FORECASTING AND DECISION SUPPORT FOR TYPE 1 DIABETES INSULIN THERAPY USING MACHINE LEARNING

Silvia Oviedo Castillo

FORECASTING AND DECISION SUPPORT FOR TYPE-1 DIABETES INSULIN THERAPY USING MACHINE LEARNING

Universitat de Girona

DOCTORAL THESIS

SILVIA OVIEDO CASTILLO

2019

Doctoral Thesis

FORECASTING AND DECISION SUPPORT FOR TYPE 1 DIABETES INSULIN THERAPY USING MACHINE LEARNING

SILVIA JULIANA OVIEDO CASTILLO

2019

Doctoral Program in Technology

Supervised by:
Dr. Josep Vehi Casellas and Dr. Iván Contreras
Tutored by:
Dr. Josep Vehi Casellas

Thesis submitted in partial fulfillment of the requirements for a doctoral degree from Universitat de Girona

Doctoral Thesis

FORECASTING AND DECISION SUPPORT FOR TYPE 1 DIABETES INSULIN THERAPY USING MACHINE LEARNING

A dissertation presented in partial fulfillment of the requirements for a doctoral degree from the University of Girona.
By:

Silvia Juliana Oviedo Castillo

Supervised by:

Dr. Josep Vehi Casellas **Dr. Iván Contreras**

FULL LIST OF PUBLICATIONS

THIS THESIS IS BASED ON A COMPENDIUM OF THE FOLLOWING PUBLICATIONS:

[1] **Silvia Oviedo**, Josep Vehi, Remei Calm, and Joaquim Armengol, "A Review of Personalized Blood Glucose Prediction Strategies for T1DM Patients." in *International Journal for Numerical Methods in Biomedical Engineering*, vol. 33, no. 6, pp. 1-21. 2017. doi:10.1002/cnm.2833

[2] Iván Conteras, **Silvia Oviedo**, Martina Vettoretti, Roberto Visentin and Josep Vehí. "Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models" in *PLoS One*, 12(11), e0187754.

[3] **Silvia Oviedo**, Iván Contreras, Carmen Quirós, Marga Giménez, Ignacio Conget and Josep Vehi, "Risk-based Postprandial Hypoglycemia Forecasting Using Supervised Learning" in *International Journal of Medical Informatics*. Aug, 2018. Submitted.

[4] **Silvia Oviedo**, Iván Contreras, Arthur Bertachi, Carmen Quirós, Marga Giménez, Ignacio Conget and Josep Vehi, "Bolus Advisor Application for Machine Learning Based Postprandial Hypoglycemia Forecasting" in *IEEE Journal of Biomedical and Health Informatics*. Dec, 2018. Submitted.

ADDITIONAL JOURNAL ARTICLES

[1] Josep Vehi, Iván Contreras, **Silvia Oviedo**, Lyvia Biagi and Arthur Bertachi, "Prediction and Prevention of Hypoglycaemic Events in Type-1 Diabetic Patients using Artificial Intelligence" in *Health Informatics Journal*, Accepted for publication with minor revisions in Jan 2019.

PEER-REVIEWED CONFERENCES

[1] **Silvia Oviedo**, Iván Contreras, Josep Vehi, Roberto Visentin and Martina Vettoretti. "Mid-Term Blood Glucose Prediction: A Hybrid Approach using Grammatical Evolution and Physiological Models," in *10th International Conference on Advanced Technologies & Treatments for Diabetes (ATTD 2017)*, Paris, France, Feb.2017. Poster presentation.

[2] Iván Contreras, Arthur Bertachi, Lyvia Biagi, **Silvia Oviedo**, Josep Vehi, "Blood Glucose Level Prediction Challenge: Using Grammatical Evolution to Generate Short to Mid-Term Prediction models," *3rd International Workshop on Knowledge Discovery in Healthcare Data, at IJCAI-ECAI 2018*, Stockholm, Sweden, 2018.

[3] **Silvia Oviedo**, Iván Contreras, Carmen Quirós, Marga Giménez, Ignacio Conget, Josep Vehi "Postprandial Hypoglycemia Prediction in Patients with Type 1 Diabetes using Insulin Pump Therapy," *Diabetes Technology Meeting (DTM 2018)*, Maryland, USA, 2018. Accepted for poster presentation.

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my advisor, Dr. Josep Vehí for giving me the opportunity to work under his guidance; for his support and optimism throughout this research. I am deeply grateful to Dr. Iván Contreras, my co-advisor, for generously sharing his knowledge and his work with me, including his code, so I could advance in my research. Also, I thank Dr. Joaquim Armengol and Dra. Remei Calm for their valuable help in the definition of my research problem.

I would like to thank to Dr. Rodolfo Villamizar and Dr. Jabid Quiroga for trusting me and introducing me to the research group.

