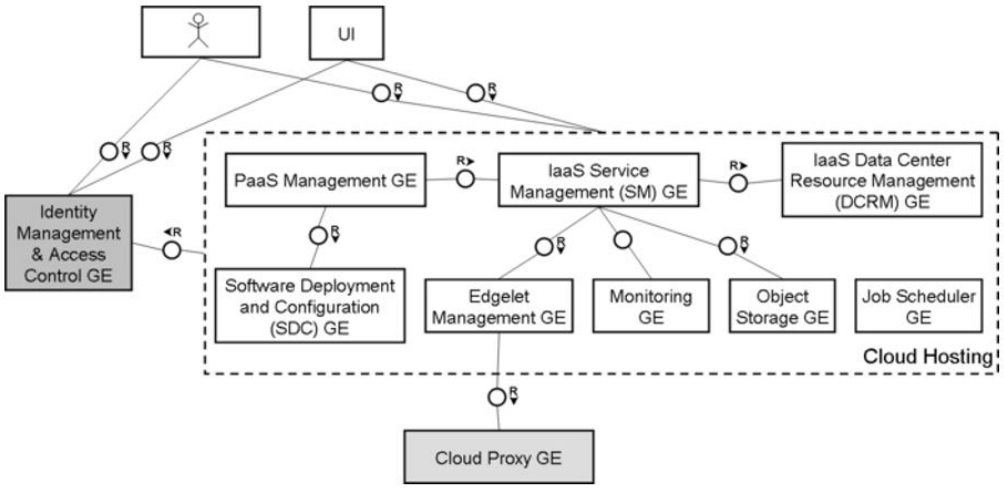


FIGURE 2.6: FIWARE Cloud architecture

- IaaS Service Management (SM) GE may invoke Aapplication Programming Interfaces (APIs) of IaaS Data Center Resource Management (DCRM) GE to perform operations on virtual resources (mainly virtual machines) which comprise the services managed by SM GE.

- Cloud chapter GEs uses Identity Management and Access Control APIs for authentication and authorization purposes

[7] UML Web Site. http://www.uml.org/ (*Last Access 20 Feb 2017*)
[8] FI-STAR Web Site. https://www.fi-star.eu/about-fi-star.html. *(Last Access 28 Feb 2017)*

- IaaS Service Management (SM) GE may use APIs of the Monitoring GE to collect metrics of the underlying resources which comprise the service, to drive service elasticity.

- PaaS Management GE will use IaaS Service Management GE to drive provisioning and auto-scaling of the Virtual Machines s composing the PaaS software stack.

- PaaS Management GE will use Software Deployment and Configuration (SDC) GE to install and configure the software components running within the individual virtual machine comprising the PaaS environment.

Apart from the Cloud hosting, there are available other kind of services, such the Services Ecosystem, which is the infrastructure to create, publish, manage and consume services across their life cycle, addressing all technical and business aspects, or the Data/Context Manager, which facilities a layer for effective accessing, processing, and analyzing massive data volumes, transforming them into valuable knowledge available to applications.

Chapter 3

Hypothesis and Objectives

3.1 Study hypothesis

T2DM incidence is increasing world wide, accounting for a decrease on the quality of life of people who develops it and an increase on both direct and indirect costs for health care systems and societies.

Randomized Clinical Trials and cost-effectiveness analysis reveals that prevention programs focused on dietary, lifestyle and pharmacological intervention on high-risk T2DM subjects increases the odds of delaying or preventing T2DM onset, with the subsequent benefits for patients, citizens and economic savings.

To this end, several risk scores for the prediction and detection of T2DM have been developed and validated in research settings, achieving good performance for identifying high risk subjects that could benefit the most from the preventive interventions.

Therefore it is reasonable to suppose that:

The integration of Existing Prediction and Detection Risk Scores for Type 2 Diabetes Mellitus based on Electronic Health Records increases the early-detection of high risk cases in La Fe Health Department in Valencia (Spain).

3.2 Objectives

To assess the study hypothesis, the research work plan needs to target one primary objective which will be supported by the accomplishment of three secondary objectives (Figure 3.1)

3.2.1 Primary Objective

The study main objective is to evaluate a software system that integrates T2DM risk scores using retrospective electronic health records and inform implementation outcomes of this in clinical practice.

To achieve this primary aim, three sub studies have been identified:

- Clinical Evaluation of the risk score models by using retrospective hospital records to identify population with risk of developing T2DM and to compare it to the outcome.

- Usability Evaluation of the tools by monitoring the use of the implemented tool performed by medical professionals as doctors, endocrinologists and head of service on the analysis of the retrospective datasets and the analysis of the results two standard User Experience and Usability questionnaires.

- Technical Evaluation of the tools by performing an analysis on standard key performance indicators on different use scenarios by monitoring system load, latency, overflow, computational resources and response.

3.2.2 Secondary Objectives

1. To create a data warehouse to integrate Electronic Health Records datasets.

2. To define the usage scenarios for risk models at a population level.

3. To implement a software platform to enable the risk model execution over electronic health records in clinical settings.

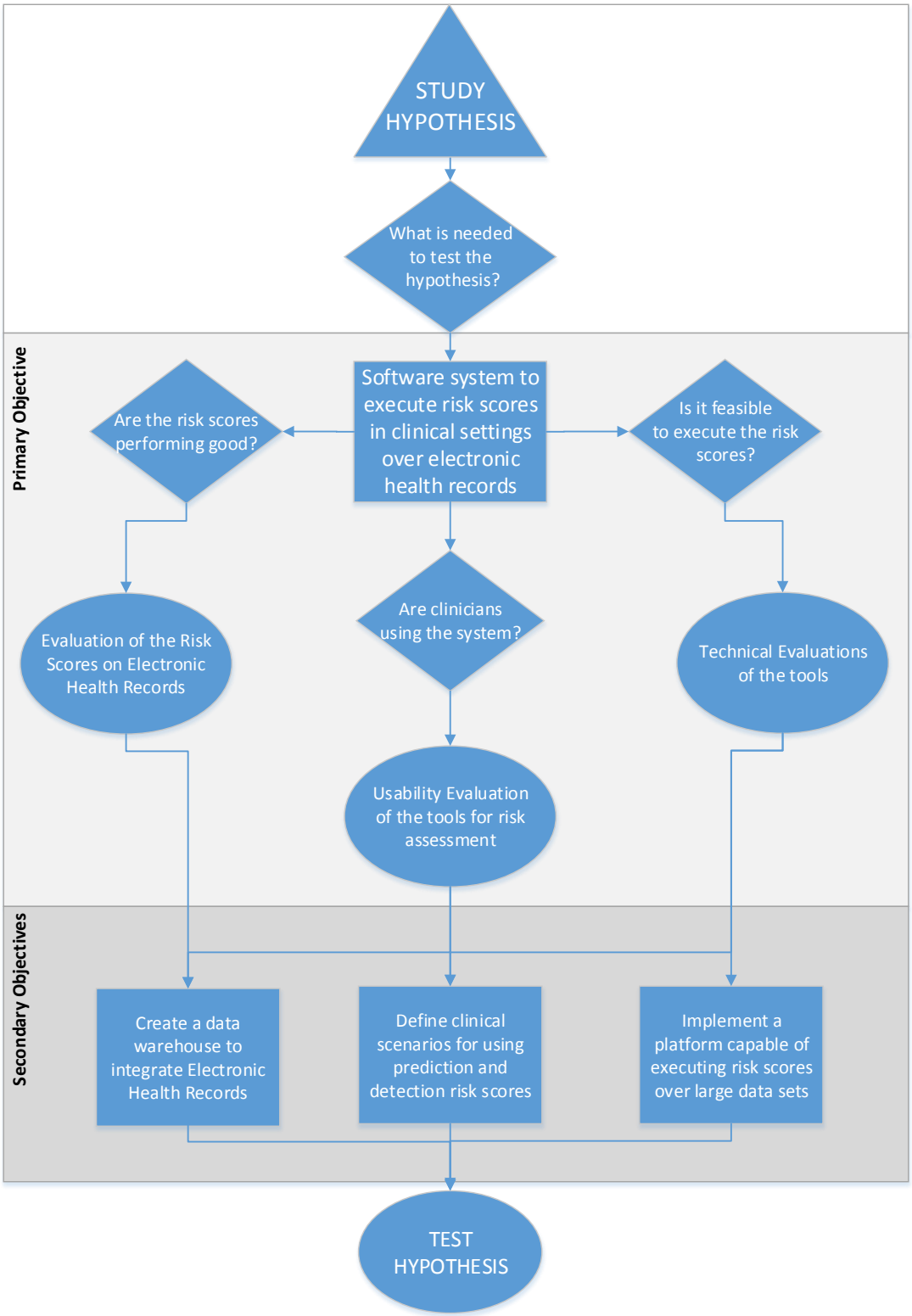


FIGURE 3.1: Conceptual view and relationship diagram of the study hypothesis, primary objective and the secondary objectives

Chapter 4

Materials and Methods

For the purpose of this research, the needed materials have been based on:

1. Electronic Health Record integration.
2. Selection of the risk scores, according to the available datasets and the clinical objectives.
3. Definition of the usage scenarios for T2DM risk assessment.
4. Software Infrastructure and tools to execute T2DM Risk models on Electronic Health Records.

To this end, risk scores and Electronic Health Records are the basement to elucidate the functional and technical specifications and user requirements to build a comprehensive solution for the assessment of T2DM risk scores in clinical foundation. This study has been driven inside the European Research Project MOSAIC, introduced in Section 4.1. The Electronic Health Records used in the study and their characteristics are described in section 4.2. A description on the T2DM prediction and detection risk scores and the evaluation criteria is done in section 4.3. The definition of the usage scenarios and the evaluation methodologies are described in section 4.4. Finally, a description of the software infrastructure used to integrate Electronic Health Records, Risk Scores for T2DM and the applications to be used by clinicians is done in section 4.5.

4.1 Context of the study

FIGURE 4.1: MOSAIC Project

The study was pursued under the execution of MOSAIC European Project [1]. The MOSAIC project is funded by the European Commission under the 7th Framework Program, Theme ICT-2011.5.2 Virtual Physiological Human (600914).

[1] http://www.mosaicproject.eu/ (Last Access March 2017)

MOSAIC provides tools that enable the implementation of better methods for advancing diagnose of type 2 diabetes and pre-diabetic states, which has the potential to improve the prognostic of the disease as it would allow the personalization and assignment of early treatments. Such innovative therapies, if followed correspondingly, can reduce significantly the complications suffered by patients and delay and/or avoid the onset of other morbidities. This also has an impact in the quality of life of the diabetic population and leads also to an increased productivity of those patients in their advanced ages.

MOSAIC consortium is constituted by 10 different partners from 5 EU countries that have collaborated for 42 months (Jan 2013-June 2016) in the definition, design, development and exploitations of the outcomes of the project.

MOSAIC aimed at generating new industrial activity by promoting developments around remote monitoring devices and systems and ICT tools for further exploitation of the MOSAIC products. MOSAIC also explored the capabilities of the models to be modified for its application in other metabolic disorders, opening the door to the creation of new research lines and also for its development and further exploitation.

Antonio Martinez-Millana was leader of the Software Architecture definition, development and implementation tasks of the Prediction and Detection Use Case of the project, which was piloted using retrospective data sets in La Fe Hospital. All the work reported herein corresponds only to the tasks developed entirely by Antonio Martinez-Martinez with the support of Jose Luis BAYO MONTÓN on the development of software components and Maria ARGENTE PLA in the case-by-case subject supervision.

4.2 Study data

The medical center participating in the research project is the Health Department Hospital La Fe in Valencia, Spain (La Fe). Within the information system of the hospital, each service and database has a different structure and different way to code variables (data types and values), as Figure 4.2 shows.To query the data in a consistent way, it is important to share a common data model with a homogeneous representation of the collected parameters. To this end, the data available was analysed with the aim of defining the most efficient strategy to gather and organize the data using a common ontology. Hospital Information System Staff have actively participated to these activities, providing detailed descriptions of their clinical databases and the necessary medical and scientific knowledge to set up the processes of data mapping and parameters selection

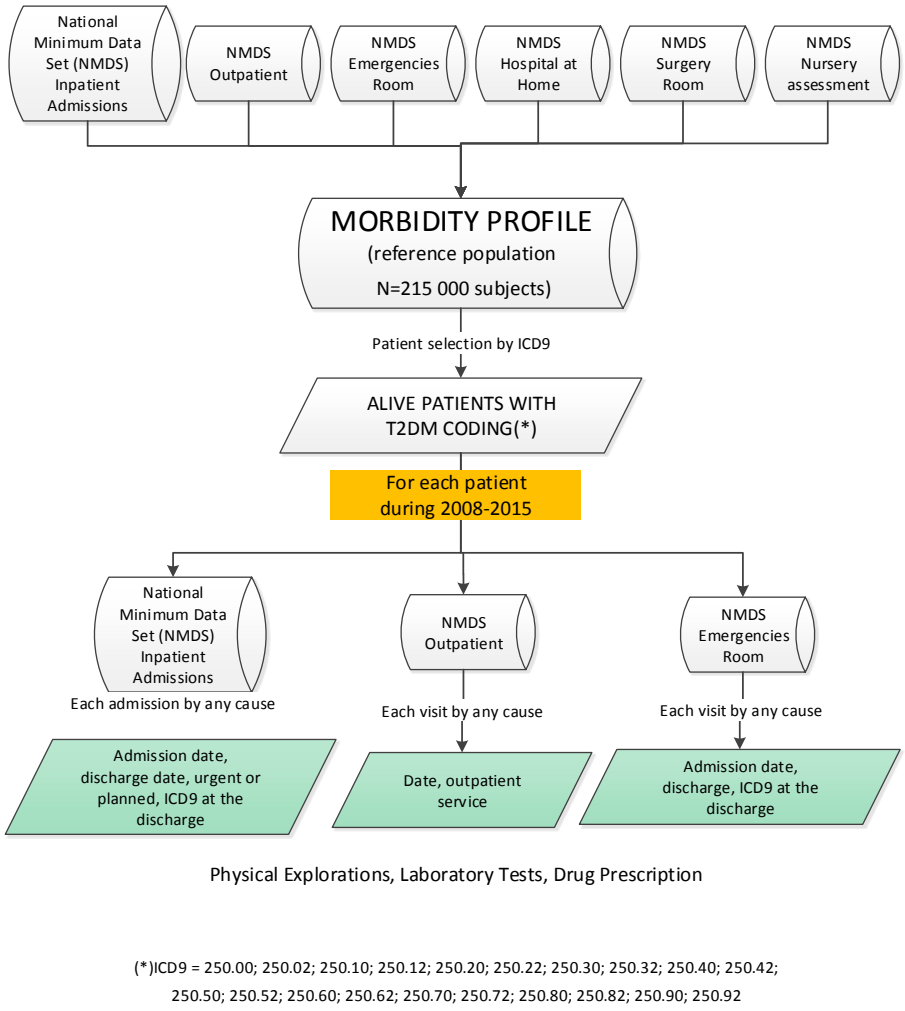


FIGURE 4.2: Hospital Electronic Records extraction from different services

La Fe Public Hospital[2] is the reference clinical setting of La Fe Health Department. La Fe Department is a geographical district that covers a population around 300,000 inhabitants and it includes a University Hospital, two specialties centers and twenty primary care centers. This health department accounts for more than 1,100 doctors, 400 residents in training and about 3,800 people in the areas of nursing whom provide universal health care services.

[2]La Fe Health Department. http://www.lafe.san.gva.es/departamento-de-salud-valencia-la-fe. *(Last Access 1 Feb 2017)*

In January 2015, after formalising the request to the Biomedical Research Ethics Committee in Hospital La Fe a Microsoft Excel (xls) document is provided with the structure in Table 4.1 and Table 4.2.

Field	**Description/Value**
PID	Patient unique identifier (alphanumeric key)
Sex	1 Male; 2 female
Diabetic	1 if the patient contains an ICD code 250.XX - 0 if not
Residence country	Code for the residence country
Residence province	Code for the residence province (46 for Valencia)
Residence city	Code for the residence city (250 for Valencia)
Postal Code	Postal code of the patient residence (46000)
Birth date	Complete date for borning date (DD/MM/YYYY hh:mm)
Birth Country	Code for the birth country (108 Spain)
Primary Care Center	Name of the patient's primary care center (string)

TABLE 4.1: Demographic descriptors

Service	**Variable**	**Description/Value**
Hospitalization	Admission date	Datetime (DD/MM/YYYY hh:mm)
	Discharge date	Datetime (DD/MM/YYYY hh:mm)
	Duration	days (Numeric)
	Diagnose and procedures	ICD9 code
	Main diagnose	1 YES;2 NO
Emergency Services	Admissiondate	Datetime (DD/MM/YYYY hh:mm)
	Dischargedate	Datetime (DD/MM/YYYY hh:mm)
	Diagnose and procedures	ICD9 code
Oputatient Services	Service ID	Outpatient service code
	Visit date	Datetime (DD/MM/YYYY hh:mm)
	Birth date	Datetime (DD/MM/YYYY)
	Birth Country	Birth country code
	Primary Care Center	Name of the patient's primary care center (string)
Laboratory Tests and Observations	Test ID	Numeric
	Test Name	String
	Timestamp	Datetime (DD/MM/YYYY hh:mm)
	Result	Double

TABLE 4.2: Observations

The strategy to map the data coming from different services had several objectives:

- Identification of the parameters that are in common between the different services.
- For such parameters, share a common representation of the variables (same units of measurement, same coding system, same type of representation);
- Define a common data structure to: build a integrated data warehouse with a common ontology to facilitate data sharing for analysis purposes.

I2B2 Workbench and Pentaho [3] open source development environment were used to accomplish this goal.

4.3 T2DM risk scores

The selection of a risk score depends strongly on the performance metrics Tables 2.2 and 2.3, but moreover on the data availability. State of the art risk models are based on regression models executed on numerical and/or categorical variable. Depending on the output, such models can provide the probability p of developing or having T2DM (Equation 4.1), or the hazard rate of developing T2DM over time (Equation 4.2.

$$p = \frac{1}{1 + \exp(-(\alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m))} \tag{4.1}$$

$$h(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m) \tag{4.2}$$

$$p = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m \tag{4.3}$$

Where:

- α is the intercept or prior probability.
- $h_0(t)$ is the intercept baseline hazard rate.
- β_x are the regression coefficient which denotes the relative weight of the corresponding predictor.
- X_x are the predictors or variables, which can be numerical (continuous) or categorical (0,1,2...).

[3] www.pentaho.com/

Risk scores output, p and $h(t)$, are assessed with discrimination and calibration metrics (Chapter 4.2.1). To be useful and properly assessed, a risk model should be calculated taking into account all the requested predictors. It is well known that the output of the risk scores (p and $h(t)$) depends on the distribution of predictors in a population, but moreover with the prevalence of T2DM in that population.

Concerning the predictors, Figure 4.3 shows the variables needed for running validated state-of-the-art risk scores, their intercepts and regression coefficients.

	FINDRISC	ARIC	San Antonio	PREDIMED	FRAMINGHAM	CAMBRIDGE
Intercept	-5.51	-9,981	-13.415		-5.427	-6.322
Age	45-54: 0.63 55-64: 0.89	0.0173	0.028		50-64: -0.010 ≥65: -0.2107	0.063
Gender			Female: 0.661		Male: 0.113	Female: -0.879
Ethnicity		African American: 0.443	Hispanic: 0.412			
Anti-Hyp. Medication	0.71			0.838	0.336	1.222
Prescribed Steroids						2.191
Fasting glucose (mg/dL)	>110: 2.14 (*)	0.140	0.079	>100: 1.929	≥100: 1.67	
BMI (kg/m2)	25-30: 0.17 >30: 1.10		0.070	≥ 27: 0.315	25-30: 0.157 >30: 0.587	25-27.49: 0.699 27.50-30: 1.970 >30: 2.518
HDL		0.006	0.039		Men <40 Women <50 0.779	
Triglyceride		6.55e-04			≥150: 0.405	
Blood Pressure		Systolic: 0.011	Systolic: 0.018	130/85 (***)	130/85 (***)	
Family History of Diabetes		0.498	0.481	0.506	0.570	0.728 (**)
Smoker				0.547		0.855 (**)
Alcohol habit				0.427		
Waist (cm)	Men 94-102 Women 80-88 0.86 Men ≥ 102 Women ≥ 88 1.35	0.0273			Men ≥ 102 Women ≥ 88 0.223	
Height (cm)		0.033				
HOMA-IR (75th) (*)					0.870	

TABLE 4.3: Risk scores predictors, intercepts and β coefficients. (*) Refers to any hipoglycemia. (**) The original model foresees two more categories not available in the study dataset. (***) The model calculates high blood pressure as an analysis of Systolic and Diastolic blood pressure or anti-hypertensive medication prescription, only one of the predictors is used.

A method to estimate the optimal threshold for the re-calibration of the model has been proposed by balancing the ratio of the False Negative (FN) and False Positive (FP)(Riley *et al.*, 2016).

One interesting model for T2DM detection is the MOSAIC model (Sambo *et al.*, 2015) which is open source and available for research [4]. This model is based on a Bayesian Network to impute missing parameters (Figure 4.3). The MOSAIC model was built to be applicable in different contexts and the performances are comparable to the FINDRISK score in scenarios where clinical data is not available. This model shows an acceptable predictive value when clinical information is available for cholesterol and fasting glucose (Sambo *et al.*, 2015). In the present study, missing variables will be imputed using MOSAIC model. Information about missing data rates and imputation performance is available in Results Chapter and in Apendix A.

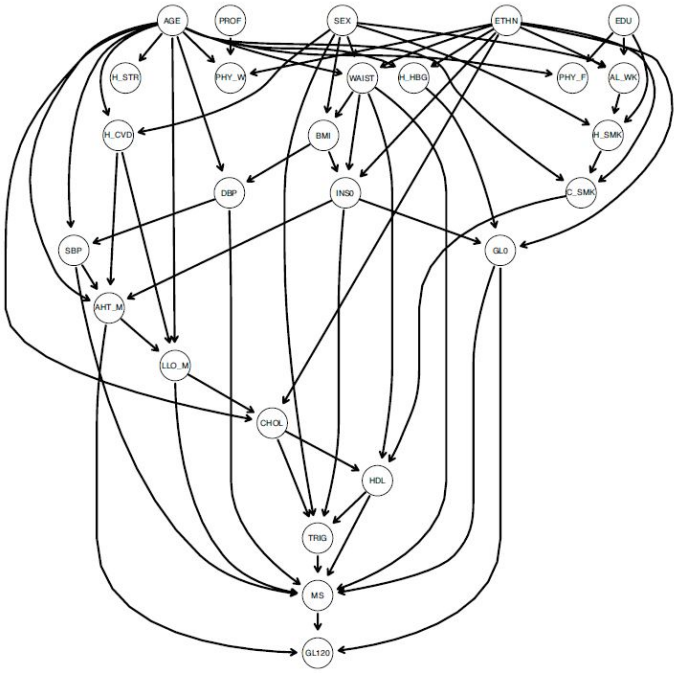


FIGURE 4.3: Cross-sectional Bayesian Network to estimate probability of the 2h-Oral Glucose Tolerance Test (2h-OGTT) outcome(GL120)

4.3.1 Assessment of risk scores performance

The performance of a risk model is assessed by Discrimination and Calibration measurements (Collins *et al.*, 2016):

- **Discrimination** is the ability of the risk prediction model to differentiate between patients who will be diagnosed with diabetes during the observation period from those who will not. Discrimination is quantified by calculating the area under the receiver operating characteristic curve statistic, where a value of 1 represents perfect discrimination. Nonetheless, although several models

[4] https://github.com/sambofra/bnstruct (Last Access 22 Feb 2017)

omit indicating their values, Sensitiviti(S), Specificity (S), Positive Predictive Value (PPV) and Negative Predictive Value (NPV) are also indicators to take into account on risk scores for predicting and detecting T2DM.

- **Calibration** refers to how closely the risk score outcome agrees with observed outcome. Calibration of the risk score can be assessed by plotting observed proportions vs. predicted probabilities; where a 45º line denotes perfect calibration. Calibration is quantified by the Hosmer-Lemershow test for the observed and expected events. The p-value can be calculated as the right hand tail probability of the corresponding χ^2 distribution for the Hosmer-Lemershow statistic. A *p-value* $\leq$0.01 indicates poor fit.

4.4 Clinical scenarios for T2DM risk assessment

The expected impact of a the study is to improve the characterization T2DM onset and target population at risk of developing T2DM in the future or which has already an undiagnosed T2DM. Given as input the available variables in a electronic health record (Ontology) for a given patient or a given population, the models can estimate the probability of being at high risk, and for MOSAIC model find out the most probable value of the variables not provided as inputs (Sambo *et al.*, 2015).

The MOSAIC model allow to enter the evidence on a subset of variables for a patient and to estimate the most probable value for the unspecified variables. This will allow not only to estimate an unspecified variable which is of interest for a better characterization of an individual or of a population, but also to use the final set of variables (both estimated and available) as input for a given risk calculator (the FINDRISK score (Lindstrom *et al.*, 2003), for instance, or some newly developed predictive models).

This means that this model can be described as the first step of an intelligent risk calculator, whose intelligence consists in the determination of missing variables to provide an estimation of the risk of developing or having T2DM. This opens the way to implement two different clinical scenarios (use cases) into the screening and risk stratification:

1. Estimate missing variables given available variables measurable with a general practitioner's visit and laboratory tests in the Electronic Health Record towards **risk stratification**.

2. Estimate the 2h-Oral Glucose Tolerance Test (2h-OGTT) glucose range given all other available variables (**supporting a diabetologist to decide whether this test is needed**).

4.4.1 Scenario 1: Risk stratification

In this case, the input data is coming from the health information system of a healthcare institution or agency. The input data are demographics variables and, when available, some other variables measurable with a general practitioner's visit and a blood test. The output will be a picture (through, say, a pie chart) of the distribution of the population most at risk of being T2DM and pre-diabetic. This output can represent a support to the following decisions:

Case 1, health care agency with limited availability of EHRs. Let's suppose that the information available to the healthcare agency is limited to demographics variables (gender, age, etc.), because the health information system is still not integrated in this settings: before asking to the hospital or to the primary care institution to provide them with phenotype and metabolic information of their served population, this system could be used to better stratify this request and narrow it only to the population which actually has the highest probability of being at risk (or, better said, by excluding the population with the lowest probability).

Case 2, health care agency with full availability of EHRs. In this case the input data for the system will be all the variables usually available in a "normal" citizen's clinical history record. In this case the output provided will be used to determine the subgroups most at risk of being T2DM or pre-diabetic; another output could be the determination of other meta-variables like being a smoker, having hyper cholesterol or not having an optimal lifestyle. In this case the tool could support decisions related to the strategic plans that an agency performs, before conducting screening campaigns and better estimate, e.g. which are the needed 2h-OGTT test, fasting glucose blood tests, screening visits foreseen for a period of interest (year, month).

Case 3, health insurance company. In this case, the system tool can be used to support the company in assessing the risk of health care expenses among a targeted group (a served company or group of individuals) and better develop routine activities such as finance forecasts, screening activities and health promotion campaigns better tailored and personalized to their clients.

4.4.2 Scenario 2: Supporting 2h-OGTT decision

In this case, the tool would have as input the EHR of a patient and the main output is to have an estimation of the 2-hours OGTT glucose range, given all other available variables. Thanks to this, the tool can support the decision of recommending or not an OGTT, with evident benefits in terms of health outcomes and cost savings.

4.4.3 Recommendations based on expected risk

All the cases should include the probability according to best performing risk normalized score

- Low risk: The risk to develop the disease in 5 years is 1%.
- Middle risk: The risk to develop the disease in 5 years is 16,5%
- High risk: The risk to develop the disease in 5 years is 33%
- Very high risk: The risk to develop the disease in 5 years is 50%

According to American Diabetes Association (ADA, 2016), screening for T2DM should be done through an informal assessment of risk factors to guide clinicians on the decision of further standard diagnostic tests, such as HbA1C. At least one annual monitoring is suggested for suspected pre-diabetic stages.

The strongest evidence on the effect of lifestyle interventions for the delay and prevention of T2DM comes from the Diabetes Prevention Program (DPP) (Lindström *et al.*, 2006), which demonstrated a significant reduction of T2DM incidence in 3 years. This study was grounded on a goal-based intervention on weight loss and moderate physical activity. Nutrition is also important for reducing the risk of developing T2DM, and data suggests that whole grain meals could help in this goal (Montonen *et al.*, 2003; Ley *et al.*, 2014).

Pharmacological interventions including metformin, α-glucosidase inhbitors, GLP-1 antagonists have shown to decrease T2DM incidence for pre-diabetic subjects. Finally, self-management and patient empowerment through education and support may be appropriate for maintaining healthy habits and behaviors that may lead to delay or even prevent the development of T2DM.

All in all, an endocrinology clinical doctor or a general practitioner should supervise each case carefully and make a decision based on the surrounding conditions, context and expertise.

With this information, subject-centered care based on T2DM risk estimation is based on 9 possible clinical recommendations:

1. Order 2h Oral Glucose Tolerance Test for this subject.
2. Order a HbA1C test for this subject.
3. Refer this subject to endocrinologist.
4. Refer this patient to General Practitioner.
5. Start pharmacological treatment.
6. Prescribe Physical Activity habits.
7. Prescribe Dietary habits.
8. Counseling and promotion of Physical Activity habits.
9. Counseling and promotion of Healthy Dietary habits.

Depending on the estimated risk and patient available data, clinical professionals from the endocrinology department will have to select none or any of the aforementioned recommendations.

4.4.4 Assessment of usability

The collection of usability-related and usefulness-related data serves the purpose of estimating the future acceptance of an eHealth system by its targeted users. For the purpose of assessing usability and usefulness of a software tool it is assumed that the perceived usability and the perceived usefulness of an eHealth System will influence the final acceptance of the system when put in routine clinical practice (Figure 4.4).

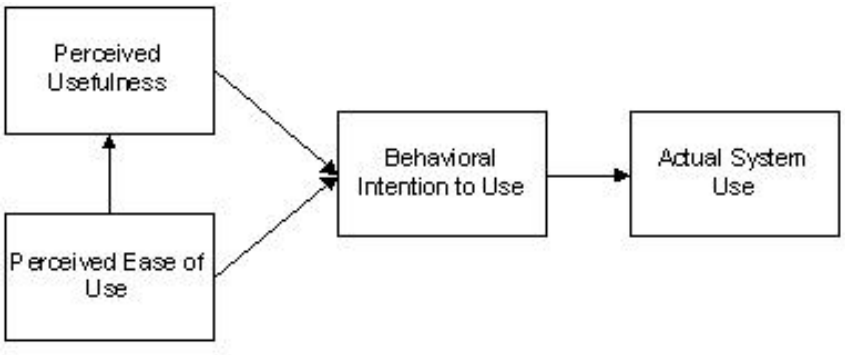


FIGURE 4.4: Technology Acceptance Model

These two concepts (Venkatesh *et al.*, 2003) are defined as:

- Usefulness, which describes the extent to which a technical system contains the necessary functions in order to achieve the specific goals which to achieve it had been created.

- Usability, which describes the extent, to which a user is able to successfully implement the functions of the technical system in order to achieve these specific goals.

The definition of the Technology Acceptance Model (TAM) (Venkatesh *et al.*, 2000), shown in Figure 4.4, also foresees the following concepts:

- Perceived usefulness: The perceived usefulness usually is measured by custom items using an adaptation of Davis scale (Davis, 1983), with some extra items ad-hoc crafted for an eHealth System. Three important constructs are related to the overall concept of perceived usefulness: Quality of Life, Adoption of Good Practices, Perceived Usefulness in its narrow sense, and Organizational Task Adequacy. Organizational Task Adequacy is oriented on the concept of task adequacy, a subordinate construct of the usability of dialogue systems - ISO 9241-10. By adding the attribute "organizational", we mean to indicate that this construct does not only measure how well will an eHealth system fit into the work process of each individual doctor but into the entire process structure of the care process as defined in the respective hospital.

- Perceived Ease of Use. With "Ease of Use", refer to the Usability of the System. According to ISO 9241, usability comprises three core measures: efficiency, effectiveness, and satisfaction. Efficiency and effectiveness are supposed to be objective measures that can be measured during a controlled user test. They are practically impossible to measure on the basis of the data collected in a test, however, in the case of acceptance it is the subjective impression of the users that counts, rather than objective usability data. This is the dimension "satisfaction", which we are going to measure by using the AttrakDiff questionnaire. The AttrakDiff is a semantic differential scale for the assessment of user experience. User experience is a concept that is broader than pure "satisfaction" and is well fit for the purpose, as it also includes social consequences of the system use, such as proposed in newer models of user acceptance.

The expected acceptance will be analyzed as:

1. User Satisfaction using **"AttrakDiff"** questionnaire the instrument consists of 4 subordinate constructs, all of which are computed separately: pragmatic quality, the two hedonic qualities stimulation and identification, and attractiveness. We will use the mean score of each construct, based on a pairwise comparison of concepts. In order to have an acceptable result, the confidence interval of the collected measures' mean value should not touch the scale's middle score. This applies to each of the three dimensions: pragmatic quality (PQ), hedonic quality (HQS/HQI) and attractiveness (ATT).

2. The perceived usability measure this instrument consists on the **System Usability Scale.** It intends to measure to which extent the users feel able to make the eHealth System do what it is supposed to be doing. The use of objective usability data from the tests is very limited, as users cannot be controlled all the time. The use of these items is of rather exploratory nature and useful to improve the interfaces and tools in the development iterations. Cronbach's alpha should be calculated to determine whether any item needs to be deleted. The arithmetic mean value of all the remaining items will be calculated. In order to have an acceptable result, the confidence interval of the collected measures' mean value should not touch the scale's middle score.

4.5 Software infrastructure to execute T2DM risk scores

4.5.1 The business context

The business context of system to support the execution of T2DM risk models in clinical settings is based on the stakeholders and the offered services (functionalities).

Client stakeholders are the abstract roles who may use the system functionality from different perspectives and for different purposes (viewpoints). Client stakeholders considered so far are:

- **End users**: non-technical end users such as health care professionals, health care managers, patients and citizens.

- **Service Providers**: Companies or organizations interested on the commercial exploitation of the system, individual tools or integration in a disease management existing system to provide services to the end users.

- **Clinical researchers**: Developers or scientist interested on the development of new technologies or models for data and process mining. Developers include domains as back-end, front-end and data base management.

System functionalities are classified under modules, which are the entities that provide services and operate within the system. These modules may offer services to be consumed among themselves or directly by stakeholders.

- **The Data Storage module** is in charge of providing a warehouse for all the data within the system. From a conceptual point of view, the data model is unique for all the system, containing Electronic Health Records (EHRs) and other kind of data (logistic and administrative), however, each site will have its own data base infrastructure, as an instantiation of the common data model. The purpose is each hospital to work only with their own data but having a common data model and data dictionary.

- **The Model Host module** is the core of the system. It is in charge of managing the client requests (user interactions), running the risk scores and querying the data warehouses. It gathers into an Application Server the tools (models) that will run the algorithms over data from hospitals and provides the services for managing them from the client side. Model Host module will also contain other modules to provide horizontal services including security features, tracking and system management.

- **The Plugin module** is the part of the system that hosts the user interfaces as defined in the use scenarios (Sections 4.4.1 and 4.4.2). These user interfaces are web pages formatted for the intended use of each type of user and scenario. The integration with existing disease management systems is articulated wrapping the interfaces within plugins, tailored for each integration case.

The list of stakeholders above is highly generalized however it provides a good division of the roles and services that build up the system architecture. Nonetheless, after having identified relevant stakeholders, we would like to look into their major expectations, i.e. expected benefits that adopting the system would provide them with. Utmost important is to point out the expectations that have been addressed in the SA for each stakeholder group:

- End users: Health professionals, including managers and policy makers and medical researchers mainly concerned with public health affairs. Good development environments and friendly interfaces will lead to better quality software, and will attract professionals to use the tools. Efficient communication with service providers and among service providers will result in services that better meet end user requirements. High quality runtime support will guarantee e.g. fault-tolerance and timely response, as well as a fair privacy protection (https). The SA will address these expectations by supporting the platform stakeholders in the three central areas shown in Figure 3 1.

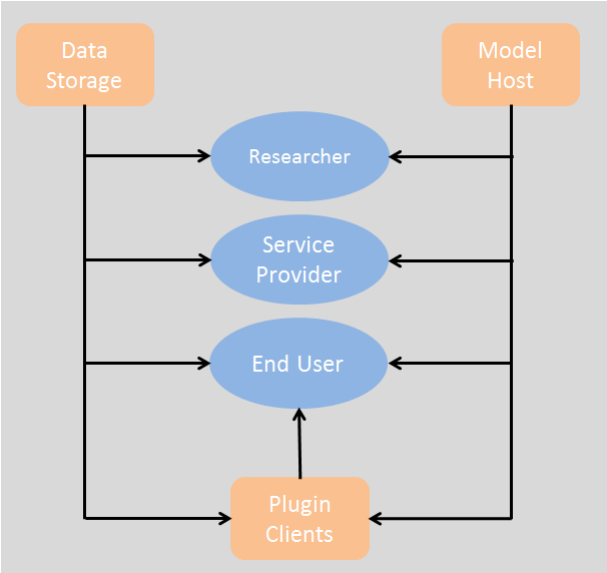


FIGURE 4.5: Business context showing the relationships among services and stakeholders

- Service providers: are concerned about the commercial exploitation of the system. They need to hold an effective communication with their end users, fluent interaction with the runtime environment to explore potential integrations and furthermore access to talented researchers for development and personalization of their services. The description provided allows service providers to deploy, monitor and otherwise manage their services as provided to their end users.

- Researchers: are mainly concerned with good development environments, a knowledgeable community of developers, and access to a market for their software and algorithms. The system should support researchers as a major stakeholder and allows them to participate in the system improvement together with service providers and end users. Two main domains of research are found within this viewpoint: Data/Process mining research and Software research. The first type is focused on the development of new algorithms and models to perform stratification, variable association and workflow analysis. The second type aims to improve the software quality of the services, interfaces and data base managers.

4.5.2 Quality Metrics

The defined business context should provide a mapping among Use Cases that evidence stakeholders' expectations and the system in terms of reference services. ISO/IEEE 1471 methodology has been used to perform the mapping between the system architecture and the stakeholders expectations. The requirements represented by the study scenarios (and their technical specifications) provide a set of measurable constraints on the architecture to measure its conformance. Emerging from the stakeholders perspectives and the scenarios, three categories have been defined:

- Category 1: A system for running algorithms on demand with a specific running environment **regardless patient health records or additional data than a set of defined parameters**.

- Category 2: A system for running algorithms on demand with a specific running environment **which needs patient health records and additional data form a huge amount of variable parameters**.

- Category 3: A system for running algorithms on demand in the client side with a specific running environment **which needs raw and pre-processed data**.

Extracted from this three categories there can be identified a set of common pathways. According to ISO/IEEE 1471 a second level of abstraction is needed to draw the common concepts or processes within these tools, naming them the Reference Success Criteria indicators, depicted in Table 4.4.

RSC ID	**RSC description**
RSC#1	Supporting rich human computer interaction
RSC#2	Supporting intelligent context management and hardware abstraction
RSC#3	Enabling system driven interaction
RSC#4	Supporting continuity of care
RSC#5	Supporting end user security and privacy
RSC#6	Supporting installation, configuration and management of system components
RSC#7	Supporting remote/local operation
RSC#8	Supporting data granted access to perform CRUD[5] operations
RSC#9	Supporting interfacing with existing information systems
RSC#10	Supporting service providers to offer system services
RSC#11	Allowing users to easily find and acquire system tools
RSC#12	Supporting exploitation of different business models
RSC#13	Capturing and utilizing user feedback
RSC#14	Supporting rapid development of new models
RSC#15	Model-based development of services through integrated model transformation tools
RSC#16	Supporting online elicitation of requirements and collection of runtime feedback from users of risk score services
RSC#17	Supporting advanced search, reuse and sharing of service components and resources
RSC#18	Supporting customization of system services

TABLE 4.4: System Reference Success Criteria

4.5.3 Business environment

The business environment is defined by mapping the business context into real deployable components. To do so it is mandatory to define the structural aspects of the components model in UML. System modules for Data Storage, Model Host and

Plugin Clients of Figure 4.5 are mapped into high-level components that will implement the services (low-level definition). For the sake of simplicity here is provided one top-level component for each of the components on the overall network. Upon this definition, an interface is the link between two parties of each service, and it needs a provider and a consumer. It is important to highlight that although there are services casted by the system modules, in further applications (concrete architectures) each service could be provided by a separate business entity, deployed and operated independently, with the only requirement of being compliant with the interoperable service protocol.

Major information concepts that are used to qualify the provided services are described as means of UML descriptors (Figure 4.6). These concepts are mainly related to the offered services, and how the architecture handles and processes those services in general, helping to contextualize their use.

The information regarding each service is stored in the platform in form of a Service Description (WSDL). As proposed by (Segagni *et al.*, 2011) this description contains references to the implementation of the service on a XML basis. In the system architecture a service is constituted by one or many components that belong to a specific system module. A service might also be constituted by other services (a composed service) and in this case the service description will have a reference to the other services' descriptions.

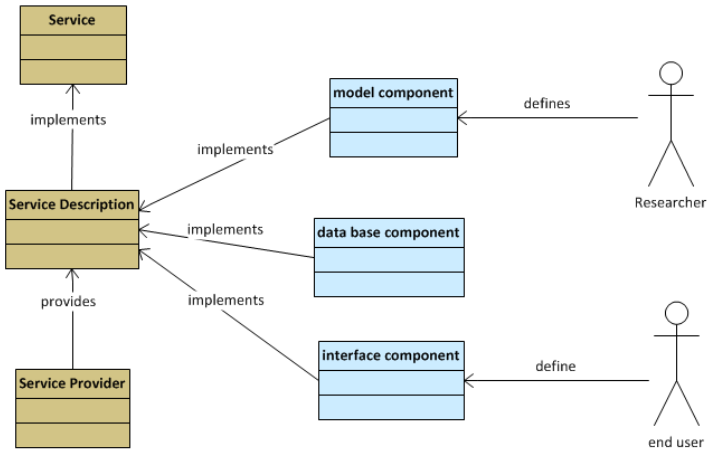


FIGURE 4.6: UML Component and service descriptors

4.5.4 Service collaboration pattern

In the following section we will look into the details of each of the three system modules and their components. The goal is to identify the high-level reference services that are provided at two different levels, as shown in Table 4.5.

The system architecture has been designed as a Service Oriented Architecture (SOA) (Newcomer *et al.*, 2004), in which the different components from different

Types of Services	Description
Module-to-Module (B2B)	Services that are provided by one system module to other module (s) of a different type (e.g. a web service provider from the models requires data from a remote database service provider).
Module-to-Client (B2C)	Services that are provided by a module to a to client stakeholders (e.g. a web service provider provides remote execution of a model).

TABLE 4.5: Type of service collaboration pattern among system components

modules access to the whole functionality of the system that may be located in different physical allocations (one or several servers) through a set of Web Services. These components interact with each other over Internet in a modality prescribed by its description using SOAP messages, conveyed using HTTPS with an XML serialization in conjunction with other Web-related standards.

Services are listed depending on their nature and purposes, for this reason they have been gathered in several different components which pertain to each of the three modules.

4.5.5 Functional view

According to IEEE 42010, the functional view describes the capabilities, structure, responsibilities and specifications of the system components and how they interact among them. The functional view categorizes the services into three types: application, interoperability and system services.

Figure 4.7 depicts the system architecture and the functional relationships among the modules. The three main modules are connected by System Services (red dotted-square), Interoperability Services (blue dotted-square) and Application Services (green dotted-square).

The term application encompasses all services that directly support one of the five main scenarios (Use Cases). Interoperability services covers all functionalities that can be reused by any component within the system (e.g. can be included in ETL processes that prepare input data for the algorithms). Finally, system services cover all logic operations (including functional logic and infrastructure) that is common to multiple scenarios.