I am very grateful to Dr. Oscar Gualdrón, for his friendship and for sharing his practical outlook in both technical and personal matters.

I thank to Dr. Carlos Borrás for encouraging me to pursue new challenges and helping and advising me personally.

I would like to express my special appreciation to my brother, Andrés, for his support both financially and technically. I thank him for looking after our parents while I was absent.

To Daniela Aguirre, for introducing me to many useful computer tools to enhance my research. But specially, for her friendship, company and kind words when I needed them the most. To Arthur and Lyvia for their companionship and help throughout our stay in Girona.

To Gina Villanueva, for her great friendship despite the distance that gave me strength to carry on.

I am thankful to Richard Zießler, for his love and support during these years.

I owe my deepest gratitude to my family for their constant interest in my well-being and their love.

Finally, I would like to acknowledge the financial support of the Spanish Ministry of Economy and Business, particularly in the award of the pre-doctorate grant BES-2014-068289 from 2015 to 2019.

Dedicated to my beloved parents, Claudia and Carlos.

LIST OF FIGURES

ACRONYMS

AP	Artificial Pancreas
ANN	Artificial Neural Network
BG	Blood Glucose
CGM	Continuous Glucose Monitoring
CEG	Clarke Error Grid
FDA	Food and Drug Administration
GE	Grammatical Evolution
GA	Genetic Algorithm
gRMSE	Glucose-specific Root Mean Square Error
IOB	Insulin On Board
ISF	Insulin Sensitivity Factor
ICR	Insulin to Carbohydrate Ratio
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MDI	Multiple Daily Injections
NB	Naive Bayes
SE	Sensitivity
SP	Specificity
SVC	Support Vector Classifier
SAP	Sensor Augmented Pump
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes

CONTENTS

ABSTRACT

Insulin therapy for Type 1 Diabetes (T1D) has several ramifications with different degrees of automation. The advances in sensors and monitoring devices have led to an increasing availability of data. Additionally, machine learning algorithms usage has sprung, allowing the development of models for Blood Glucose (BG) forecasting with relative ease. Nevertheless, BG forecasting is still a challenging task for prediction horizons beyond 30 min and, even more so, with missing or erroneous data, which is a common burden in the field.
This thesis is devoted to generate machine learning models that forecast either BG levels using regression algorithms or postprandial hypoglycemia using classification algorithms. The application of these models range from Multiple Daily Injections (MDI) therapy up to Sensor Augmented Pump (SAP) therapy.
On one hand, this work focuses on the prediction of BG values by proposing a hybrid model that uses Grammatical Evolution (GE), an insulin on board model, and a glucose rate of absorption model to predict BG values with a prediction horizon of 120 min. The algorithm relies on the construction of a set of rules that determine the search space for an optimization algorithm based on a Genetic Algorithm (GA). A glucose-specific fitness function leads the evolution of the solution while penalizing deviations based on their clinical harmfulness and a tailored evolutionary grammar.
On the other hand, this study delves into the methods to forecast hypoglycemic events aiming to contribute to decision-making tools in T1D therapy. For this reason, a method for training classification models that predict postprandial hypoglycemia is also proposed and validated for MDI and SAP applications, using real patients' data in free-living conditions. The method relies on well-known machine learning algorithms and, in some cases, a combination of them to anticipate hypoglycemia using an entirely data-driven approach with carbohydrate content estimation, insulin bolus and BG level as common inputs. The aforementioned approaches are evaluated using clinically meaningful metrics that provide insights regarding the practical use of the proposed methods. The obtained results are promising and contribute to the advances in the development of technologies for the management of type 1 diabetes.