From left to right in Figure 4.7 the schema shows the data storage module, based on the I2B2 technology. This module is composed by several single data entities each of them gathering data from different sources: Hospitalization, Laboratory Tests, Outpatient Services, etc. From a logical point of view, the Data Storage Module is a unique conceptual part from the overall system structured according to the Common Ontology, however, from a physical point of view, each warehouse and data set is an isolated virtual machine located elsewhere and reachable though the Internet. The connection of the Data Storage Module and the System Application Server

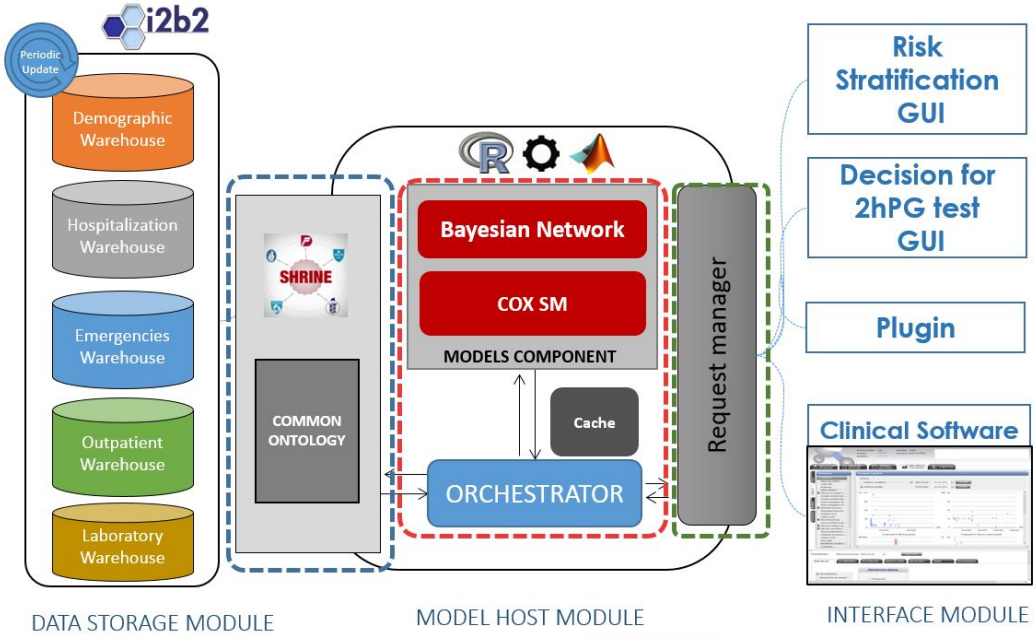


FIGURE 4.7: System architecture functional view

is performed by the SHRINE service layer (green shaded), a set of interoperability services that allow performing federated queries to the whole data storage warehouses, regardless its physical location. This configuration permits researchers and clinicians to choose which the target population of the queries is, extend and reduce the focus and, in case of availability, connect new warehouses/datasets compliant with the ontology.

These interoperability services are used by the system services (blue square) to perform ETL and storage operations. These services are gathered within the Orchestrator Component and the Models Component. The first component is in charge of executing the predefined work flows for each tool and model. As mentioned before, the requirements for providing the input parameters and running specific algorithms involve many software components within the system that must be able to work in a distributed and controlled way. This kind of complex process execution is solved by using process orchestration which assumes that the processes are able to exchange data to execute processes in a distributed way (Martinez-Millana *et al.*, 2015).

This means that the orchestrated processes are independent and can communicate with each other, in what we know as the "defined execution flow" (work flow). Using this model, it is possible to connect and disconnect components and modules dynamically. Components can provide their functionalities and services can consume them without the necessity of knowing the concrete architecture of the deployed service. This facilitates the creation of more independent and flexible services, fault tolerant and able to deal with different kind of components and different configurations. Nonetheless, the orchestration paradigm requires the use of a common interchange language that allows components to understand the purpose of the services available in the system architecture.

Rather than using syntactical models with common message formats (which ensures that the services are able to read the services data format) we plan to enhance the service descriptors using semantics. This is because the syntactical data format limits the capacity of services to understand the data content. This limitation can affect the independence of the services, which must be prepared to read data in all the possible formats, and requires a strict subscription process to make sure that the component information is sent to the listening services. This is a problem in distributed architectures where the modules are exposed to a very aggressive and stressful environment and the in which the inner component configuration is continuously changing (e.g.: a revision of a Bayesian Network module to improve the model classification outcomes). The use of semantics as an alternative to syntactical models provides ad-vantages for the services in the understanding of data structures and enables an intelligent subscription process to provide dynamism to the deployment of components.

Figure 4.8 shows a picture of the system orchestrator. The core of the component is a message dispatcher engine (Choreographer) and a data base that contains the services that are registered (declared) within it. The services may be connected to the core locally, when the services are allocated in the same computer of the choreographer core (e.g.: Model Services), or remotely by using a TCP protocol service wrapper (e.g.: ETL services). An ontology reasoner is connected to the orchestrator. This reasoner is able to infer knowledge from registered services where semantic information is available through the core. Connected to the choreographer there is the Orchestrator service, which allows the use of work flows to describe processes in a graphical way. Using this approach, the end users and researchers can create their own solutions by using graphical tools to describe processes as a tidy and causal combination of other services.

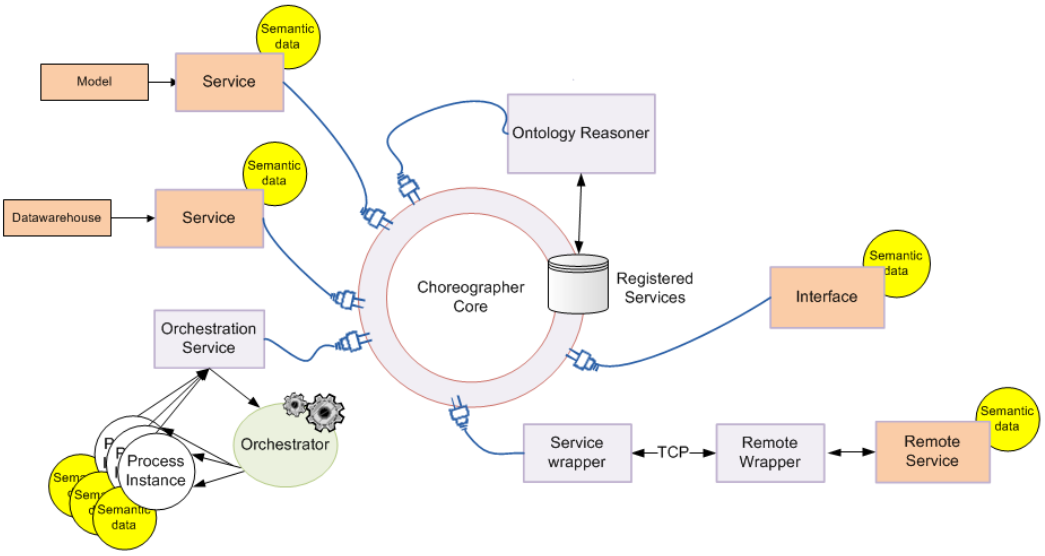


FIGURE 4.8: System Orchestrator functional schema

The reasoners are software pieces that allow performing semantic search in the ontology. The semantical description of the services must provide a reliable shot of the functionalities and actions they provide so the core can detect automatically

which services are available. However, the key point of this component is the orchestrator. It enables the execution of a predefined work flow. A work flow is the formalization of a process as matters of an automation process. Work flows may be described graphically to be interpreted by humans, but in this case we are referring to formalization in the manner it can be executed by a work flow engine. The second component within the Models Host Module is the models grid which contains the system services to run the algorithms (screening and pre-diabetes classification) using the required running environment (R and matlab). Finally, the application services are located in the right part of the schema. They are the services consumed by the system standalone tools and disease management systems that may integrate these functionalities.

The choreographer in the orchestrator component dispatches messages among the modules using a specific XML message protocol called XMSG. This protocol is based on the combination of FIPA (Site, 2017) and SOAP (Newcomer *et al.*, 2004) protocols. The classic FIPA protocol, defined for Multi Agent Systems communication allows sharing knowledge using several protocols. XMSG is based on FIPA headers to route and characterize the messages. The content in XMSG is based in the SOAP protocol. SOAP is a well-known and widely used protocol to perform service calls. The XMSG protocol allows broad and multi-cast as well as P2P message calls using custom symbols in the destiny address.

An example of XMSG message is shown in Figure 4.9. The message is sent from the ServiceA, whose logical address is java.ServiceA, to the NightService, whose logical address is java.Choreographer.NightService. Both sender and receiver information and the type of message sent (request, inform, event...) are defined in the message header. Following it, in the content part of the message, the call to the specific method of the service is defined. In this example, the method invoked is Execute Process, which needs the lights addresses as input parameter.

The communication among the services is made via peer to peer communications. This means that each service must know in each moment what services and methods are alive and the kind of information that they are able to deliver. The use of semantic information in the service registry allows the services to know the semantic meaning of the data sent and the type of the expected answer that may be returned. This casuistic of information, formally defined in an ontology, can be used by the reasoner to allow the services to use semantically-driven searches to improve search accuracy by understanding the contextual meaning of service terms. Each service will provide information using semantic languages as Ontology Web Language (OWL)[6]. This information is used by the reasoner to discover services that matches, not only the syntactical information but also the meaning, with the high level query of the user or the service. This system allows retrieving context based search results that make the system more dynamic and powerful, helping computers to perform automated information gathering and research. In our example, the light address must be specified by the sender. Nevertheless, the use of the ontology reasoner can help in this problem. The ontology reasoner can be used to infer what are the lights installed in the system at the current moment allowing services to be auto-configurable.

[6]OWL Specifications Web Site https://www.w3.org/TR/owl2-overview/ (Last Access 23/02/2017)

```
<xmsg>
    <message-id>a6dfaea7-5ceb-4e70-8010-ee09529ea05a</message-id>
    <sender>java.ServiceA</sender>
    <receiver>Java.Choreographer.NightService</receiver>
    <protocol>request</protocol>
    <language>XMSG</language>
    <content>
    <Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">
        <Body>
            <ExecuteProcess xmlns="http://tempuri.org">
                <ligth1 type="xs:string">
                    Java.Choreographer.ligth.1
                </light1>
                <ligth2 type="xs:string">
                    Java.Choreographer.ligth.2
                </light2>
                <ligth3 type="xs:string">
                    Java.Choreographer.ligth.3
                </light3>
                <ligth4 type="xs:string">
                    Java.Choreographer.ligth.4
                </light4>
            </ExecuteProcess>
        </Body>
    </Envelope>
    </content>
</xmsg>
```

FIGURE 4.9: XMSG sample

4.5.6 Model Host Component

Figure 4.10 shows the central part of the system architecture which hosts the engines to execute risk scores. This section describes which services are provided from the components shown in the model host of Figure 4.7. As the components wrapping the models are being tested and developed currently, the services they offer are listed as matters of the functionalities they provide, not entering to list the type and name of input/output parameters. With respect of the Orchestrator and horizontal components (security and track), the services are being already defined.

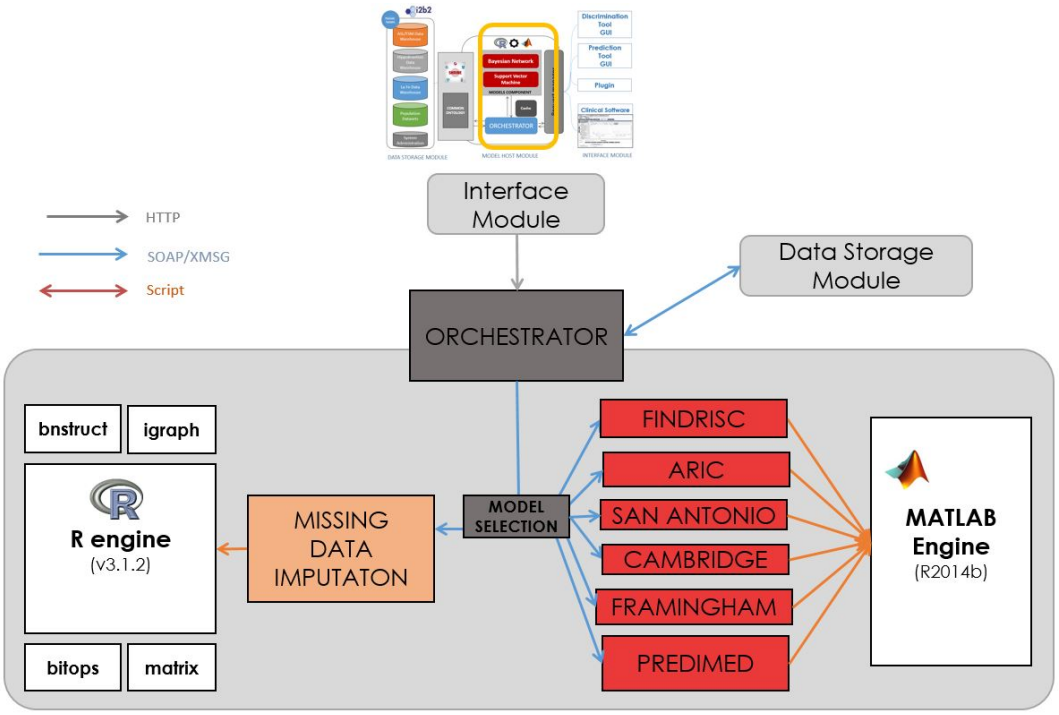


FIGURE 4.10: Execution of the risk scores equations using mathematical engines

4.5.7 Security Component

The Security Component is in charge of providing secure horizontal features for all the services (most of them in the Service Provider layer as it is the service access point from outside to the system) based on four dimensions:

- Authentication: It must be possible for the service provider to ascertain the identity of the service requester.
- Authorization: The service provider must be able to determine whether the requester has the appropriate rights to invoke the service.
- Message Confidentiality: Message contents must only be visible to the intended recipient.
- Message Integrity: It must be possible to guarantee that a message has not been altered or tampered with in transport between the service consumer and the service provider.

Authentication is supported through the use of client-side x.509 certificate, credentials (username and password) for each professional end-user, and a Security Assertion Markup Language (SAML) certificate. All web services are offered in a Secure Socket Layer (SSL), the system implements this security feature encrypting the information exchanged between the end points and thus the message confidentiality is guaranteed. Only certified connections will be accepted by this component. Each end user will be provided by a set of credentials (Username and password), and they will be mandatory to log into the web applications and furthermore to authenticate the connection.

4.5.8 Track Component

Every system must provide a record tack of the executed services, their results, timestamps and other audit information. The track component is in charge of recording the trace of all the activities that take place during the performance of the system (in both test and deployment phases). The records must be standardized (or even normalized), understandable and be ready to be parsed and mined. Therefore this component will record all the interaction events among the modules and components (Figure 4.11). As the user interaction deserves special attention and opens a brand new study field, all the interactions in the Interface Module will be recorded in a special format and place in a basic txt file (to make easy the access for the information). A file named *LOGusername.txt* will be automatically generated upon first launch of a user. A main class controls the interaction events during a session and track them in that file. Each interaction event will be written in a line with the following format:

<Timestamp>, <Module>, <control>, <free text>

- Time stamp: dd/mm/yyyy hh:mm:ss.
- Module: the module (view or form) where the patient is currently in.
- Control: the control used: button, label, picture, graph, chart etc.
- Free Text: free text that indicates the interaction or notes for the usability expert.

Regarding the programming track, it is important to state the six masking levels of priority to perform the trace that the Log4J library provides:

- FATAL: Used for critic system messages. After this message the program/process will be aborted.
- ERROR: Used for minor errors while the application is running. These errors will not cause an application break, but may affect to the performance. For instance, if a configuration parameter is not set and the application must load the default values.
- WARN: These messages do not affect to the application performance but should be highlighted to avoid programming mistakes.
- INFO: Information messages in "verbose" mode.
- DEBUG: Used to report messages at debug level. This level should not be used in a release.
- TRACE: Used to show messages with a more detailed level tan DEBUG.

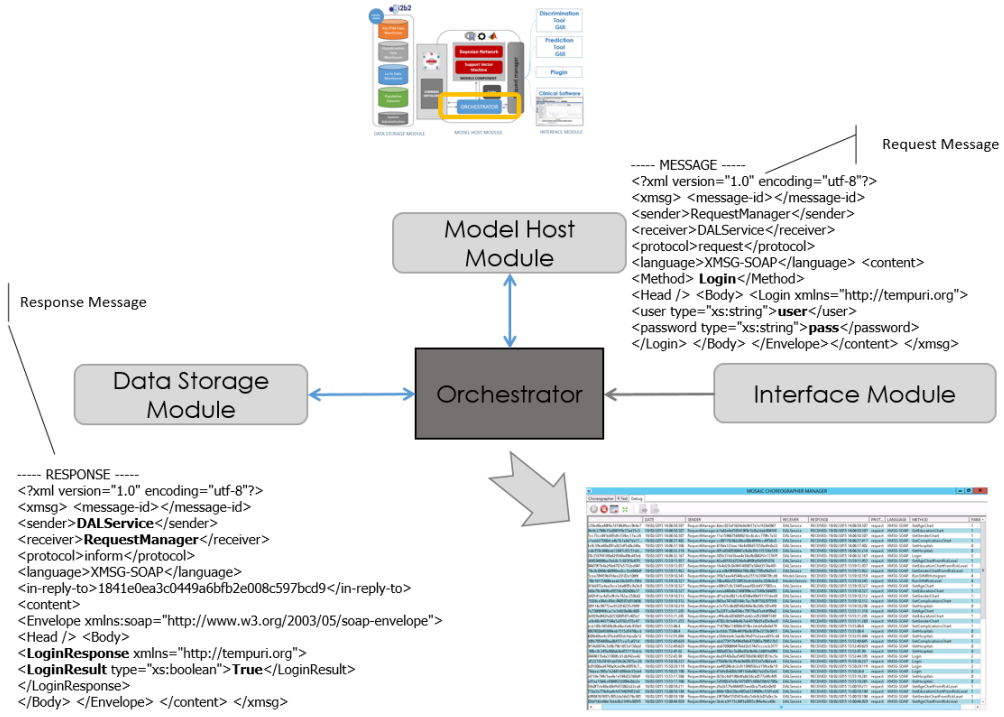


FIGURE 4.11: Track of the system service messages

These messages are broad-casted to one or more destinations, called *appenders*. There are a wide range of appenders, however anyone can create their custom appenders, adding new information as the time stamp, running variables and extra information. Beyond the functionalities provided by third parties libraries, such Log4J, Log4Net and Google Log (Glog), the system offers two services to perform the programming track and user interactions.

The assessment of information technology systems is a complex operation in which individual components are assessed both separately and in the context of their intended use. Part of this evaluation must be conducted while they are communicating or interacting with other system components, and possibly at the same time. Determining their fitness for use involves both an evaluation of whether they meet technical criteria while the components are running, and a technical verification, in which component inputs and outputs are analysed and compared to the designed flows.

For each model, a scenario for the Best and Worst case has been defined according to the specification and behaviour of the operation. For the prediction model the Best Case is the execution for a single patient, and the worst is the execution for the highest available population, which is 8080. In the case of the detection model, it can be executed only for a single patient, so the worst case is when the model does not have any input variable (i.e.: it has to estimate the 21 missing parameters, and the best case when it has 20 input parameters and only has to estimate one).

The track component will be recording the information on the selected KPIs to perform the technical evaluation.

4.6 Study design

The study design was based on a single center randomized study investigating the acceptance and usability of the tools for the prediction and detection of T2DM, and comparing the effect of the risk score evaluation during nine consecutive weeks based on retrospective Electronic Health Records (EHRs). The biomedical research ethics committee of the Hospital La Fe approved on January 2015 the formal request of data and the study design. No further considerations were given by this committee.

All de-identified patients who fulfilled inclusion and exclusion criteria entered into a first evaluation. The first evaluation consisted on a collection of clinical data into the EHRs to check availability and completeness. Patients were divided into two groups: cases for patients having a T2DM ICD-9 coding and controls for patients whom had not that coding.

At the end of the evaluation patients were gathered and risk scores were evaluated among them. The system was evaluated in the Endocrinology Service of Hospital La Fe during a continuous period of three months involving endocrinologists and head of service whom used the tool during 2 hours per session. Three training sessions were planned with the participants prior to the utilization (Figure 4.12). Participants were blindly randomized and assigned were assigned with patients from both groups.

	DECEMBER/2015					JANUARY/2016				FEBRUARY/2016				MARCH/2016	
	W1	W2	W3	W4	W5	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2
Training		1h	1h	1h											
Risk Stratification							2h	2h	2h						
Decide if 2hPG is needed							2h	2h	2h	2h	2h	2h	2h	2h	2h
Questionnaires				2h									2h	2h	2h

FIGURE 4.12: Gantt chart of the evaluation study

The study plan consisted of three stages:

1. Training sessions: Three group sessions with the clinical professionals who signed the informed consent for introducing into the tools and learning the actions to visualize data and execute the risk models (blue shaded grids).

2. Evaluation of risk scores and clinical evaluation: Evaluation of the tools during sessions of 2 hours long during 9 weeks. The two clinical scenarios defined in Chapter 4.4 were assessed in parallel (green and yellow shaded grids)

3. Questionnaires: Acquisition of profiling information for clinical participants and the usability and acceptance questionnaire fulfillment was performed in sessions specially devoted to this end (red shaded grids)

4.6.1 Safety and withdrawals

No safety issues arose during the study design and preparation. Data included in the study is retrospective and de-identified so any unexpected finding did not affect current patient treatment nor health-status perception at the present. Clinical professionals using the tools were recruited according to their role into the Endocrinology Service in Hospital la Fe, and after signing the informed consent to participate in the study were included into the acceptance and usability sub-study. No participant withdrew or discontinued during the study.

4.6.2 Study Population

The criteria for the diagnose of T2DM was based on the American Diabetes Association guidelines ADA, 2016 for fasting blood glucose, HbA1C and random blood glucose cut-off points.

Computational models were be executed on a dataset containing patients with retrospective EHRs for the study variables with the following criteria:

INCLUSION CRITERIA

- Age>45 years
- To have a confirmed diagnose of T2DM within years 2014 – 2015 AND to have observed variables with a time stamp of one AND/OR five years before the T2DM diagnose.
- To have no confirmed diagnose of T2DM within years 2014 – 2015 AND to have observed variables with a time stamp of one AND/OR five years before the T2DM diagnose.

Subjects with EHRs available for a timespan of 5 years were included into the prediction study, whereas subjects with EHRs available for one year were included into the detection study.

EXCLUSION CIRTERIA

- T2DM originated by other reasons than ageing and lifestyle (pancreatic cancer, transplant, immuno-suppression).
- Type 1 Diabetes Mellitus.
- Patients with steroids prescription.
- No data availability 5 years or 1 year before T2DM onset (for cases).
- Use of anti-diabetic preventive medication (for controls).

With respect to the medical professional using the tool, they will be profiled according to their age, years of professional experience, gender, role in the endocrinology service and ICT Literacy and compliance level.

4.6.3 Outcome measurements

1. **Retrospective validation**

   Comparison of the risk score VS the clinical history of the patient (T2DM/not T2DM). The applied comparison criteria will be done through a revision of every case with the clinical history of every subject. To do so the following data will be gathered:

   (a) RISK SCORES Comparison: Discrimination and Calibration performance of the predictive risk score calculated for every selected case using FINDRISK, ARIC, Framingham, PREDIMED, Cambridge and San Antonio without calibration.

   (b) CURRENT DIAGNOSE COMPARISON: The proportion of individuals with an HbA1c of 6.0–6.4% or a FPG of 110-126mg/dL, thereby being eligible for a preventative intervention and the proportion of subjects in high risk for MOSAIC Detection model.

   Missing data will be imputed using the MOSAIC variable imputation Bayesian Network (Sambo *et al.*, 2015).

2. **Intervention analysis** For those cases of diagnosed T2DM identified through screening we will assess the recommendations given by the medical staff: Proportion of cases identified as High Risk of T2DM (HRT2DM) or T2DM who would be offered a preventive lifestyle intervention.

   (a) Proportion of cases identified as HRT2DM or T2DM who would be offered a GP appointment.

   (b) Proportion of cases identified as HRT2DM or T2DM who would be offered a Specialist appointment.

   (c) Proportion of cases identified as HRT2DM or T2DM who would be offered a therapeutic intervention.

   (d) Proportion of cases identified as HRT2DM or T2DM who would be offered a HbA1C test.

   (e) Proportion of cases identified as HRT2DM or T2DM who would be offered a 2h-OGTT.

   (f) Proportion of cases identified as HRT2DM or T2DM who would be offered life style intervention

3. **Usability of the tool and suitability of the screening process**

   (a) Quantitative measurements for the pragmatic quality, hedonic quality and attractiveness.

   (b) Global System Usability Scale (Brooke, 1996).

4. **Technical throughput of the tool**

   (a) Key performance indicators for best and worst scenario.

4.6.4 End-points

1. Clinical Outcomes

    - High risk of T2DM cases or T2DM cases.
    - A cut-off point for high risk of T2DM cases that would not require blood testing.
    - AUC of ROC curve of the prediction and detection risk tool on the study dataset.

2. Usability Outcomes

    - User experience of the tool according to pragmatic quality, hedonic quality and attractiveness.
    - Usability scale and dependency factors.

3. Technical Outcomes

    - Computational Load (Memory footprint in the server).
    - Response delay to service request (s).
    - Access time to main DB/Caché (ms).
    - Time usage span (s).
    - Maximum response Delay.

Chapter 5

Results

5.1 Clinical Outcomes

Spanish incidence rate of diabetes (Valdés *et al.*, 2007) is 10.8 cases/1000 person-years. For a total population of 215,000 subjects observed during 6 years, the expected overall T2DM population would be 13,932 subjects for Hospital La Fe Health Department. After the extraction process, the study dataset was comprised by of 10,730 subjects (77,03%) with data of regular laboratory tests and hospital visits from 2008 to 2015 and a confirmed diagnose code of T2DM (ICD9:250.0).

The incidence rate (cases/persons-year) is 1,532 patients (2014-2015). The minimum sample size Kenny, 1987 adjusted to a 15% of expected loss and a 95% confidence level is n_T =160. A total of 159 subjects who met the inclusion and exclusion criteria defined in Section 4.6.2 were collected.

Figure 5.1 shows the results of the prediction model for the 159 patients collected, where each line represents the prediction for a single patient. As we can see, there are some patients for which the predictions are meaningful; while for other patients, the predictions were wrong (as all these patients have currently developed T2DM).

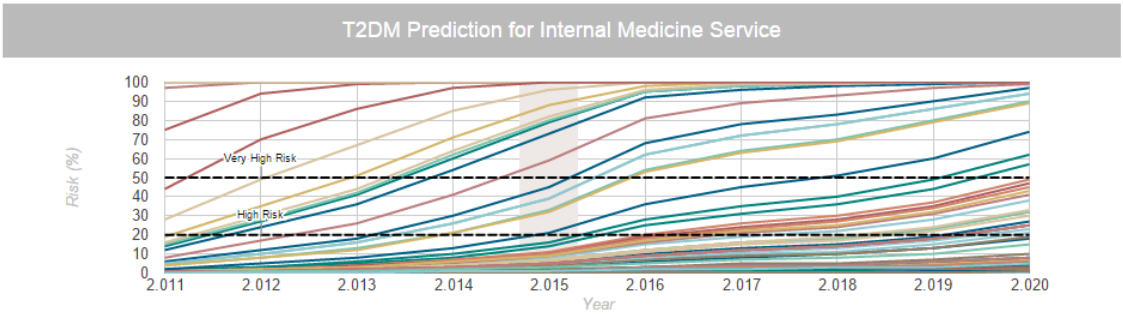


FIGURE 5.1: Prevalence adjusted T2DM risk Score prediction outcome for 10 years

After conducting an individual analysis of the hospital records for each patient in the cohort, that were supposed to be diagnosed during years 2014-2015, clinicians found out that the ICD codifications for T2DM were erroneous and the majority

of the patients have developed diabetes several years before, ranging from 1988 to 2008.

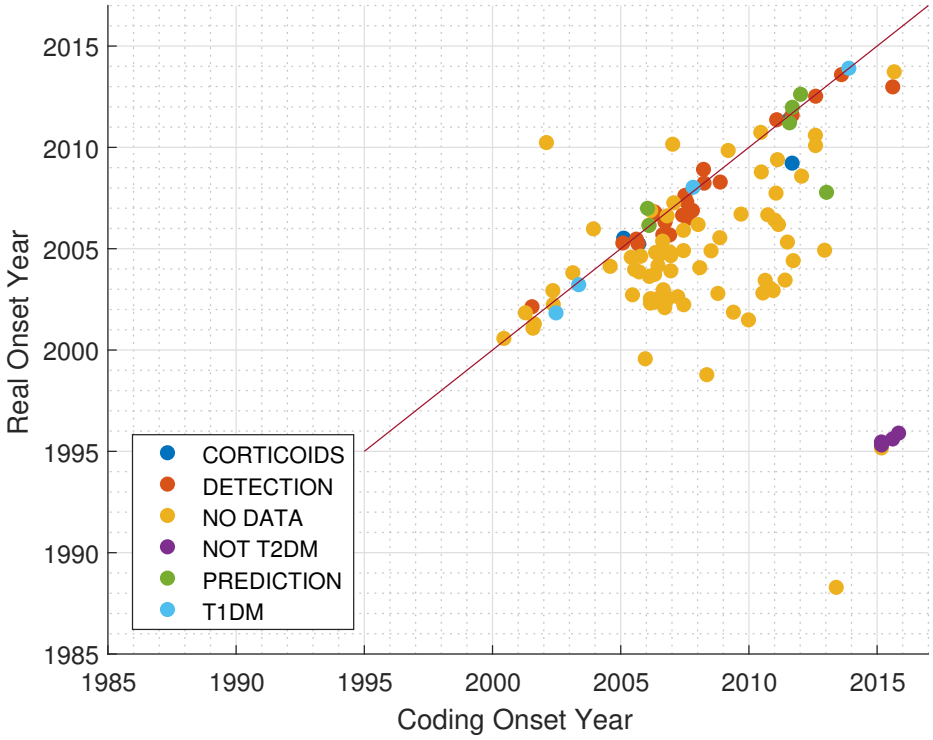


FIGURE 5.2: Inconsistencies in the T2DM onset date and coding time stamp in the Electronic Health Record

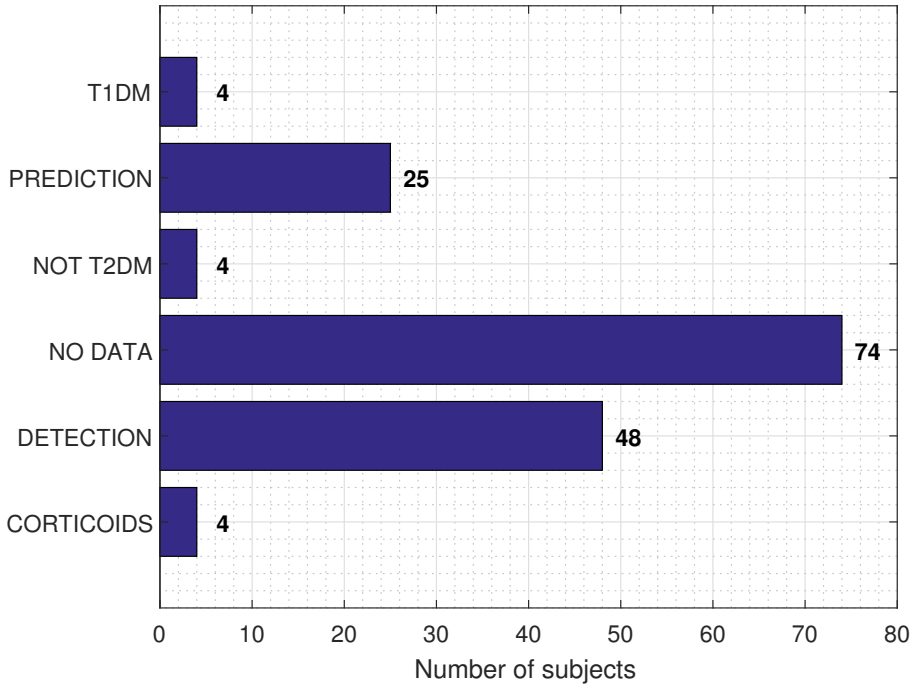


FIGURE 5.3: Results of the clinical case by case revision

After analyzing the 159 subjects, n=73 patients were eligible to be included in the pilot and were recorded into the system database. The reason of having such a low incidence rate is due to a lack of quality in the disease coding of the electronic medical record (ICD9). Case-by-case revision patients were selected according to established criteria for WHO T2DM Diagnose (Table 2.1). The main limitation was to discover patients that had developed diabetes and had clinical records of at least 5 years before the real disease onset (Figure 5.2).

This challenge was very difficult to overcome because information systems in la Fe are split into two platforms. Hospital la Fe was moved from 2010 to 2013 from its former location in the west side of Valencia City, to a brand new facility in the south side of the city.

Laboratory systems and electronic health records were duplicated to keep the compatibility of legacy systems and therefore currently there are two data repositories in which the clinical software management relies. This fact was a key issue when discovering T2DM patients and the availability of records that could fulfill the criteria defined in the study.

Finally n_P=25 subjects for prediction analysis and n_D=48 subjects for the detection analysis met the inclusion criteria and were included in the database and used to conduct clinical, usability and technical analysis (Figure 5.3).

5.1.1 Evaluation of prediction risk scores for T2DM performance

A total of n_P=25 subjects (13 controls and 12 cases of T2DM) were recorded to assess both discrimination and calibration. Independence of variables was assessed by a two-sided t-Student test at IC=95%. All variables were independently distributed to the class, except *Diastolic Blood Pressure*, which is not identified as a predictor in any of the state of the are T2DM risk scores. Table A.1 and Figures A.2 and A.3 in Appendix A show the distribution of numeric and categorical variables.

After the execution of the selected risk scores on the study sample, the distribution of the outcome was analyzed with respect to the class (Figure 5.4). Only Framingham ($p = 0.005$), San Antonio ($p = 0.018$) and Findrisc ($p = 0.048$) achieve a significant difference for the observed outcome. Table 5.1 depicts discrimination and calibration performance for the suggested cut-off points and for the re-calculated cut-off points (those which maximize the AUC ROC). Calibration slope for suggested and re-calculated cut-off points are shown in Figure 5.5.

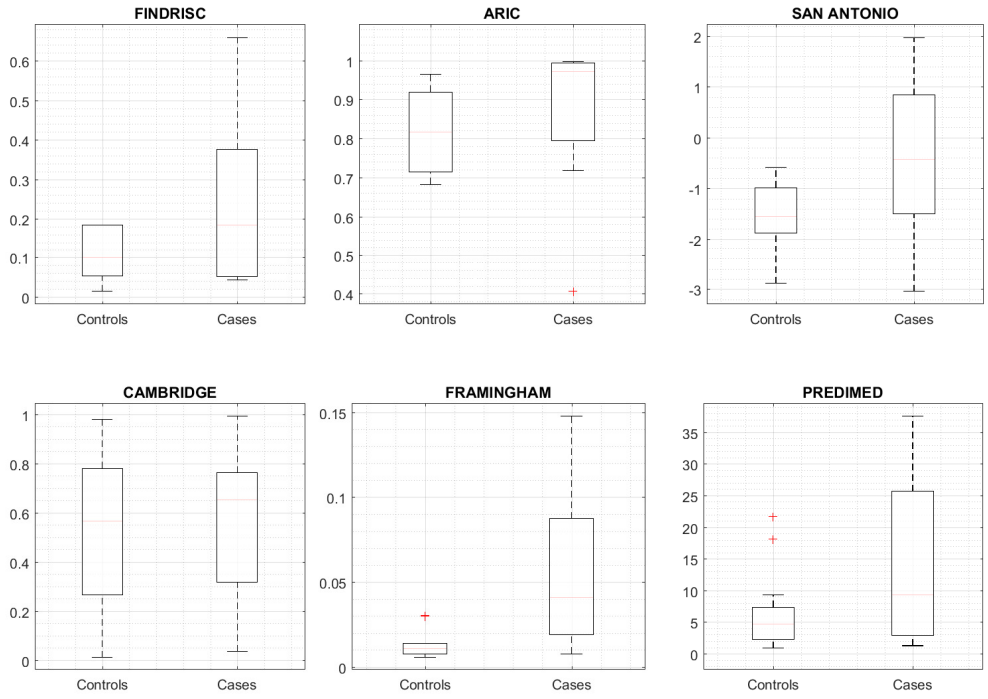


FIGURE 5.4: Distribution of the risk scores outcome with respect to the class. Only Findrisc, Framingham and San Antonio scores show a statistically significant difference ($p < 0.05$).

PERFORMANCE DESCRIPTORS FOR THE SUGGESTED CUT-OFF POINTS

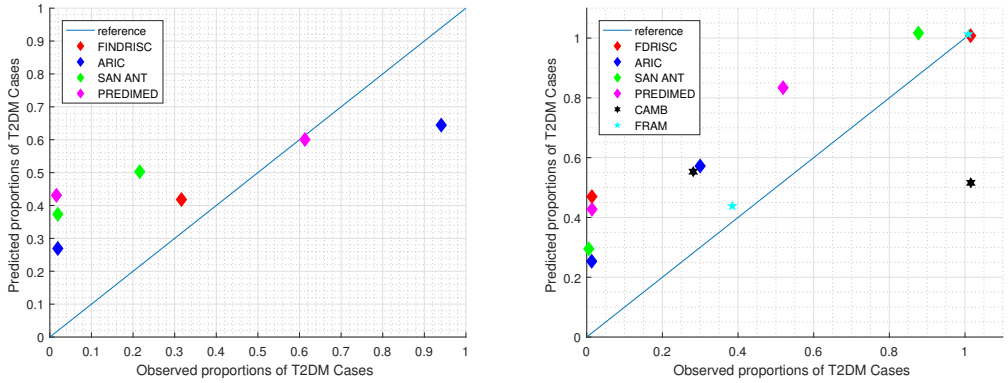
	S	Sp	PPV	NPV	AUC	Cut-off	HL score	*p value*
FINDRISC	0.54	0.67	0.64	0.57	0.69	0.1800 (*)	0.004	0.002
ARIC	0.69	0.5	0.6	0.6	0.73	0.8210 (*)	0.11	0.054
SAN ANT	0.46	1	1	0.63	0.76	0.0650	0.019	0.001
PREDIMED	0.38	0.91	0.83	0.57	0.66	19 (*)	0.511	**0.225**

PERFORMANCE DESCRIPTORS FOR RECALCULATED CUT-OFF POINTS

	S	Sp	PPV	NPV	AUC	Cut_off	HL Score	*p value*
FINDRISC	0.38	1	1	0.6	0.69	0.28211	0.003	0.0426
ARIC	0.53	1	1	0.67	0.73	0.97343	0.271	**0.397**
SAN ANT	0.61	1	1	0.71	0.76	-0.475	0.018	**0.107**
PREDIMED	0.54	0.91	0.83	0.57	0.66	16.297	0.049	**0.175**
CAMBR.	0.76	0.33	0.55	0.57	0.53	0.345	0.288	**0.408**
FRAM.	0.85	0.83	0.84	0.83	0.875	0.034	<<0.01	0.02

TABLE 5.1: Discrimination and calibration performance of the risk models for suggested and recalculated cut-off points

According to these outcomes the **Framingham risk score** model performs better to predict subjects developing T2DM using a threshold = **0.034**.



(A) Calibration plot for suggested cut-off points (B) Calibration plot for re-calculated cut-off points

FIGURE 5.5: Calibration performance of risk scores with suggested and calculated cut-off points. Cambridge and Framingham scores do not suggest a cut-off points so the performance descriptors are not applicable in the top table.

5.1.2 Support on detecting T2DM

Detection of pre-T2DM and T2DM cases are done using the Bayesian Network from MOSAIC model (Sambo *et al.*, 2015) on the n_D=48 population (23 cases and 25 controls). This model calculates the probability of having a 2h-OGTT test low (<140 mg/dL), medium (140-199 mg/dL) or high (>200mg/dL).

Table A.2 and Figure A.5 and Figure A.6 in Appendix A show the distribution of numeric and categorical variables.

MOSAIC model provides a probability for each range (for instance: 80% LOW, 15% MEDIUM and 5% HIGH), so the purpose is to find which are the thresholds for these probabilities that perform a better classification among subject that developed T2DM and those who not. Figure 5.7 shows the distributions for each probability compared by cases and controls. Only the probability of a high 2h-OGTT result has a significant difference ($p << .01$).

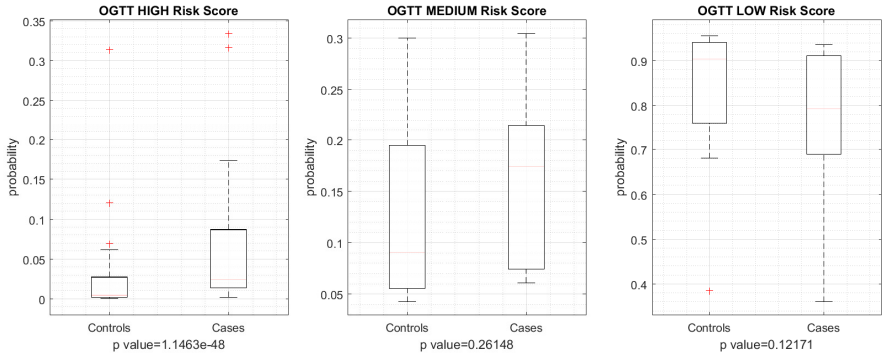


FIGURE 5.6: Comparative distribution of the probability of having a high, medium and low 2h-OGTT result among cases and controls

Criteria	Cut-off	S	Sp	PPV	NPV	AUC
High 2h-OGTT Risk	0.021	0.74	0.6	0.63	0.71	0.69
FG (ADA)	126	0.17	1	1	0.57	0.74
A1C (ADA)	6.5	0.00	1	-	0.59	0.81
FG calibrated	111	0.61	0.8	0.74	0.69	0.74
A1C calibrated	6.2	0.36	0.875	0.67	0.67	0.81

TABLE 5.2: Detection of T2DM stages performance

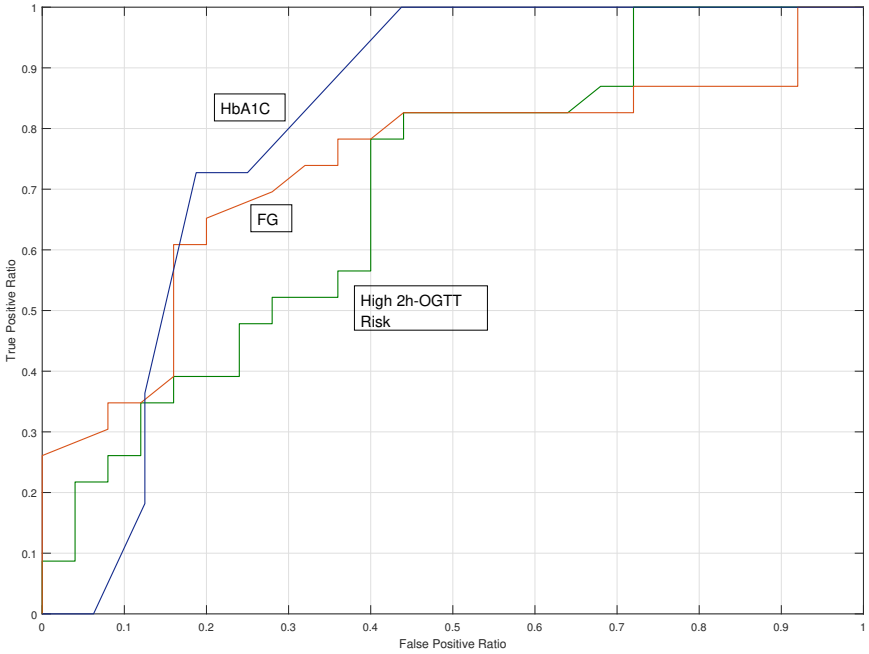


FIGURE 5.7: Comparison of the C statistic (Area Under the ROC curve) for the 2h-OGTT High risk probability and the two gold standard procedures (FG and A1C)

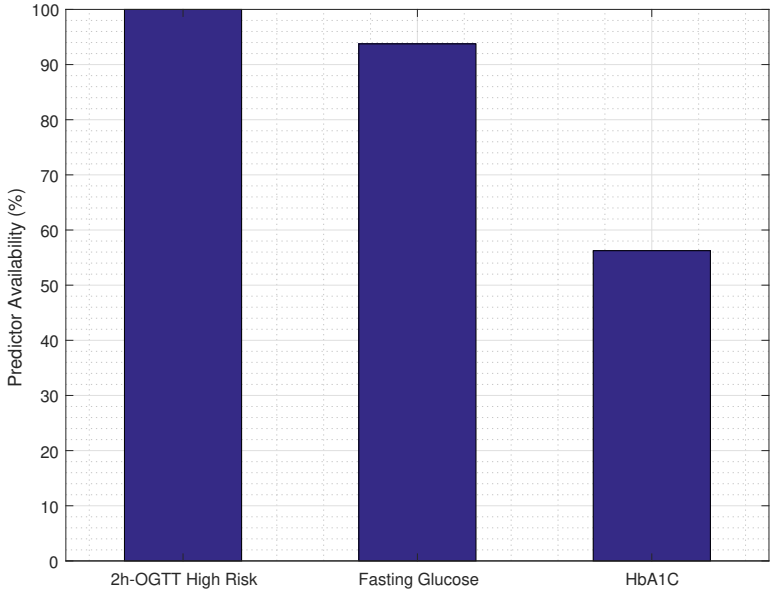


FIGURE 5.8: Data availability (%) for each of the assessed predictors. 2h-OGTT High Risk is on 100 %, whereas Fasting Glucose is 94% and A1C is 56%.