RESUMEN

La terapia con insulina para pacientes con T1D tiene varias ramificaciones con diferentes grados de automatización. Los avances en sensores y dispositivos de monitorización conllevan un incremento en la disponibilidad de datos. Adicionalmente, el uso de algoritmos de aprendizaje automático se ha popularizado, facilitando el desarrollo de modelos para pronosticar Glucosa en Sangre (GS) con mayor facilidad. Sin embargo, predecir los niveles de GS es una tarea compleja para ventanas de predicción más allá de 30 min, y más aún, con datos erróneos o faltantes, una limitación muy frecuente en este campo. Esta tesis está dedicada a la generación de modelos basados en aprendizaje automático para predecir ya sean niveles de GS usando algoritmos de regresión o, hipoglicemia postprandial usando algoritmos de clasificación. La aplicación de estos modelos va desde terapia de múltiples inyecciones diarias (MID) hasta la terapia SAP.
Por una parte, este trabajo se enfoca en la predicción de valores de GS proponiendo un modelo híbrido que emplea gramáticas evolutivas (GE), un modelo de insulina a bordo, y un modelo de absorsión de glucosa para predecir los niveles de GS con un horizonte de predicción de 120 min. El algoritmo se fundamenta en la construcción de un conjunto de reglas que determinan el espacio de búsqueda de un algoritmo de optimización basado en algoritmos genéticos. Una función de costo especial para medidas de glucosa conduce la evolución de la solución mientras penaliza las desviaciones basándose en el impacto clínico de las mismas y una gramática evolutiva a medida.
Por otra parte, este trabajo profundiza en métodos para la predicción de eventos de hipoglicemia apuntando a contribuir con el desarrollo de herramientas de apoyo a decisiones terapéuticas. Por esta razón, en este trabajo también se propuso y validó un método para el entrenamiento de algoritmos de clasificación para la predicción de hipoglicemia postprandial, aplicada a las terapias MID y SAP usando datos de pacientes reales. El método consiste en algoritmos comunes de aprendizaje automático y, en algunos casos, la combinación de algunos de ellos para anticipar hipoglicemias usando un enfoque de modelo basado en datos, con entradas relacionadas con la estimación de carbohidratos, dosis de insulina y nivel de glucosa en sangre como entradas comunes a los modelos. La metodologías mencionadas fueron evaluadas usando métricas con significancia clínica que permiten evaluar el uso práctico de los métodos propuestos. Los resultados obtenidos son prometedores y contribuyen a los avances en el desarrollo de tecnologías para el manejo de la diabetes tipo 1.

RESUM

La teràpia amb insulina per a pacients amb T1D tenen vàries ramificacions amb diferents graus d'automatització. Els avenços en sensors i dispositius de monitorització comporten un increment en la disponibilitat de dades. A més a més, l'ús d'algoritmes d'aprenentatge automàtic s'han popularitzat, facilitant així el desenvolupament de models per pronosticar Glucosa en Sang (GS) amb major facilitat. No obstant això, preveure els nivells de GS és una tasca complexa per a finestres de predicció més enllà de 30 minuts, i més encara, amb dades errònies o absents, la qual cosa és una limitació molt freqüent en aquest camp.

Aquesta tesis està dedicada a la generació de models basats en aprenentatge automàtic per predir ja siguin nivells de GS utilitzant algoritmes de regressió o hipoglucèmia postprandial utilitzant algoritmes de classificació. L'aplicació d'aquests models van des de la teràpia de múltiples injeccions diàries (MID), fins a la teràpia SAP.

Per una banda, aquest treball es focalitza en la predicció de valors de GS proposant un model híbrid que utilitza gramàtiques evolutives (GE), un model d'insulina a bord, i un model d'absorció de glucosa per predir els nivells de GS en un horitzó de predicció de 120 minuts. L'algoritme es basa en la construcció d'un conjunt de regles que determinaran l'espai de cerca d'un algoritme d'optimització basat en algoritmes genètics. Una funció de cost especial per mesures de glucosa que condueix l'evolució de la solució mentre penalitza les desviacions basant-se en l'impacte clínic de les mateixes i una gramàtica evolutiva a mida.

Per altra banda, aquest treball profunditza en mètodes per la predicció d'esdeveniments de la hipoglucèmia apuntant a contribuir en el desenvolupament d'eines de suport a decisions terapèutiques. Per aquesta raó, en aquest treball també s'ha proposat i validat un mètode per l'entrenament d'algoritmes de classificació per la predicció de la hipoglucèmia postprandial, aplicada a les teràpies MID i SAP utilitzant dades de pacients reals. El mètode consisteix en algoritmes comuns d'aprenentatge automàtic i, en alguns casos, la combinació d'alguns d'ells per anticipar hipoglucèmies utilitzant una visió de model basat en dades, amb entrades relacionades amb l'estimació de carbohidrats, dosis d'insulina i nivell de glucosa a la sang com entrades més comuns en els models.

Les metodologies esmentades han sigut evaluades utilitzant mètriques amb significança clínica que permeten evaluar l'ús pràctic dels mètodes proposats. Els resultats obtinguts són prometedors i contribueixen en els avenços en el desenvolupament de tecnologies pel tractament de la diabetis tipus 1.

1 INTRODUCTION

1.1 MOTIVATION

Diabetes is a chronic disease characterized by the body's inability to produce the necessary amount of insulin to regulate Blood Glucose (BG) levels. According to the International Diabetes Federation (IDF), in 2017 there were 425 million people with diabetes in the world, making diabetes one of the largest global health emergencies of the 21st century [1]. Figure 1 shows the worldwide distribution of the prevalence (%) of diabetes for men and woman in 2017. According to the same report, diabetes prevalence for both men and women will rise to nearly 10% in 2045.