5.1.3 Clinical advice for high risk subjects

Analysis on the recommendation with respect to the predicted risk (expected class) and the real subject situation (observed class).

The system calculated a risk for each subject and presented through the we interface to the clinician, who had to make an assessment based on the real available clinical data, the inferred parameters with MOSAIC model, and the estimated risk. Based on this assessment, the clinician had the option of selecting one of the nine recommendations suggested by the American Diabetes Association (ADA, 2016).

Table 5.3 shows the selected recommendations classified for the estimated risk (Risk Outcome column) and the real subject situation (Real Situation), which depicts patients who developed T2DM afterwards or not.

19 out of 23 cases (82,6%) identified as high risk (true positives), were assigned to pursue an HbA1C analysis which is the most specific test for discriminating the diagnose, whereas only 13 out of 23 (56,2%) of the real cases were assigned to do the test. This reflects that the endocrinology service in Hospital La Fe is aligned with the recommendations from the American Diabetes Association (ADA, 2016) in the way they need a value for HbA1C to discriminate whether the subject is pre-T2DM or has developed T2DM.

	Risk Outcome		**Real Situation**	
Recommendation	**LOW RISK**	**HIGH RISK**	**NO T2DM**	**T2DM**
Order an 2h-OGTT for this patient.	4	6	3	5
Order an HbA1C test for this patient.	15	19	19	13
Refer to General Endocrinologist .	1	2	0	3
Refer to General Practitioner.	11	12	7	11
Start Pharmacological Treatment.	1	8	0	7
Start Dietary Activity hhabits.	5	12	3	9
Start Physical Activity hhabits.	6	11	4	8
Counseling about healthy lifestyle.	15	11	12	9
Counseling about diet, physical activity and weight control.	6	11	3	13

TABLE 5.3: Number of recommendations for each subject according to the risk outcome, compared to the real situation of the patient (blind for doctors)

5.1.4 Missing data influence on risk score outcome

A paramount issue is the level of reliability of the scores depending on the data availability. To this end, based on the data availability of the predictors (Appendix A), it is needed to find out the extent to which the missing parameters percentage is affecting the classification.

Prediction analysis

The percentage of missing data is not a factor related to the class (Figure 5.9), as the independence test does not rejects the null hypothesis. Figure 5.10 shows the distribution of missing parameters for each of the risk score outcomes. Missing data factor affects only to Framingham score (p= 0.049).

Figure 5.11 shows the missing parameters distribution (mean and standard deviation) for each of the categories in the confusion matrix: True Positives (TP), False Positives (FP), False Negatives (FN) and True Negatives (TN). On this analysis it is confirmed that the missing data is not affecting any of the classification performance indicators.

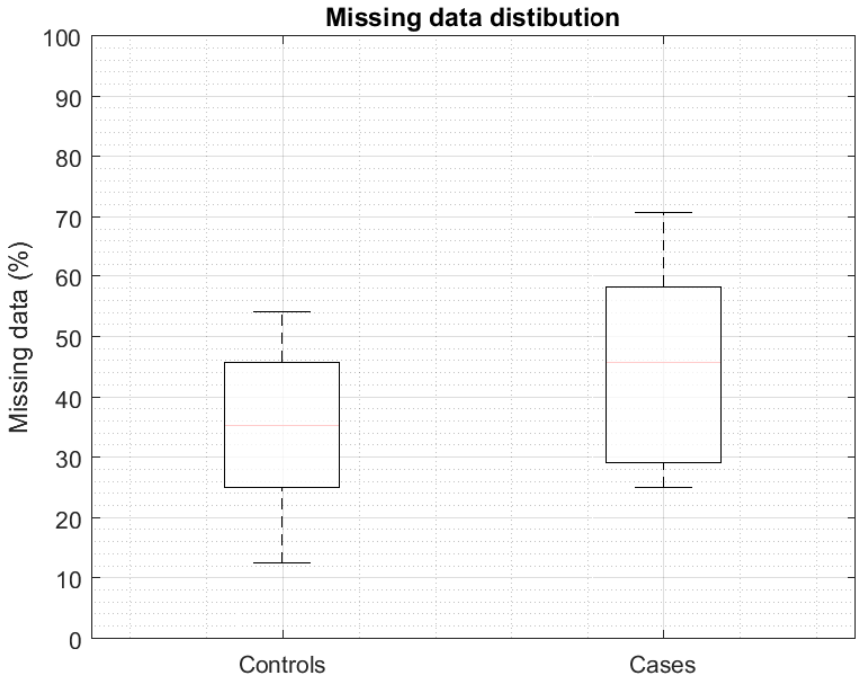


FIGURE 5.9: Distribution of the percentage of missing data among cases and controls for the prediction analysis. there is not statistical significant difference ($p > 0.05$)

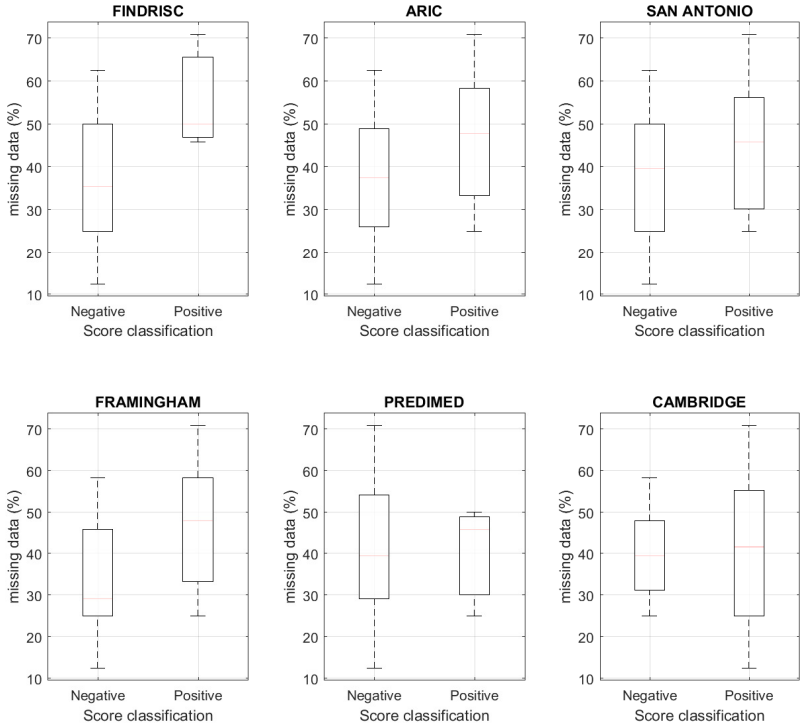


FIGURE 5.10: Distribution of the missing data with respect to the output risk score. Only Framingham shows a statistical significant difference with p=0.049

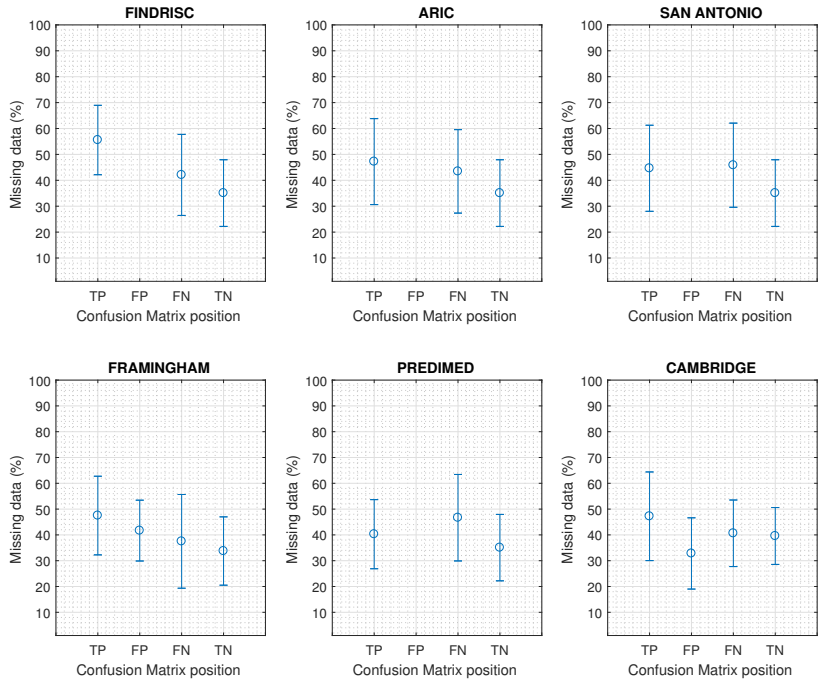


FIGURE 5.11: Distribution of the missing data for the categories in the confusion matrix

Detection analysis

As Figure 5.12 shows, the missing parameters factor is not defining the class for the detection of T2DM.

For the detection of T2DM missing parameters can be an issue, as the American Diabetes Association (ADA, 2016) defines diagnostic cut-off points for HbA1C, fasting glucose and 2h-OGTT and of these, the first and the third may not be recorded unless a doctor specifically ordered that test. Moreover the 2h-OGTT is less available than the HbA1c, as the former can be determined in a regular laboratory test and the first requires the patient to appoint for a 120 minutes duration test.

For the data set of this study, missing HbA1c accounts for the 54% of the cases, whereas missing fasting glucose accounts only for the 6% (Figure 5.8).The risk estimated for a high 2h-OGTT is available for all the patients, even though the classification under performs compared with HbA1C and fasting glucose.

Nevertheless, an analysis on how the missing parameters affected the classification was worth to know if the estimation of the 2h-OGTT high risk is performing better than the fasting glucose cut-off classification. Figure 5.13 shows the distribution of cases and control with respect to each of the three indicators. Fasting Glucose and HbA1C have a good classification of the T2DM patients.

Figure 5.14 shows the distribution of the High 2h-OGTT Risk estimation and the distribution of the Fasting Glucose with respect to the class for the subjects who had

not an HbA1c value in their Electronic Health Records. The two-sided t-Student test for fasting glucose distributions rejects the null hypothesis with a $p << 0.05$, whereas the null hypothesis is not rejected for the high 2h-OGTT risk.In this case, the AUC achieved by the fasting glucose indicator with a th=126 mg/dL is 77% and for the high 2h-OGTT risk is 55%.

This analysis confirms the results obtained in the detection model analysis in the way that the 2h-OGTT estimator does not perform a better classification when HbA1C or Fasting Glucose are available.

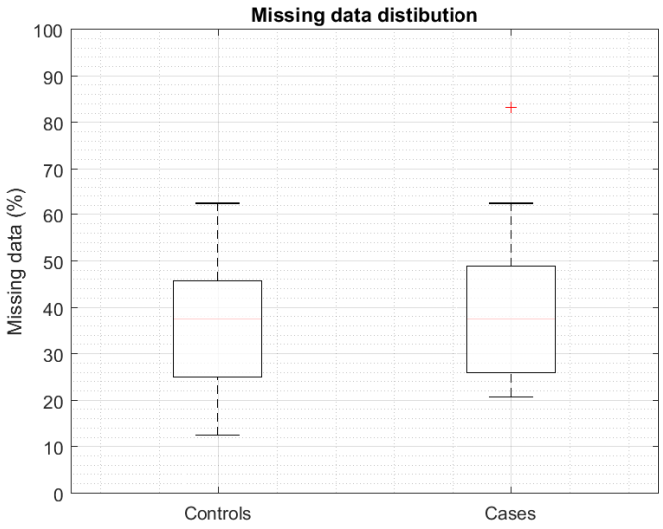


FIGURE 5.12: Distribution of the missing data to the class for detection. There is no statistical significant difference among groups $p > 0.05$

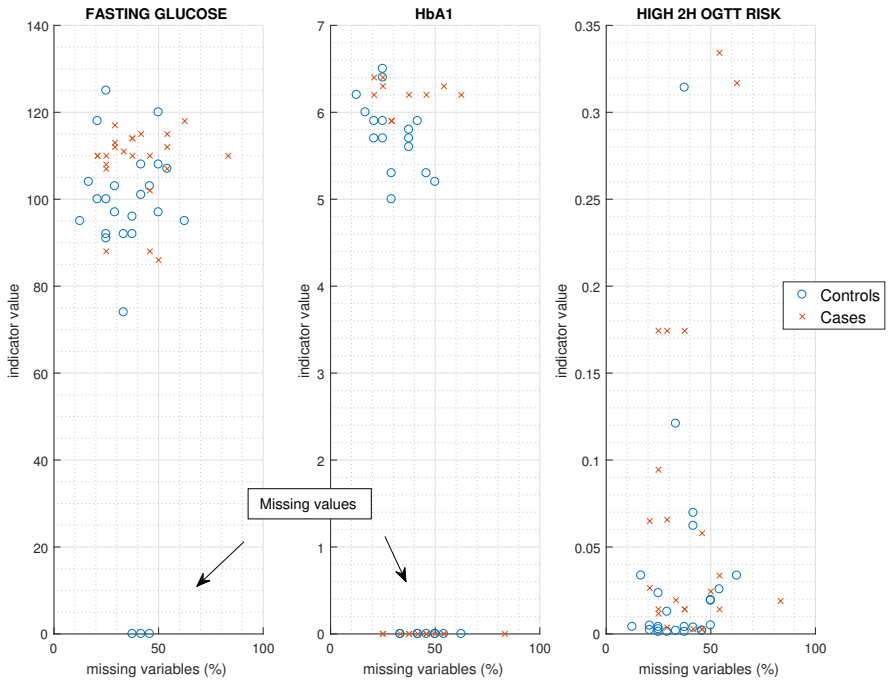


FIGURE 5.13: Distribution of the missing data rate with respect to the class and the indicator for discrimination

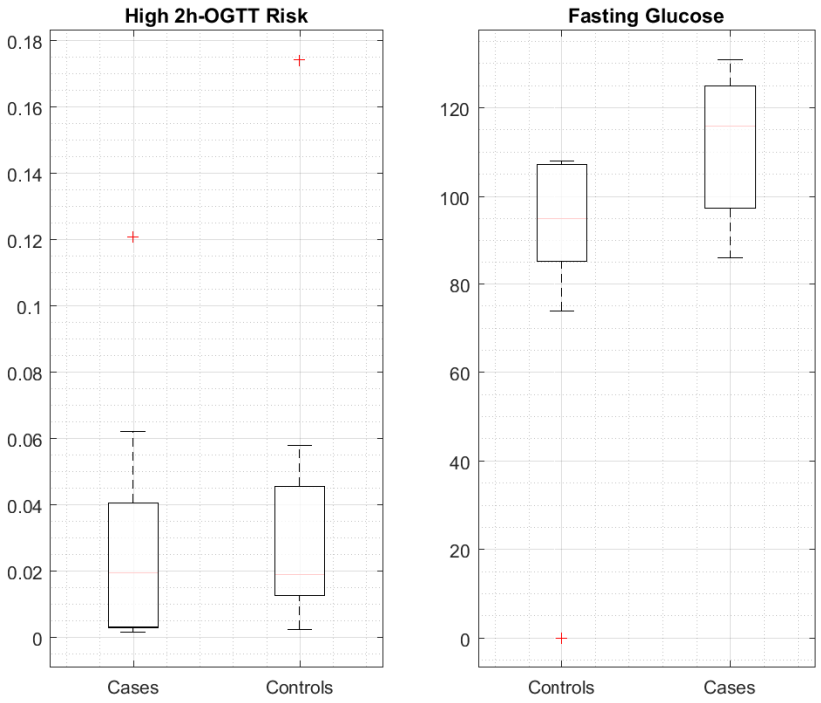


FIGURE 5.14: Distribution of the high 2h-OGTT Risk and Fasting Glucose values for subjects with no HbA1C data.

5.2 Usability Analysis

Usability analysis was arranged for the two clinical scenarios defined in Chapter 4.3. Clinical staff from Hospital La Fe (Table 5.4) used the system to identify risk sub-groups and analyze high-low risk subjects during two consecutive weeks (Table 5.5).

Gender		Male (2) /Female (6)
Age (Years)		42±13
Professional Experience (Years)		14±10
ICT Literacy (Self-reported)		High=3; Medium=3; Low=2;
Number of Patients Assisted	Overall	319.33±247.66
	T2DM Patients	127.44±75.22
	High risk of developing T2DM	48.00±33.79

TABLE 5.4: Clinicians evaluation the two scenarios

User satisfaction was measured with AttrakDiff questionnaire (Apendix B), which contains 4 subordinate constructs all of which are computed separately: pragmatic quality, the two hedonic qualities stimulation and identification, and attractiveness. The standard quality criteria consisted in having the confidence interval of the collected measures' mean value above the scale's middle ($Score \geq 3$).

With respect to the usability, the System Usability Scale (SUS) is a ten-item *Likert* scale questionnaire (Apendix B) created more than 30 years ago, it is the most used questionnaire to measure perceptions of usability, being also used as industry standard. SUS result can be interpreted by converting the average sum of each item-score to a percentile rank by normalization and comparing it to the reference curve provided by the SUS creators. The average SUS scores from about 500 studies are reported as a reference in Figure 5.15b and Figure 5.16b. A SUS score above a 68 would be considered above average and anything below 68 is below average. The graph below this figures shows how the percentile ranks associate with SUS scores and letter grades (from A to E).

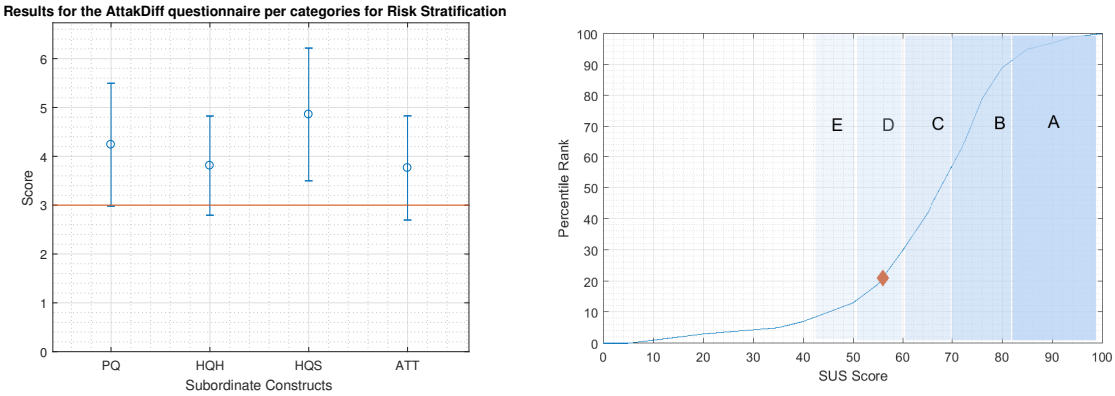
Results for the user experience and the usability are depicted in Figure 5.15 and Figure 5.16. For "Scenario 1: Risk Stratification" the AttrakDiff results show that the standard deviation for Hedonic QUality for Identification (HQH) and the Attractiveness (ATT) is below the reference score, which indicates that the system is not frustrating but has floor for improvement. For this scenario, the SUS score achieves

Indicator of Use	***mean***	***sd***	***min***	***max***
Number of users per day	2.5	16.43	1	4
Duration of sessions (min)	26.16	13.72	0.25	45.93
Number of patients evaluated per doctor	6.25	4.97	1	15
Number of patients evaluated per day	10.71	12.18	0	26
Number of sessions per doctor (user)	1.82	1.16	1	5

TABLE 5.5: Distribution of the evaluation sessions (number, duration, number of patients per day and per session)

a percentile rank of **58.30**, this value belongs to a percentile rank of 30%, which has higher perceived usability than 30% of all products tested to define the previous curve.

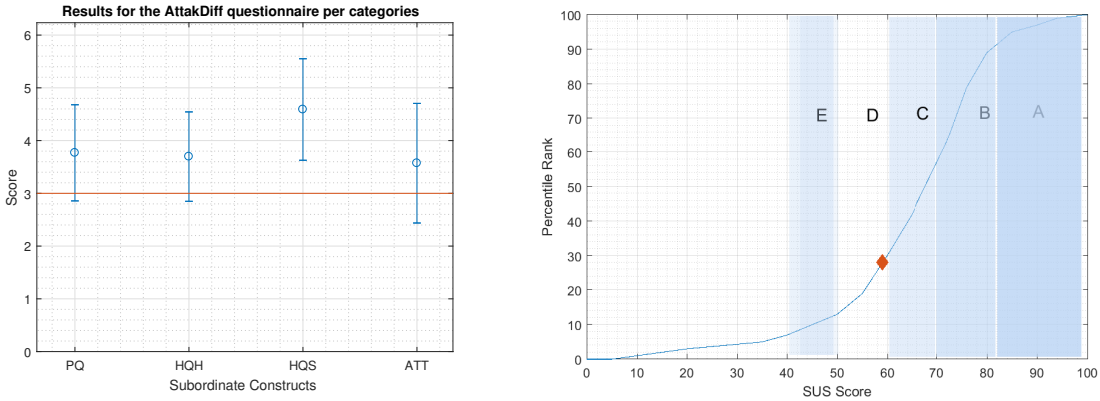
For "Scenario 2: Supporting 2h-OGTT Decision" the AttrakDiff results show that the standard deviation for Pragmatic Quality (PQ) and Hedonic Quality for Identification (HQH) are slightlhy on the cut-off score, whereas the Attractiveness (ATT) is deep below the reference score, which indicates that the system could be frustrating for the clinical users. For this scenario, the SUS score achieves a percentile rank of **68.30**, this value belongs to a percentile rank of 30%, which has higher perceived usability than 40% of all products tested to define the previous curve.



(A) AttrakDiff results per category

(B) SUS Score

FIGURE 5.15: Scenario 1: Risk stratification user experience and usability results



(A) AttrakDiff results per category

(B) SUS Score

FIGURE 5.16: Scenario 2: Supporting 2h-OGGT Test Decision user experience and usability results

Similar needs have been identified in the two scenarios. They are:

- To increase filtering and data-insights functionalities.
- To provide more visualization options: automatically build and send reports.

- To include more contextual information of the selected groups and risks.
- To provide aggregated and less complex information.

Specifically for Scenario 2, the most common identified issues were to avoid duplication of information and too long contents, while increasing the presence of "Help" boxes, contextual tags, and also to increase the clarity through top screen or pop up messages. Also in this case, users were asking for more filtering criteria and search criteria (e.g. searching patients by laboratory test results).

In Figure 5.15 and Figure 5.16 the results of the user experience while using the tools (Attrakdiff questionnaire) are satisfactory as they meet the quality criterion of having the confidence interval of the collected measures' mean value above the scale's middle score (value equal or higher than 3).

The tools used for Scenario 1 and Scenario 2 induce to perform an active risk stratification of T2DM in hospital settings which represent a disruption in terms of organizational and procedural aspects. If we want to introduce risk stratification tools for T2DM, we have to work together with the health care organizations that are available to implement this innovation, and this work consists of generating data warehouses specific for this purposes (upon a process of data pooling and quality checks) and in training the health care professionals that would be involved in this process.

The personnel involved in the evaluation provided feedback on how improving some specific information (e.g. BMI categories should meet the WHO recommendations; the decision making process for risk predictions should be based on the comparisons with what professionals have detected; recommendations should be in check boxes, so the clinician may select any/all of them depending the risk estimation and the imputed variables; other filtering options should be included, for instance Erectile Dysfunction).

Users' insights were gathered upon conducting usability tests (results reported in the next section). As a general comment the tools have shown good usability results but have way on for improvement.

5.3 System Technical Evaluation

The technical assessment of the components while running, has been evaluated with the deployed version of the system for pilots. The Models Host is running into a Windows Server 2012 R2 Standard, with an Intel® Xeon ® processor E5405 2GHz. with a RAM memory of 2,35GB. Performance and resources utilization has been monitored using the Choreographer Logger Service and default Windows/Ubuntu Performance Analysis Tools. A routine for the execution of each model was launched ten times while Key Performance Indicators were recorded. Highest and Lowest values have been removed, the average of the following eight has been calculated and reported in this section.

The system used in the study was successful in enabling the collection of data for the clinical, technical and usability validation. Some minor technical issues were raised at the start of the evaluations, however, thanks to the approach of providing a distributed Service Oriented Architecture, these could be quickly resolved without affecting the pilot execution.

Performance results are shown for different scenarios, evaluating relevant Key Performance Indicators (KPI) as memory usage, data base growth, time delays and other features.

5.3.1 Map of evaluations

The system is a Service Oriented Architecture system composed by three main modules: 1) Data Storage Module; 2) Model Host Module; and 3) Interface Module. Several components deployed in different technologies conform each of these modules, and the collaboration and perfect communication among them was a critical issue to guarantee the proper execution of the defined workflows. The evaluations have been done as the study clinical scenarios (Section 4.4), but more specifically the components affected are:

- Data Warehouses (DW)
- Data Access Layer – Query Engine (QE) – multiple/single subject
- Missing Data Imputation (MDI)
- Risk Score Module (RSM)
- Orchestrator (O)
- Interface Module

The execution of the mentioned components do not follow a subsequent schema, as there are some of them that operate in the background and update new information or model outcomes as they are ready to be sent to related components (for instance the QE checks if result data are already available from previous request and displays cached results, without invoking MDI/RSM again).

Technical performance has been done on the mentioned components and looking for the following indicators:

- Verification of the model execution:
  - Appropriateness of the Query
  - Units homogenization
  - Handling the Results and Storage
- Performance of the model execution:
  - Best Case Vs Worst Case
  - Latency (time delay of the response)
  - Memory Load in the System Server
  - Central Process Unit load
  - Network resources

5.3.2 Verification of the models execution

The evaluated system has integrated the state of the art statistical models (risk scores and data imputation) as their own scripts and not as executable files. By this, the system can overtake hot-updates (without stop-reset) and minor modifications easily (re-calibration).Figure 5.17 shows an example of the missing data imputation model integration. On the left side the original code, on the right side, the integration script which implements a call to the R engine and the raw script file. As matter of integrating raw code, there were some verifications to be done in the way that the risk scores were implemented, and moreover, in the way the variables have to be given as input.

Model Integration

The first step is to check that the script (or set of scripts) that was going to execute the statistical engine (R and/or Matlab) was correctly formatted. To check this, the track service into the choreographer provides a trace of the messages exchanged among system components and their content. Prior to the system release and in the development version of the system, a query for each of the models is executed and the trace message is analysed, as described in Figure 5.17 and Table 5.6.

Bayesian Network Model

R Console for testing the model

```
> library(bnstruct)
input <- list(
 SEX = 2, AGE = 69, C_SMK=2,
 H_SMK = 2, H_STR = 2, H_HBG = 2, BMI = 28.50,
 WAIST = 92, SBP = 154, DBP = 96,
 AHT_M = 1, LLO_M = 1, CHOL = 237,
 TRIG = 9.8904 , HDL = 32 , GL0 = 4.88, MS = 2
 )
> source("BNscript.R")
>v1;v2;v3;v4;v5;v6;v7;v8;v9;v10;v11;v12;v13;v14;v15;v16;
v17;v18;v19;v20;v21

[1] 0.58741910 0.25497897 0.02094656 0.11775952
0.01889585
[1] 0 1
[1] 0 1
[1] 0.68429253 0.28048591 0.03522156
[1] 0.7040993 0.2959007
[1] 0.3743414 0.6256586
[1] 0 1
[1] 0 1
[1] 0 1 0
[1] 0 0 1
[1] 0 1
[1] 0 1
[1] 1 0
[1] 1 0
[1] 0 1
[1] 0 1
[1] 1 0
[1] 0.02868447 0.21274401 0.75857153
[1] 1 0 0
[1] 0 1
[1] 0.895596644 0.098982698 0.005420658
```

Model execution in Model Host Module

```
 "IdMenssage": "a5c02448239c47fe8be2a17af23ff9e8",
 "sender": "ModelsService",
 "receiver": "ModelService.R",
 "language": "XMSG-SOAP",
  "Content": {
   "method": "LaunchRScript",   "parameters": {
    "script": {       "type": "xs:string",
     "value":
setwd("C:\MOSAIC\Workspaces\MOSAIC\bin\Release\Model
s\")
library(bnstruct)
input <- list(\nAGE = 69 , \n SEX = 2 , \n ETHN = 1 , \n C_SMK
= 2 , \n H_SMK = 2 , \n H_STR = 2 , \n H_HBG = 2 , \n BMI =
28.50 , \n WAIST = 92 , \n SBP = 154 , \n DBP = 96 , \n AHT_M
= 1 , \n LLO_M = 1 , \n CHOL = 237 , \n TRIG =
9.89048210455007 , \n HDL = 32 , \n GL0 =
4.88888888888889 , \n MS = 2 )

 v1<-posterior$AL_WK
v2<-posterior$C_SMK
v3<-posterior$H_SMK
v4<-posterior$PHY_W
v5<-posterior$PHY_F
v6<-posterior$H_CVD
v7<-posterior$H_STR
v8<-posterior$H_HBG
v9<-posterior$BMI
v10<-posterior$WAIST
v11<-posterior$SBP
v12<-posterior$DBP
v13<-posterior$AHT_M
v14<-posterior$LLO_M
v15<-posterior$CHOL
v16<-posterior$TRIG
v17<-posterior$HDL
v18<-posterior$INS0
v19<-posterior$GL0
v20<-posterior$MS
v21<-posterior$GL120"
    },
    "resultObjects": {  "type": "xs:string",
     "value": "v1, v2, v3, v4, v5, v6, v7, v8, v9, v10, v11, v12,
v13, v14, v15, v16, v17, v18, v19, v20, v21"
   }}}}

RESPONSE:
{
 "IdMensaje": "2f78fa985d854536a485cc6b77a1b8b3",
 "sender": "ModelService.R",
 "receiver": "ModelsService",
 "language": "XMSG-SOAP",
 "inresponserespuestaa":
"a5c02448239c47fe8be2a17af23ff9e8",
 "Content": {
  "method": "LaunchRScriptResponse",
  "parameters": {    "LaunchRScriptResult": {
    "type": "xs:string",
    "value":    "[\"0,58741910    0,25497897    0,02094656
0,11775952    0,01889585\",\"0    1\",\"0    1\",\"068429253    ,
0,28048591                    0,03522156\",\"0,7040993
0,2959007\",\"0,374341423298932  0,625658576701068\",\"0
1\",\"0 1\",\"0 1 0\",\"0 0 1\",\"0 1\",\"0 1\",\"0 1\",\"1 0\",\"1 0\",\"0 1\",\"0
1\",\"1 0\",\"0,02868447 0,21274401 0,75857153\",\"1 0 0\",\"0
1\",\"0,895596643647435                    0,0989826984480582
0,00542065790450731\"]"
   }}}}
```

FIGURE 5.17: Comparison of the isolated and integrated execution of the Data Imputation R script

NAME OF THE PREDICTOR	UNITS IN DATA BASE	UNITS IN THE RISC SCORE	UNITS IN THE INTERFACE	DB -> Model	Model -> DB	Interface
PatientID	Alphanum	Alphanum	Alphanum	none	none	none
BIRTH-DATE	YYYY-MM-DD	Numeric (Years)	Numeric (Years)	YES	none	Age
GENDER	MALE FEMALE	1/2	MALE FEMALE	YES	none	M/F
ETHNIA	Caucasian Nordic African	1/2/3	Caucasian Nordic African	YES	YES	YES
HISTORIC CVD	Yes/No	1/2	YES/NO	YES	YES	YES
HISTORIC STROKE	Yes/No	1/2	YES/NO	YES	YES	YES
HISTORIC HIGH BG	Yes/No	1/2	YES/NO	YES	YES	YES
METAB. SIND.	Yes/No	1/2	YES/NO	YES	YES	YES
ANTI-HYP MED.	Yes/No	1/2	YES/NO	YES	YES	YES
LIDIP LOWERING MED.	Yes/No	1/2	YES/NO	YES	YES	YES
FAMILY HISTORY DIABETES	No/ 1st/2nd	0/1/2	No/ 1st/2nd	0/1/2	-	Yes
CURRENT SMOKER	Yes/No	1/2	YES/NO	YES	YES	YES
HABITUAL SMOKER	Yes/No	1/2	YES/NO	YES	YES	YES
BMI	Numeric (Kg/m^2)	Numeric ((Kg/m^2)	Numeric (Kg/m^2)	None	none	none
WAIST	Numeric (cm)	Numeric (cm)	Numeric (cm)	none	none	none
Systolic B.P.	Numeric (mmHG)	Numeric (mmHG)	Numeric (mmHG)	None	none	none
Diastolic B.P.	Numeric (mmHG)	Numeric (mmHG)	Numeric (mmHG)	None	none	none
PULSE	Numeric (bpm)	Numeric (bpm)	Numeric (bpm)	None	none	none
CHOL. TOTAL	Numeric (mg/dL)	Numeric (mg/dL)	Numeric (mg/dL)	mg/dL - mmol/l	mg/dL - mmol/l	mg/dL
TRYGL.	Numeric (mg/dL)	Numeric (mg/dL)	Numeric (mg/dL)	mg/dL - mmol/l	mg/dL - mmol/l	mg/dL
HDL	Numeric (mg/dL)	Numeric (mg/dL)	Numeric (mg/dL)	mg/dL - mmol/l	mg/dL - mmol/l	mg/dL
FASTING GLUCOSE	Numeric (mg/dL)	Numeric (mg/dL)	Numeric (mg/dL)	mg/dL - mmol/l	mg/dL - mmol/l	mg/dL

TABLE 5.6: Predictors Mapping across System Modules

5.3.3 Technical Assessment

Technical assessment was done for two boundary scenarios (Best and Worst case depicted in Table 5.7). Results are provided in tables and figures, which stand for a 60 seconds time-window of the described operations.

Prediction Risk Score					
	n	**Latency (s)**	**CPU(%)**	**Memory (kB)**	**Bandwith (kbps)**
Best Case	1	0.016	20.20	374,012	9.8
Worst Case	8080	25.876	60.50	463,853	173.35
Data Imputation Model					
	Input Vars	**Latency (s)**	**CPU(%)**	**Memory (kB)**	**Bandwith (kbps)**
Best Case	20	1.486	48.50	360,416	40.23
Worst Case	0	1.860	49.5	360,748	63.56

TABLE 5.7: Results of the Technical assessment for the Best and Worst scenario in the prediction risk score and the data imputation model

In Table 5.7, the worst case for data imputation model happens when there are not imputation parameters so the Bayesian network has to perform all the operations to estimate unknown variables. Whereas, if the model has all the input variables (20 for the best case), no estimating operation is needed.

Figures 5.18, 5.19 and 5.20 show the execution performance of a Risk Score for the worst case (execution over 8080 subjects). The CPU is used for an average of 60.5% during 25.876 seconds. Not interruptions are produced by memory allocations, network issues or CPU overflow.

Figures 5.21, 5.22, 5.23 and 5.24 show the performance of the data base engine for each services (Table 5.8). The CPU average use is 43.70% and the latency depends on the number of subjects that have to be uploaded. The worst case is found for loading laboratory data for 6402 subjects, which takes 248 minutes for the setup loading and 3.462 seconds for subsequent queries.

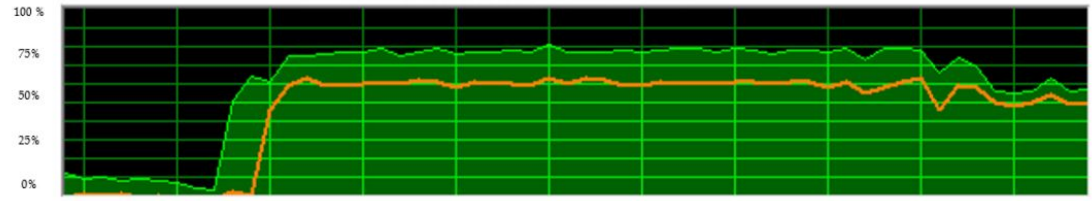


FIGURE 5.18: CPU relative use(%) for the Prediction Risk Score execution under the worst case (Orange Line).

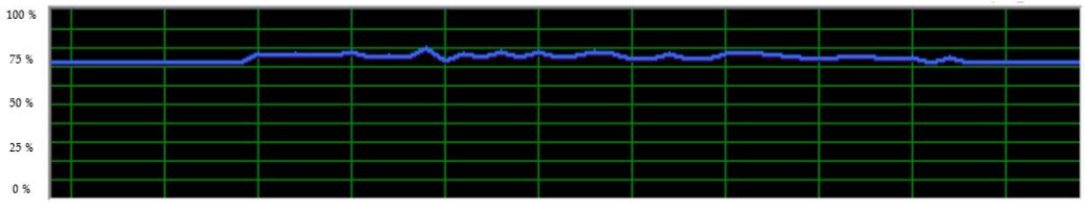


FIGURE 5.19: Memory use of the Prediction model execution for the worst case. Memory burst occurs to pagination when the model is executed

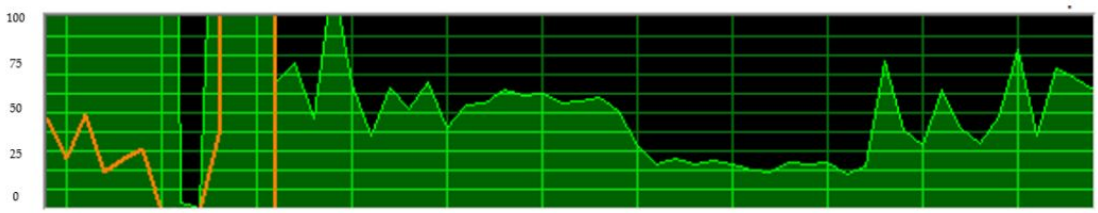


FIGURE 5.20: Network resources of the Prediction Risk Score Execution for Worst Case (Orange line). Over-buffering occurs due to the auto-scale mode of the monitor. Peak= 175,296 kpbs

Data Base module Performance						
Service	**Number of subjects**	**Time to Setup (min)**	**Latency per patient (s)**	**CPU (%)**	**Memory (kb)**	**Bandwidth (kbps)**
Emergency	658	79	7.412	43.70	137,733	720
Outpatient	1020	67	1.766			
Laboratory	6402	248	3.462			
Regular Queries	-	-	0.254	60.20	80,457	72,459

TABLE 5.8: Performance for the Data Base Management Module among different services and regular queries

FIGURE 5.21: Central Processing Unit relative use (%) for Loading Patients (Orange line)

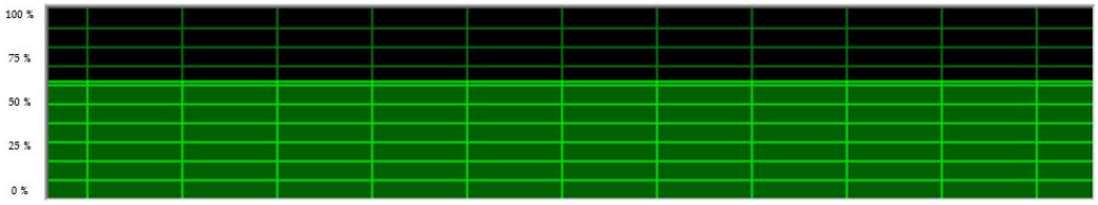


FIGURE 5.22: Random-Access Memory (RAM) relative use (%) for Patients Load).

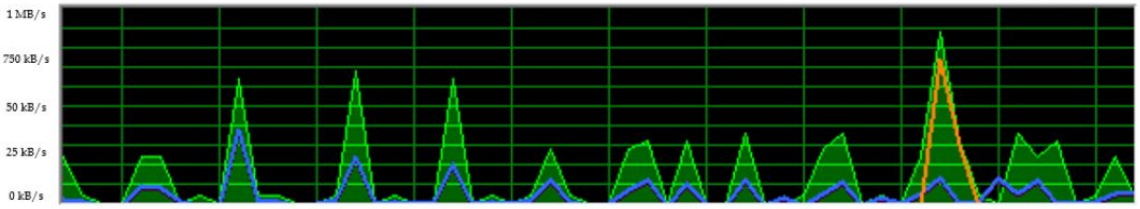


FIGURE 5.23: Hard-disk operations for the Choreographer writing/reading (Orange Line)

FIGURE 5.24: Network Resources for Loading Patients (Orange line).

Chapter 6

Discussion

6.1 Advancing the prediction and diagnose of T2DM

This study has assessed six externally-validated risk scores for the prediction of T2DM: Findrisc (Lindstrom *et al.*, 2003), ARIC (Schmidt *et al.*, 2005), San Antonio (Stern, 2002), Cambridge (Rahman *et al.*, 2008), Framingham (Wilson, 2007) and PREDIMED (Guasch-Ferré *et al.*, 2012). All these models show C statistic values ranging from 66% to 85%, either on internal and external validation. Framingham risk score has achieved an area under the ROC curve of 87,5%, which improves previous studies, whereas the rest remain within the aforementioned range.

Moreover, the study of these models draws a high variability on the number of parameters to use (predictors) and their relative weight. However, the description of these parameters (Table 4.3) confirms that T2DM is mainly an environmental disease, and therefore more and better indicators about lifestyles behaviors need to be included in future risk tools.

The results on the application of these models in clinical settings confirms their usefulness to discriminate high risk T2DM patients. Nevertheless, data quality is a paramount shortcoming that affects the scalability of this type of solutions for high-risk subjects identification.

Risk scores have been incorporated in new designed web-based tools to advance in the diagnosis and prediction of T2DM. The multilevel intervention program has proposed two use scenarios:

- Risk Stratification at clinical population level (hospitals), to be used by clinicians for targeted- early- screening combining algorithms of prediction and detection of pre-diabetes and high risk individuals.

- Supporting the clinical decision for a 2h-OGTT, at an individual level. To be used by clinicians for subject-targeted T2DM risk assessment.

The work that has been done in collaboration with the Endocrinology Service and the Quality Department of Hospital La Fe. Moreover, the collaboration among the Universitat Politècnica de València and key opinion leaders in T2DM, have filled up the gap between recommendations and the real life context.

Currently, there does not exist a homogeneous program for the screening and prevention of T2DM. Clinical guidelines for T2DM done in Spain confirms that there is not an established method to identify people at high-risk. This thesis introduced the concept of proactive search that allow the identification of high-risk population and associate clinical actions. For example, our proposed screening strategy use the MOSAIC (Sambo *et al.*, 2015) detection model to discriminate between different available screening tests. Our findings suggest that the integration in the clinical process of the risk score based screening in combination with subject-oriented lifestyle intervention could reduce the risk of T2DM. This approach had to face a different reality because there are not assigned resources to perform targeted screening and lifestyle intervention in most clinical settings . The experience of the study represents a relevant case study to illustrate the viability of such a screening strategy. Although it was not possible to measure the real clinical impact of the tool, because of shortcomings related to data quality and inaccuracies, it was possible to explore the barriers to consolidate the proposed process. The viability and usefulness in the clinical practice is confirmed by the usability and technical results.

Numerous studies have confirmed that, compared with FPG and A1C cut points, the 2-h PG value diagnoses more people with diabetes (ADA, 2016). The implementation of an accurate model for estimating the risk of having a 2h-OGTT will drive to the implementation of cost-effective precise interventions to delay or even prevent the onset of T2DM.

6.2 Prediction and detection of T2DM in clinical settings

It is possible to use the evaluated system using available electronic health records data. Data quality and availability is a critical issue that should be examined in the Information Technology service of a hospital to clean and ensure the consistency of the records prior to the risk evaluation. Based on this, it is feasible to define a proactive screening strategy based on risk scores and models, which have shown hereby acceptable accuracy results.

American Diabetes Association (ADA, 2016), which are the guidelines followed int he Endocrinology Service of Hospital La Fe, recommend the screening of all the adults with more than 45 years and in the patients with BMI $\geq$25 kg/m^2. In case that the test is negative (no diabetes or pre-diabetes states are diagnosed) the recommendation is to screen every year. By adding risk scores (Wilson, 2007) and imputation models (Sambo *et al.*, 2015), we are able to suggest a more proactive screening strategy in which a process of selective screening could be done using available data, without the need of complex data analytics or new laboratory tests. The screening strategy aims to cover two aims:

- Meet the recommendation of the American Diabetes Association in terms of diagnostic cutt-off points and frequency of screening for high risk population selection.

- Use state-of-the-art externally validated risk scores to stratify the population in three different risks and imputation models to infer the most probable result of a 2h-OGTT.

Figure 6.1 shows the proposed work flow for proactive T2DM screening in clinical settings.

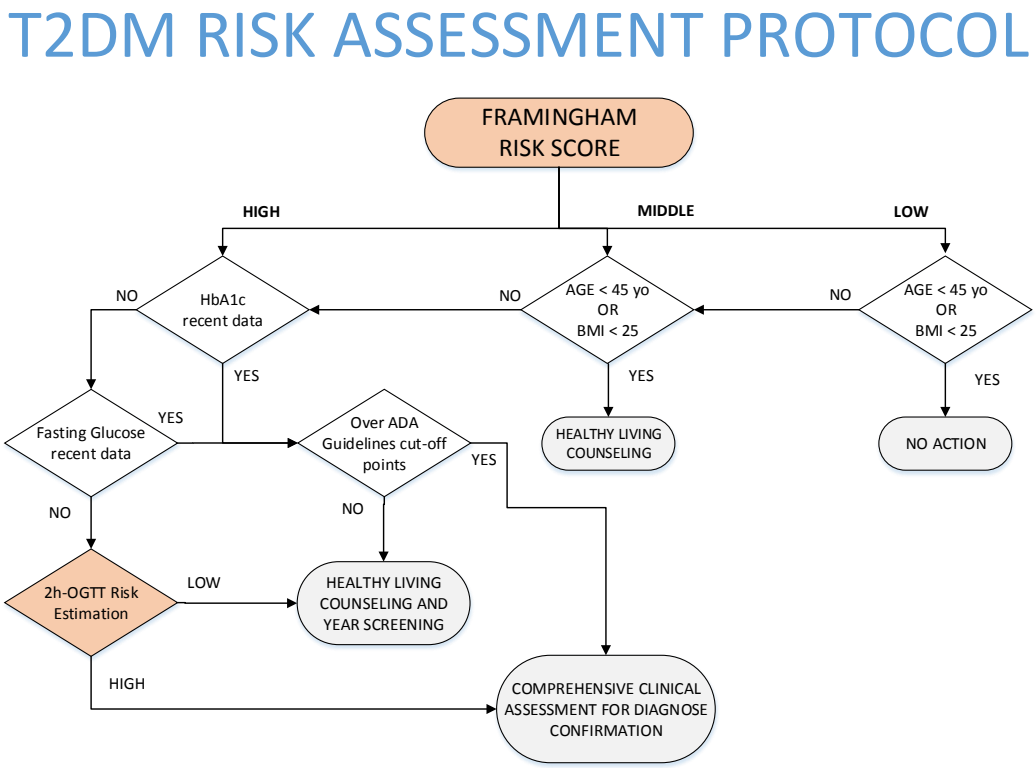


FIGURE 6.1: Work flow for proactive T2DM screening in clinical settings

The proposed risk stratification needs to be revised by the endocrinology doctors before a formal decision is made. In the hospital settings the high risk cases identified by the Framingham model can be reviewed by a nurse from the endocrinology and compared with the Electronic Health Record data.

Risk scores have been tested in the hospital de la Fe (Spain) in the endocrinology service, giving the possibility to experts in diabetes to assess the tools in the clinical practice. Unfortunately, due to technical reasons it was not possible to access data from other departments like Cardiology, Internal Medicine, Neurology, and only retrospective data of endocrinology have been used. Technicians had also to face another problem: the quality of the data. It was discovered that ICD-9 (International Classification of Disease version 9) was not correctly attributed to cases, and this required the creation of a new dataset that was manually inserted by the health professional. This was not the ideal setting for the study but from this experience, it was possible to draw the following recommendations of use:

- Proactive search can be used to select high risk population: from the discussions with medical experts and the real user of this tools the proposed tool is a novel opportunity to identify new cases of TD2M using existing data.

- It requires interdepartmental coordination: the use of the tools generate new clinical processes that could not be previously present in the clinical centres or hospital. Potential barriers that should be managed are the access to the data, and the allocation of sufficient resources for all the four actions of the screening tools (proactive search, risk stratification, case revision and actions of screening and prevention). These need to be addressed in order to consolidate the process in the real clinical practice.

- The process should be automatic as much as possible. The risk stratification should be done automatically in background and integrated in the health care records as additional clinical information to be presented in the Patient Health records. An existing example of tools are the Clinical Risk Grouping tool (provided by 3M) used in the local health care agency of Valencia that classifies patients in 1080 groups and 7 level of severity of the disease.

- Quality of the data: in order to successfully implement the tools in health care settings it is strongly recommended to assess the quality of the data and verify possible missing data, errors in the codifications etc.

Risk scores can be fitted into the current T2DM prevention and detection campaigns to define patient cohorts. To this end, the assessment on the effectiveness of a public health campaign, clinical protocol or medical technology (drug, combination of drugs, recommendations or monitoring system) will be driven for specific high risk subjects enhancing success odds.

6.3 Future Work

This Thesis has shown a inconvenient truth: Clinical records are not prepared to the Big Data era, but, the confirmation of this extent is a paved way to new research and future work.

Postdoctoral research will focus on the continuation of the main gaps found in each of the studies performed in this thesis:

1. Identification of consistent codifications on clinical records: Elaboration of new algorithms to validate clinical records without the need of a case-by-case revision.

2. Usability: Design of new organizational frameworks and tools for the integration and evaluation of preventive protocols for T2DM.

3. Technical: Implementation of distributed architectures relying on cloud computing resources such as Infrastructure as a service for a fast and efficient execution of risk scores and sustainable data storage.

However, the applicability of the proposed work-flow for the integration of prediction and detection risk scores for T2DM should need to carry out a longitudinal two-cohort study using the models suggested in this Thesis in one arm and using the current prevention interventions in the other to compare afterwards the clinical impact of the proposed work-flow. The ideal scenario would be to involve primary health care centers to perform a long-term prospective study to measure the benefits in terms on newly identified high-risk individuals and how effective the pharmacological and lifestyle intervention could be in the Mediterranean population.

The study domain of this thesis has a high level of exigency, as it requires to understand and deal with issues from several different knowledge domains: clinical (endocrinology), statistics, technology acceptance and ICT systems. This high level is demanding but extremely satisfactory when the hard work transforms into valuable results. Moreover if such results have a high potential for benefiting patients, doctors and the society in which we live in.

Chapter 7

Conclusion

7.1 Study Objectives Revision

The integration of Existing Prediction and Detection Risk Scores for Type 2 Diabetes Mellitus based on Electronic Health Records enables the detection of high risk cases whereas detection models under-perform with respect to state of the art clinical guidelines.

7.2 Objectives

7.2.1 Clinical Evaluation

- Electronic Health Records are not prepared to execute predictive risk scores as deficiencies in the quality of the data. Main shortcoming is found in the inaccuracy on the disease-specific coding time stamp, which is not correspondent with the actual onset date. Second shortcoming is found on the lack of data (missing predictors) needed to execute predictive and detection risk scores.

- After recalibration, Framingham risk score has properly classified a significant cohort of the study sample as diabetic (AUC=85%), enabling the oriented preventive treatment for delaying the onset of T2DM. Without recalibration and using the suggested cut-off points, only Frindrisc (AUC=69%) and San Antonio (AUC=73%) provide an acceptable classification accuracy.

- The risk of having a positive 2h-OGTT has under-performed (AUC=69%) with respect to the Fasting Glucose test (AUC=74%) and the HbA1C test (AUC=81%). Fasting Glucose data availability was close to 100% which suggest that the tool for supporting clinicians to decide if a 2h-OGTT is needed can be based on this indicator without the need of having a model to simulate the mos probable outcome of a 2h-OGTT. HbA1C data availability was under 50% which suggest that in this case the tool could be useful.

- Clinicians are likely to chose among pharmacological preventive interventions and healthy lifestyle recommendations for high risk subjects whereas the recommendations decrease for low risk subjects.

7.2.2 Usability Evaluation

- Risk classification and decision support tools for pre-diabetic stages have good levels of user' satisfaction but low levels in terms of usability. Such tools are introducing a breaking and disruptive routine.The active search of possible T2DM cases requires efforts to be done in terms of user training, in the definition of the organizational and procedural aspects.

- Predictive and detection tools have to increase filtering and stratification functionalities (make them more automatized, visualize predictions according to the selected strata), to provide more visualization options and to automatically build and sent reports It is needed the inclusion of more contextual information and explanations, and to provide more aggregated and less complex information (unless necessary).

7.2.3 Technical Evaluation

- The components of the designed system perform in an adequate way and present a reasonable light use of the server resources. A comprehensive and detailed description of models allowed to properly integrate different risk scores and models. The inclusion of a central component enables the description and testing the accuracy and properness of input predictors and risk outputs.

- Vulnerabilities (such as the units of the parameters across the several datasets) can be minimized by performing a comprehensive mapping of the variables and the different values they should take when used in the three big system modules: Data Base, Model Host and User Interfaces (Table 5.6). Results on the CPU, Memory and Bandwidth are acceptable and confirm that the System can be used without any exception, memory fail or interruption. Time delays of the model execution are reasonable but may be too long for the on-demand execution into the clinical daily basis practice (Table 5.7).

- The system was verified and has demonstrated to work properly with populations around 1000 patients. If the targeted population increases (10.000-100.000), further tests should be driven to check this favorable behavior.

7.3 Scientific contribution

7.3.1 Journal Publications

Martinez-Millana A, Fico G, Fernández-Llatas C and Traver V. Performance assessment of a closed-loop system for diabetes management. *Medical and Biological Engineering and Computing* (2015). Vol. 53(12), pp. 1295-1303. Springer Berlin Heidelberg.

7.3.2 International Conferences

Martinez-Millana A, Fernandez-Llatas C, Sacchi L, Segagni D, Guillen S, Bellazzi R and Traver V (2015). From data to the decision: A software architecture to integrate predictive modelling in clinical settings. *37th Annual International Conference of the IEEE In Engineering in Medicine and Biology Society (EMBC)*, Aug, 2015, pp. 8161-8164.

Fico G, Cancela J, Arredondo M, Dagliati A, Sacchi L, Segagni D, **Martinez-Millana A**, Fernandez-Llatas C, Traver V, Sambo F, Facchinetti A, Verdu J, Guillen A, Bellazzi R and Cobelli C (2015), User Requirements for Incorporating Diabetes Modeling Techniques in Disease Management Tools, In 6th European Conference of the International Federation for Medical and Biological Engineering. Vol. 45, pp. 992-995. Springer International Publishing.

Fico G, Cancela J, Arredondo M, Dagliati A, Sacchi L, Segagni D, **Martinez-Millana A**, Fernandez-Llatas C, Traver V, Sambo F, Facchinetti A, Verdu J, Guillen A, Bellazzi R and Cobelli C (2015), User Requirements for Incorporating Diabetes Modeling Techniques in Disease Management Tools, In 6th European Conference of the International Federation for Medical and Biological Engineering. Vol. 45, pp. 992-995. Springer International Publishing.

Dagliati A, Sacchi L, Bucalo M, Segagni D, Zarkogianni K, **Martinez-Millana A**, Cancela J, Sambo F, Fico G, Meneu Barreira M, Cerra C, Nikita K, Cobelli C, Chiovato L, Arredondo M and Bellazzi R (2014), A data gathering framework to collect Type 2 diabetes patients data. *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, June, 2014. pp. 244-24.

7.3.3 International Seminar

As part of the doctoral learning plan, Antonio Martinez has performed a three month research stay in the Faculty of Science and Technology of Universidade de Coimbra (Portugal). During this stay, Antonio was invited to give a seminar to disseminate the results of the research presented in this doctoral Thesis (Figure 7.1)

FIGURE 7.1: Seminar poster

7.4 Contributions to the professional development

The study presented in this Thesis was part of a European Project funded by the European Comission. Antonio Martinez has leaded the tasks related to the development, deployment and validation of the architecture presented in Chapter 4 and Chapter 5. A part from the technical work, he has represented the Small and Medium Company *Tecnologias para la Salud y el Bienestar*, in which he was working for two years. The representation has been held in the Consortium Meetings and in the Periodic Annual Technical Reviews to defend the work in front of the Project Officer and the Experts nominated by the European Comission to monitor the project progress.

As a technical leader in the intersection of medical professionals and engineers of information technology, the work developed during this research has provided an unique framework for acquiring experience and developing technical and personal skills. Even though the research question was clear, the author and supervisors have worked together to understand, collect and draw a comprehensive state of the art in T2DM risk scores and the limitations they have with respect to the applicability in real clinical scenarios, outside well-controlled clinical research trials. The collaboration among Hospital La Fe and the Universitat Politècnica de València made possible the two supervisors of this work to held a collaboration to pursue a study on how applicable these risk scores are and moreover evaluate their performance in three different dimensions of the same problem: clinical, usability and technical.

At the halfway of the research plan, when the study raw data was analyzed, the author faced a break-point that could have dropped down the entire research: the ICD-9 coding inaccuracies issues. Nevertheless, the author and supervisors stood forward by analyzing the possible corrective actions and the alternatives pathways to finalize the hypothesis testing. Thanks to the commitment of Dr. Juan Fco MERINO TORRES and María ARGENTE PLA, from the endocrinology department of Hospital La Fe, inconsistencies in the data set where analyzed case-by-case, delivering a sufficient sample size to perform the prediction and detection analysis for the three study dimensions.

The three month research stay in the Univeristy of Coimbra (Portugal) under the supervision of Professor Paulo de CARVALHO has been helpful to experiment new ways of understanding research, enhance the skills on data modeling and establish the seed for future research and innovation collaborations.

Last but not least, the active participation of the author and the supervisor Vicente TRAVER SALCEDO in the European Project MOSAIC and research related activities in the Universitat Politècnica de València has been a high-intensity skill development practical course on project management, reporting and research transference.

7.5 Impact: Awards and media

The work performed in this study was awarded with the first prize of VLC/IDEA 2015, an annual research competitive contest for pre- and post- doctoral researchers (Figure 7.2.

Moreover, this work has led to three press note releases:

- Report in a research magazine of the Spanish Television: *La aventura del saber.TELEVISIÓN ESPAÑOLA*. Link: http://www.rtve.es/m/alacarta/videos/la-aventura-del-saber/aventura-del-saber-05-04-16/3554781/?media=tve From minute 43 onwards.

- Press release and video report in the Media Services of the Universitat Politècnica de València. http://www.upv.es/noticias-upv/noticia-7898-diabetes-de-tip-es.html

- Press release from EFE in a newspaper: http://www.deia.com/2015/12/20/sociedad-/estado/un-nuevo-sistema-detecta-antes-la-diabetes-tipo-2-con-la-historia-clinica-del-paciente

FIGURE 7.2: First Prize diploma of VLCIDEA Contest on the ICT Category

Appendix A

Appendix A: Study Data Descriptive Analysis

A.1 Prediction Data Description

VARIABLES	CONTROLS (n=13)		CASES (n=12)		P VALUE	MISSING DATA (%)
GENDER	4 M / 9 F		5 M / 7 F			
	Mean	*SD*	*Mean*	*SD*		
AGE	65.76	8.20	59.41	9.28	0.082	0
BMI	28.78	5.20	32.16	8.46	0.433	56
WAIST	98.66	5.13	92.00	0.00	0.377	**84**
SBP	130.00	12.94	136.67	21.82	0.451	36
DBP	75.30	9.86	89.83	12.30	**0.020**	36
Pulse	70.85	8.78	74.00	12.20	0.613	**52**
Cholesterol	198.31	48.62	208.50	31.53	0.544	0
Triglyceride	149.23	60.63	175.75	61.96	0.290	0
HDL	45.58	17.16	49.11	13.67	0.618	16
Fasting Glucose	101.55	12.34	98.27	10.51	0.510	12
HbA1C	5.89	0.37	5.58	0.40	0.132	32

TABLE A.1: Descriptive distribution, dependency analysis and missing data rate for Cases and Controls of the prediction analysis

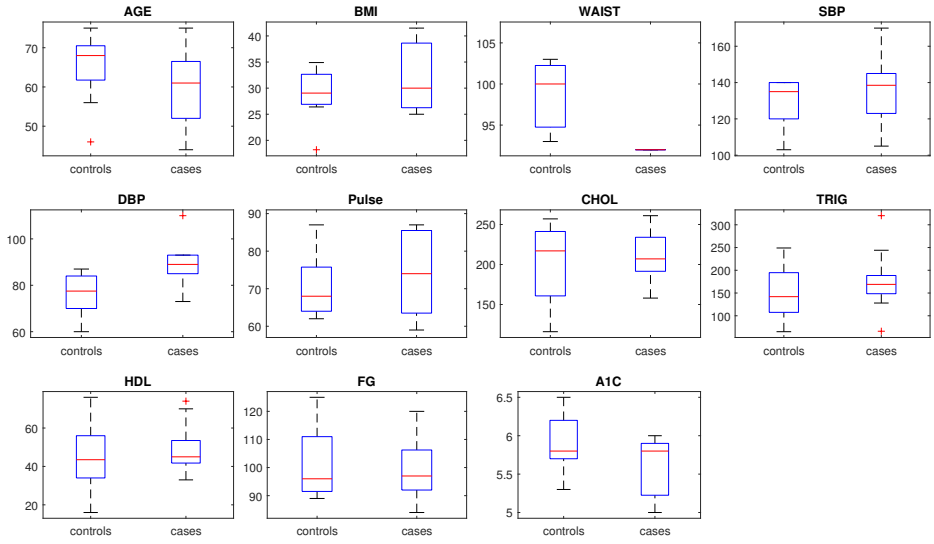


FIGURE A.1: Raw numerical data comparison for prediction risk scores

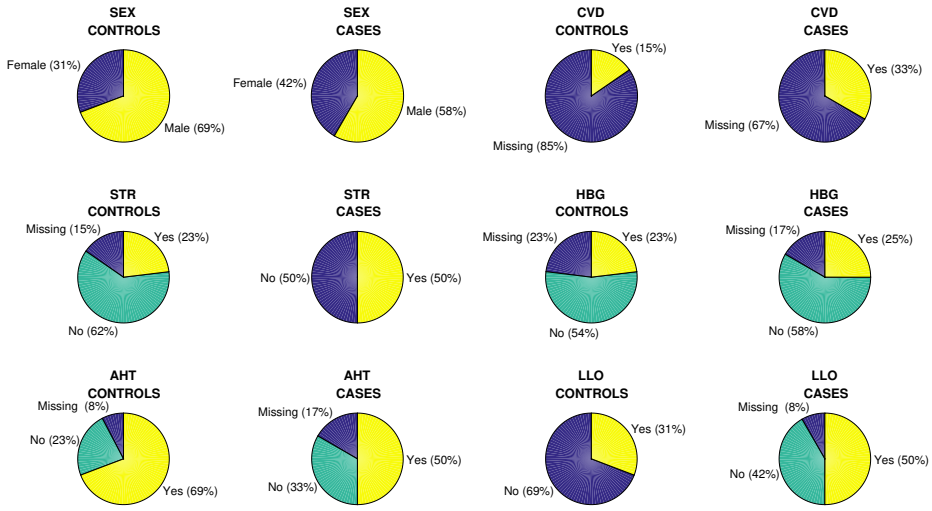


FIGURE A.2: Raw categorical data comparison for prediction risk scores(1)

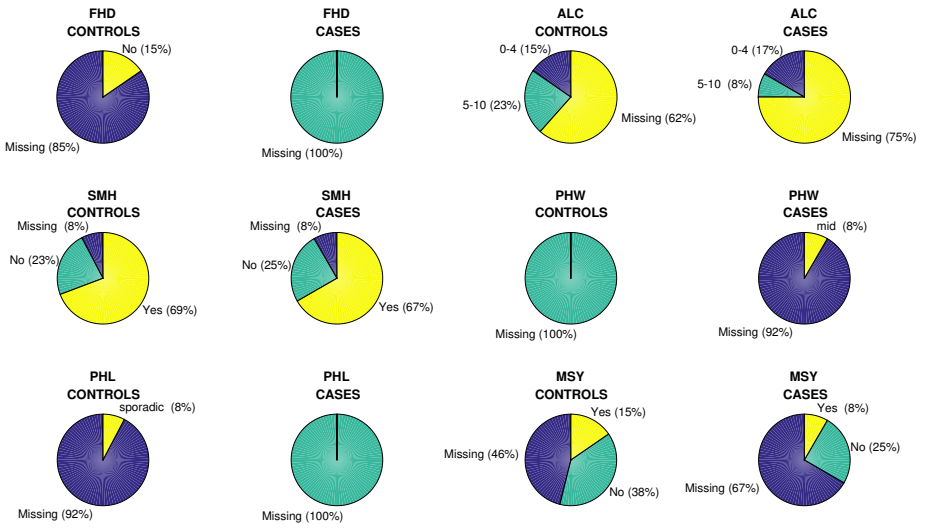


FIGURE A.3: Raw categorical data comparison for prediction risk scores (2)

A.2 Detection Data Description

VARIABLES	CONTROLS (n=25)		CASES (n=23)		P VALUE	MISSING DATA (%)
GENDER	12 M / 13 F		13 M / 10 F			
	Mean	*SD*	*Mean*	*SD*		
AGE	61.6	8.98	62.35	11.18	0.800	0
BMI	29.22	6.14	32.13	7.87	0.319	45.8
WAIST	96	6.10	115	24.95	0.262	**85.4**
SBP	135.41	18.514	128	16.749	0.237	31.25
DBP	82.41	12.76	79.5	9.07	**0.020**	36
Pulse	71.25	10.83	81.92	12.62	0.030	45.83
Cholesterol	204.76	41.43	203.23	41.75	0.900	2.08
Triglyceride	177.52	94.29	195.9	68.36	0.290	0
HDL	45.58	17.16	49.11	13.67	0.643	4.16
Fasting Glucose	100.82	11.083	108.13	8.95	**<0.05**	6
HbA1C	5.75	0.41	6.17	0.19	**<0.05**	44

TABLE A.2: Descriptive distribution, dependency analysis and missing data rate for Cases and Controls of the detection analysis

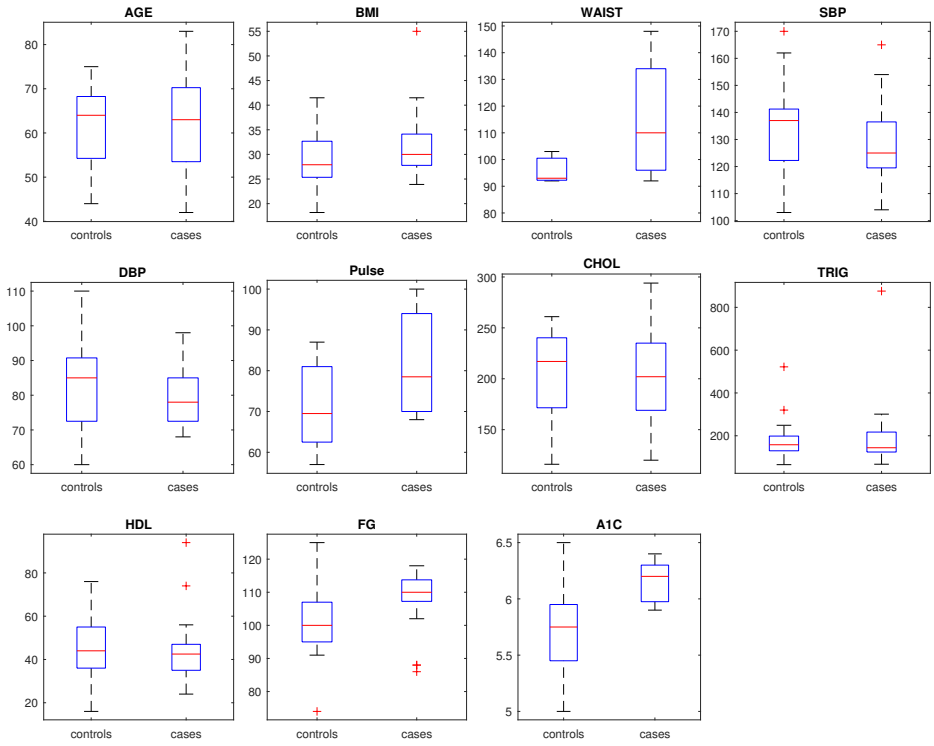


FIGURE A.4: Raw numerical data comparison for detection risk scores

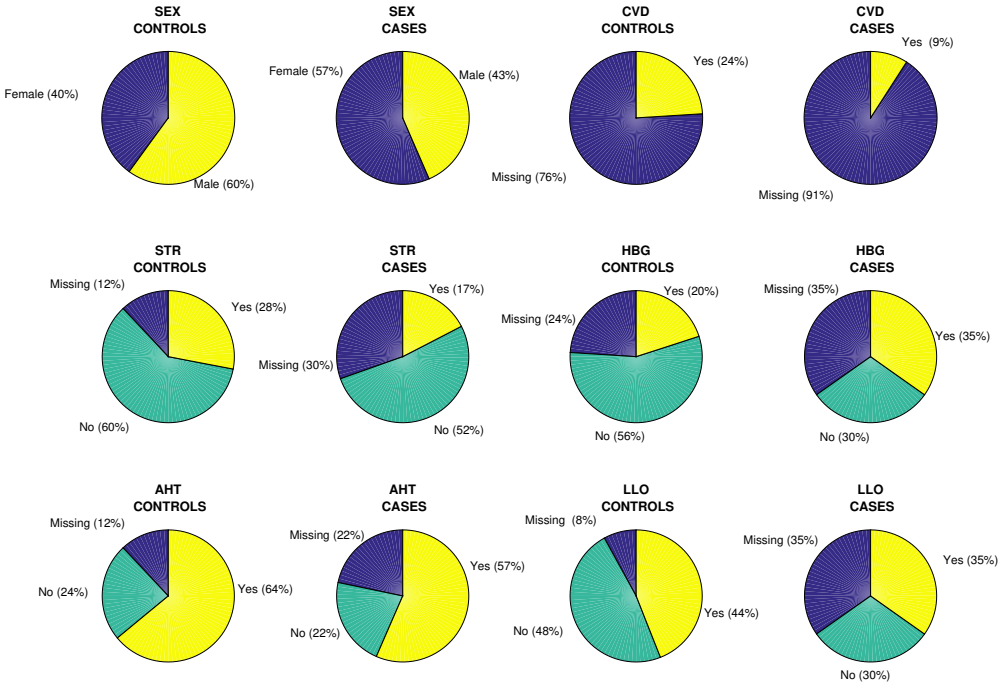


FIGURE A.5: Raw categorical data comparison for detection risk scores(1)

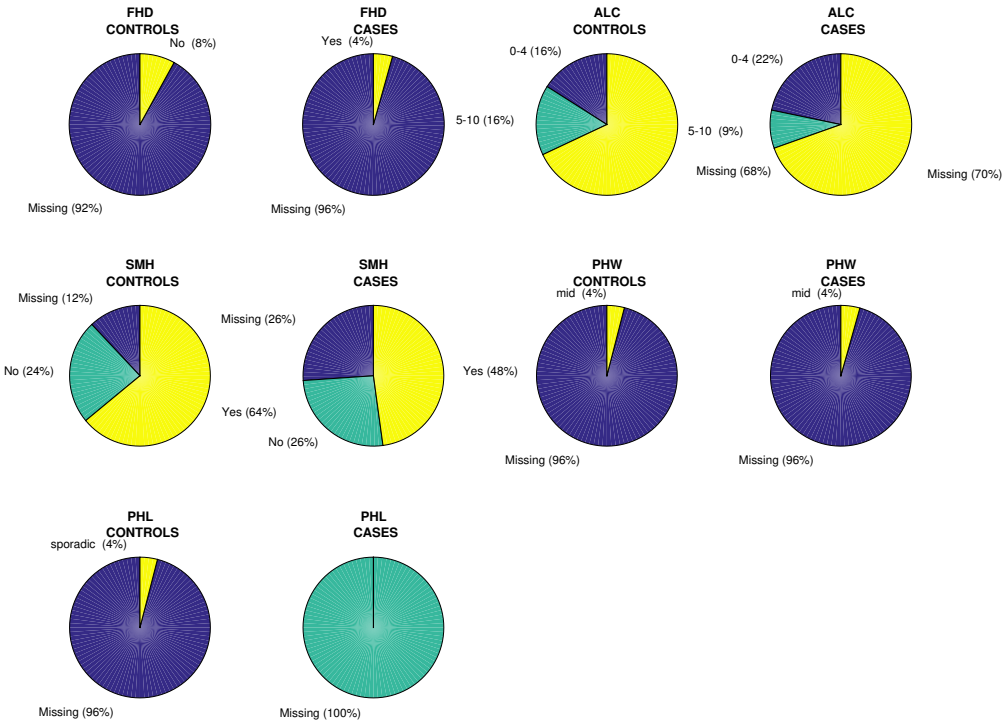


FIGURE A.6: Raw categorical data comparison for detection risk scores (2)

Appendix B

Appendix B: Usability questionnaires

B.1 AttrakDif questionnaire

Dimension	Pairwise Concepts
Pragmatic Quality	technical-human
	complicated-simple
	impractical-practical
	cumbersome-straighforward
	unpredictable-predictable
	confusing-clearly structured
	unruly-manageable
Hedonistic Quality -Human	isolating-connective
	unproffesional-proffesional
	tacky-stylish
	cheap-premium
	alienating-integrating
	separates me-brings me closer
	unpresentable-presentable
Hedonistic Quality - System	conventional-inventive
	unimaginative-creative
	cautious-bold
	conservative-innovative
	dull-captivating
	undemanding-challenging
	orginary-novel
Attractiveness	unpleasant-pleasant
	ugly-attractive
	disagreable-likeable
	rejecting-inviting
	bad-good
	repelling-appealing
	discouraging-motivating

TABLE B.1: AttrakDiff cuestionnaire. User has to choose in a 6 item Linkert Scale each of the pairwise concepts. 1 was assigned for totally agree on the left-side concept and 6 was assigned for totally agree on the right-side concept.

B.2 SUS questionnaire

Item	Question
1	I think that I would like to use this system frequently
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system
7	I would imagine that most people would learn to use this system very quickly
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

TABLE B.2: SUS cuestionnaire based on ten questions to be anwsered in a 5 item Linkert Scale: Completely Disagree - Mostly Disagree - Neutral - Mostly Agree - Strongly Agree

Bibliography

Abdul-Ghani, M. A. *et al.* (2011). "Two-Step Approach for the Prediction of Future Type 2 Diabetes Risk". In: *Diabetes Care* 34.9, pp. 2108–2112.

ADA (2015). "Classification and Diagnosis of Diabetes". In: *Diabetes Care* 39.Supplement 1, S13–S22.

— (2016). "Classification and Diagnosis of Diabetes". In: *Diabetes Care* 40.Supplement 1, S11–S24.

Agostino, R. B. D *et al.* (2008). "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study". In: *Circulation* 117.6, pp. 743–753.

Alberti, G. *et al.* (2004). "Type 2 Diabetes in the Young: The Evolving Epidemic: The International Diabetes Federation Consensus Workshop". In: *Diabetes Care* 27.7, pp. 1798–1811.

Alssema, M. *et al.* (2010). "The Evaluation of Screening and Early Detection Strategies for Type 2 Diabetes and Impaired Glucose Tolerance (DETECT-2) update of the Finnish diabetes risk score for prediction of incident type 2 diabetes". In: *Diabetologia* 54.5, pp. 1004–1012.

Altman, D. G *et al.* (2009). "Prognosis and prognostic research: validating a prognostic model". In: *BMJ* 338.may28 1, b605–b605.

Arrieta, Francisco *et al.* (2016). "Diabetes mellitus y riesgo cardiovascular: recomendaciones del Grupo de Trabajo Diabetes y Enfermedad Cardiovascular de la Sociedad Española de Diabetes (SED, 2015)". In: *Atención Primaria* 48.5, pp. 325–336.

Barry, Eleanor *et al.* (2017). "Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: systematic review and meta-analysis of screening tests and interventions". In: *BMJ*, p. i6538.

Beagley, Jessica *et al.* (2014). "Global estimates of undiagnosed diabetes in adults". In: *Diabetes Research and Clinical Practice* 103.2, pp. 150–160.

Bernstein, Rebecca S. *et al.* (2015). "Diabetes and Hypertension Prevalence in Homeless Adults in the United States: A Systematic Review and Meta-Analysis". In: *American Journal of Public Health* 105.2, e46–e60.

Brooke, John (1996). *SUS: A quick and dirty usability scale*.

Bueno H. Hernáez R., Hernández A (2008). "Diabetes mellitus tipo 2 y enfermedad cardiovascular en España: una revisión descriptiva". In: *Revista Española de Cardiología* 8.61, pp. 53–61.

Buijsse, B. *et al.* (2011). "Risk Assessment Tools for Identifying Individuals at Risk of Developing Type 2 Diabetes". In: *Epidemiologic Reviews* 33.1, pp. 46–62.

Collins, Dylan *et al.* (2016). "Simple and adaptable R implementation of WHO/ISH cardiovascular risk charts for all epidemiological subregions of the world". In: *F1000Research* 5, p. 2522.

Collins, G. S. and D. G. Altman (2011a). "External validation of QDSCORE® for predicting the 10-year risk of developing Type 2 diabetes". In: *Diabetic Medicine* 28.5, pp. 599–607.

Collins, G. S. and K. G. M. Moons (2012). "Comparing risk prediction models". In: *BMJ* 344.may24 2, e3186–e3186.

Collins, Gary S *et al.* (2011b). "Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting". In: *BMC Medicine* 9.1.

Collins, Gary S. *et al.* (2015). "Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement". In: *Annals of Internal Medicine* 162.1, p. 55.

Dam, Rob M. van (2003). "The epidemiology of lifestyle and risk for type 2 diabetes". In: *European Journal of Epidemiology* 18(12), pp. 1115–1126.

Davis, Mark H (1983). "Measuring individual differences in empathy: Evidence for a multidimensional approach". In: *Journal of personality and social psychology* 44.1, pp. 113–126.

Dorresteijn, J. A. N. *et al.* (2011). "Estimating treatment effects for individual patients based on the results of randomised clinical trials". In: *BMJ* 343.oct03 1, pp. d5888–d5888.

Drake, A J (2002). "Type 2 diabetes in obese white children". In: *Archives of Disease in Childhood* 86.3, pp. 207–208.

E. Codd, S. Codd and C. Salley. (1993). *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate*. E.F. Codd Associates.

EASD (2013). "Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD". In: *European Heart Journal* 34.39, pp. 3035–3087.

Eriksson, K-F and Folke Lindgärde (1998). "No excess 12-year mortality in men with impaired glucose tolerance who participated in the Malmö Preventive Trial with diet and exercise". In: *Diabetologia* 41.9, pp. 1010–1016.

Gillies, Clare L *et al.* (2007). "Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis". In: *Bmj* 334.7588, p. 299.

Gloy, V. L. *et al.* (2013). "Bariatric surgery versus non-surgical treatment for obesity: a systematic review and meta-analysis of randomised controlled trials". In: *BMJ* 347.oct22 1, f5934–f5934.

González, Paloma, Eduardo Faure, Almudena del Castillo, *et al.* (2006). "Coste de la diabetes mellitus en España". In: *Medicina clinica* 127.20, pp. 776–784.

Guariguata, L. *et al.* (2014). "Global estimates of diabetes prevalence for 2013 and projections for 2035". In: *Diabetes Research and Clinical Practice* 103.2, pp. 137–149.

Guasch-Ferré, Marta *et al.* (2012). "A Risk Score to Predict Type 2 Diabetes Mellitus in an Elderly Spanish Mediterranean Population at High Cardiovascular Risk". In: *PLoS ONE* 7.3. Ed. by Noel Christopher Barengo, e33437.

Guide, MATLAB User's (2014). "The MathWorks Inc". In: *Natick, MA* 4, p. 382.

Help, Matlab (2017). *Techniques to Improve Performance*. 20.02.2017.

Hemingway, H. *et al.* (2013). "Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes". In: *BMJ* 346.feb05 1, e5595–e5595.

Herman, I. (2011). *Semantic Web. W3C*. 17.02.2017.

Hippisley-Cox, J. *et al.* (2009). "Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore". In: *BMJ* 338.mar17 2, b880–b880.

Hornik, Kurt (2012). "The comprehensive R archive network". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 4.4, pp. 394–398.

ICER (2016). *Diabetes Prevention Programs: Effectiveness and Value. Evidence Report*. Institute for Clinical and Economic Review.

Kahn, Steven E., Rebecca L. Hull, and Kristina M. Utzschneider (2006). "Mechanisms linking obesity to insulin resistance and type 2 diabetes". In: *Nature* 444.7121, pp. 840–846.

Kenny, David A. (1987). *tatistics for the social and behavioral sciences*. Boston: Little, Brown.

Knowler (2002). "Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin". In: *New England Journal of Medicine* 346.6, pp. 393–403.

Komatsoulis, George A. *et al.* (2008). "caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability". In: *Journal of Biomedical Informatics* 41.1, pp. 106–123.

Ley, Sylvia H *et al.* (2014). "Prevention and management of type 2 diabetes: dietary components and nutritional strategies". In: *The Lancet* 383.9933, pp. 1999–2007.

Lindstrom, J. and J. Tuomilehto (2003). "The Diabetes Risk Score: A practical tool to predict type 2 diabetes risk". In: *Diabetes Care* 26.3, pp. 725–731.

Lindström, Jaana *et al.* (2006). "Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study". In: *The Lancet* 368.9548, pp. 1673–1679.

Little, Randie R. (2003). "Glycated Hemoglobin Standardization – National Glycohemoglobin Standardization Program (NGSP) Perspective". In: *Clinical Chemistry and Laboratory Medicine* 41.9.

Lluch-Ariet, Magí (2016). *The MOSAIC System - Contributions to efficient and secure exchange of networked clinical data*.

Mann, D. M. *et al.* (2010). "Comparative Validity of 3 Diabetes Mellitus Risk Prediction Scoring Models in a Multiethnic US Cohort: The Multi-Ethnic Study of Atherosclerosis". In: *American Journal of Epidemiology* 171.9, pp. 980–988.

Martinez-Millana, A. *et al.* (2015). "Performance assessment of a closed-loop system for diabetes management". In: *Medical & Biological Engineering & Computing* 53.12, pp. 1295–1303.

Massi-Benedetti, M. (2002). "The Cost of Diabetes Type II in Europe The CODE-2 Study". In: *Diabetologia* 45.S1, S1–S4.

Mazouzi, Hamza, Amal El Fallah Seghrouchni, and Serge Haddad (2002). "Open protocol design for complex interactions in multi-agent systems". In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems part 2 - AAMAS02*. Association for Computing Machinery (ACM).

McElduff, P (2004). "Will changes in primary care improve health outcomes? Modelling the impact of financial incentives introduced to improve quality of care in the UK". In: *Quality and Safety in Health Care* 13.3, pp. 191–197.

Meigs, James B. *et al.* (2008). "Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes". In: *New England Journal of Medicine* 359.21, pp. 2208–2219.

Montonen, Jukka *et al.* (2003). "Whole-grain and fiber intake and the incidence of type 2 diabetes". In: *The American journal of clinical nutrition* 77.3, pp. 622–629.

Moons, Karel G M *et al.* (2012). "Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker". In: *Heart* 98.9, pp. 683–690.

Moreno, Antonio *et al.* (2006). "Applying Agent Technology to Healthcare: The GruSMA Experience". In: *IEEE Intelligent Systems* 21.6, pp. 63–67.

MOSAIC (2014). *Models and simulation techniques for discovering diabetes influence factors*. 5.02.2017.

Murphy, S. *et al.* (2009). "Instrumenting the health care enterprise for discovery research in the genomic era". In: *Genome Research* 19.9, pp. 1675–1681.

Murphy, S. N. *et al.* (2010). "Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)". In: *Journal of the American Medical Informatics Association* 17.2, pp. 124–130.

Newcomer, Eric and Greg Lomow (2004). *Understanding SOA with web services (independent technology guides)*. Addison-Wesley Professional.

Noble, D. *et al.* (2011). "Risk models and scores for type 2 diabetes: systematic review". In: *BMJ* 343.nov28 1, pp. d7163–d7163.

Norlund, A *et al.* (2001). "Cost of illness of adult diabetes mellitus underestimated if comorbidity is not considered". In: *Journal of internal medicine* 250.1, pp. 57–65.

Pan, X.-R. *et al.* (1997). "Effects of Diet and Exercise in Preventing NIDDM in People With Impaired Glucose Tolerance: The Da Qing IGT and Diabetes Study". In: *Diabetes Care* 20.4, pp. 537–544.

Paulweber, B. *et al.* (2010). "A European Evidence-Based Guideline for the Prevention of Type 2 Diabetes". In: *Hormone and Metabolic Research* 42.S 01, S3–S36.

Phillips, Lawrence S. *et al.* (2014). "We Can Change the Natural History of Type 2 Diabetes". In: *Diabetes Care* 37.10, pp. 2668–2676.

Pyykkonen, A.-J. *et al.* (2011). "Depressive Symptoms, Antidepressant Medication Use, and Insulin Resistance: The PPP-Botnia Study". In: *Diabetes Care* 34.12, pp. 2545–2547.

R. Roset M. Lurgi, M. Croitoru B. Hu M. Lluch-Ariet and P. Lewis. (2008). "Visual mapping tool for database interoperability: the healthagents case". In: *Proceedings of the Third Conceptual Structures Tool Interoperability Workshop*. CEUR, pp. 44–54.

Rahman, M. *et al.* (2008). "A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study". In: *Family Practice* 25.3, pp. 191–196.

Reilly, Brendan M. (2006). "Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions". In: *Annals of Internal Medicine* 144.3, p. 201.

Richardson, Leonard and Sam Ruby (2008). *RESTful web services*. " O'Reilly Media, Inc."

Riley, Richard D *et al.* (2016). "External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges". In: *BMJ*, p. i3140.

Rubin, Donald B (1996). "Multiple imputation after 18+ years". In: *Journal of the American statistical Association* 91.434, pp. 473–489.

Ruiz-Ramos, Miguel *et al.* (2006). "La diabetes mellitus en España: mortalidad, prevalencia, incidencia, costes económicos y desigualdades". In: *Gaceta Sanitaria* 20, pp. 15–24.

Ryden *et al.* (2013). "ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD". In: *European Heart Journal* 34.39, pp. 3035–3087.

Sacks, D. B. (2012). "Measurement of Hemoglobin A1c: A new twist on the path to harmony". In: *Diabetes Care* 35.12, pp. 2674–2680.

Sambo, F. *et al.* (2015). "A Bayesian Network analysis of the probabilistic relations between risk factors in the predisposition to type 2 diabetes". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Institute of Electrical and Electronics Engineers (IEEE).

Schmidt, M. I. *et al.* (2005). "Identifying Individuals at High Risk for Diabetes: The Atherosclerosis Risk in Communities study". In: *Diabetes Care* 28.8, pp. 2013–2018.

Segagni, D. *et al.* (2011). "R Engine Cell: integrating R into the i2b2 software infrastructure". In: *Journal of the American Medical Informatics Association* 18.3, pp. 314–317.

Simmons, R. K., J. B. Echouffo-Tcheugui, and S. J. Griffin (2010). "Screening for type 2 diabetes: an update of the evidence". In: *Diabetes, Obesity and Metabolism* 12.10, pp. 838–844.

Site, FIPA Web (2017). *Foundation for intelligent physical agents*. 22.02.2017.

Soriguer, Federico *et al.* (2012). "Validación del FINDRISC (FINnish Diabetes Risk SCore) para la predicción del riesgo de diabetes tipo 2 en una población del sur de España. Estudio Pizarra". In: *Medicina Clínica* 138.9, pp. 371–376.

Stern, Michael P. (2002). "Identification of Persons at High Risk for Type 2 Diabetes Mellitus: Do We Need the Oral Glucose Tolerance Test?" In: *Annals of Internal Medicine* 136.8, p. 575.

Steven M. Haffner, and Seppo Lehto *et al.* (1998). "Mortality from Coronary Heart Disease in Subjects with Type 2 Diabetes and in Nondiabetic Subjects with and without Prior Myocardial Infarction". In: *New England Journal of Medicine* 339.4, pp. 229–234.

Steyerberg, Ewout W. *et al.* (2013). "Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research". In: *PLoS Medicine* 10.2, e1001381.

Talmud, P. J *et al.* (2010). "Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study". In: *BMJ* 340.jan14 1, b4838–b4838.

Thomas, Louis H. Philipson (2015). "Update on Diabetes Classification". In: *Medical Clinics of North America* 99.1, pp. 1–16.

Tuomilehto, Jaakko (2014). "Prevention of type 2 diabetes–where is the evidence?" In: *EMHJ* 20.11.

Tuomilehto, Jaakko *et al.* (2001). "Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance". In: *New England Journal of Medicine* 344.18, pp. 1343–1350.

Umpierrez, G. E. and M. Korytkowski (2013). "Is Incretin-Based Therapy Ready for the Care of Hospitalized Patients With Type 2 Diabetes?: Insulin therapy has proven itself and is considered the mainstay of treatment". In: *Diabetes Care* 36.7, pp. 2112–2117.

Valdés, Sergio *et al.* (2007). "Population-based incidence of type 2 diabetes in Northern Spain". In: *Diabetes care* 30.9, pp. 2258–2263.

Venkatesh, Viswanath and Fred D Davis (2000). "A theoretical extension of the technology acceptance model: Four longitudinal field studies". In: *Management science* 46.2, pp. 186–204.

Venkatesh, Viswanath *et al.* (2003). "User Acceptance of Information Technology: Toward a Unified View". In: *MIS Quarterly* 27.3, pp. 425–478. ISSN: 02767783.

Wilson, Peter W. F. (2007). "Prediction of Incident Diabetes Mellitus in Middle-aged Adults". In: *Archives of Internal Medicine* 167.10, p. 1068.