Diabetes can be classified in three main groups: Type 1 Diabetes (T1D), Type 2 Diabetes (T2D), and gestational diabetes. Specifically, T1D is caused by an autoimmune response that destroys β-cells. People with T1D highly rely on external insulin in order to regulate their BG levels. As stated in [2] the reason for the increasing number of people who develop T1D is still unclear; hypothetical reasons include changes in environmental risk factors and/or viral infections. The aforementioned motivated several efforts in biological solutions, such as encapsulated islet transplantation, where encapsulated islets are transplanted using a minimally invasive intervention. The encapsulation protects the pancreatic islets within a immunoisolation device to avoid the immune response and posterior destruction of the graft without the need for toxic immunosuppression. The major limitation of this method is the fibrotic overgrowth surrounding the graft, which leads to oxygen and nutrients depletion in the transplanted islets[3]. Another alternative aims at the restoration of the β-cells function via gene therapy [4]. So far, in a mouse model, blood glucose was reestablished for 4 months, prior to further destruction of the β-cells.
Insulin therapy solutions include Multiple Daily Injections (MDI) therapy, Sensor Augmented Pump (SAP) therapy and the most ambitious so far: The Artificial Pancreas (AP). The AP is an automated system designed to control BG and reduce T1D associated events such as hypo/hyperglycemia, which are life-threatening situations for the T1D patients. As shown in Figure 2, the AP is a closed-loop control system relying on a Continuous Glucose Monitoring (CGM) sensor that continuously approximates BG values as BG_m, a control algorithm, an insulin pump that administers the bolus and basal insulin I as control action and the patient, who receives the control

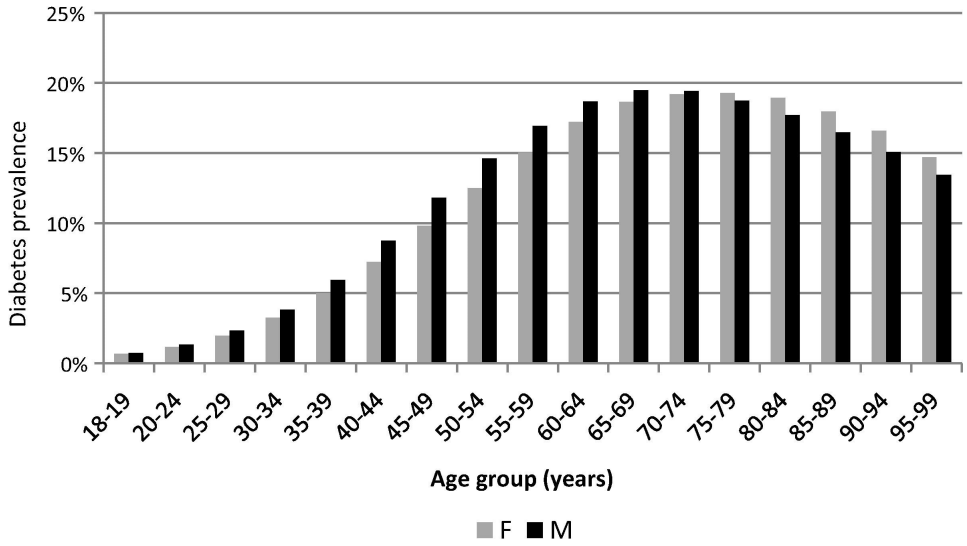

Figure 1: Worldwide prevalence (%) of diabetes by age and sex, 2017. Image from [1]

actions as insulin therapy. The control actions depend on the deviation e of the measured BG from the set point, r. The CGM has a trans-dermal glucose sensor that measures glucose concentration in the tissue fluid. To date, there are two licensed AP devices, with their components depicted in Figure 3. The fist one is the MiniMed 670G closed-loop system. It has been licensed by the Food and Drug Administration (FDA) for commercial use. It is designed to reduce/stop or increase insulin delivery when it detects low or high BG levels in the patient, respectively. The insulin pump delivers an insulin dosage through an infusion cannula. The pump is able to update the amount of insulin to be delivered according to the CGM information and the controller output. This system has been assessed in some clinical studies [5], demonstrating its capacity to reduce the risk of hiper/hypoglycemia, nocturnal hypoglycemia and increase the time in normal BG range. In addition, the Guardian™ Connect CGM sensor users can use a separate app for iPhone to assist them finding blood glucose trends in the patients' data[6]. The Sugar.IQ performs a continuous analysis of the BG sensor data to provide insights about the relationship between BG and food, insulin and lifestyle factors.

The second licenced system is the DBLG1 Diabeloop system. It has been granted the CE marking for medical devices. This system uses a CGM, a patch insulin pump and a controlling algorithm in a handset device to monitor blood glucose every five minutes and determine the correct dose of insulin based on a algorithm using past data and patient's physiology. In a two-centre, randomised, crossover clinical trial [7], Diabeloop significantly improved glycaemic control while reducing the risk of hypoglycaemia in 29 adults with T1D.

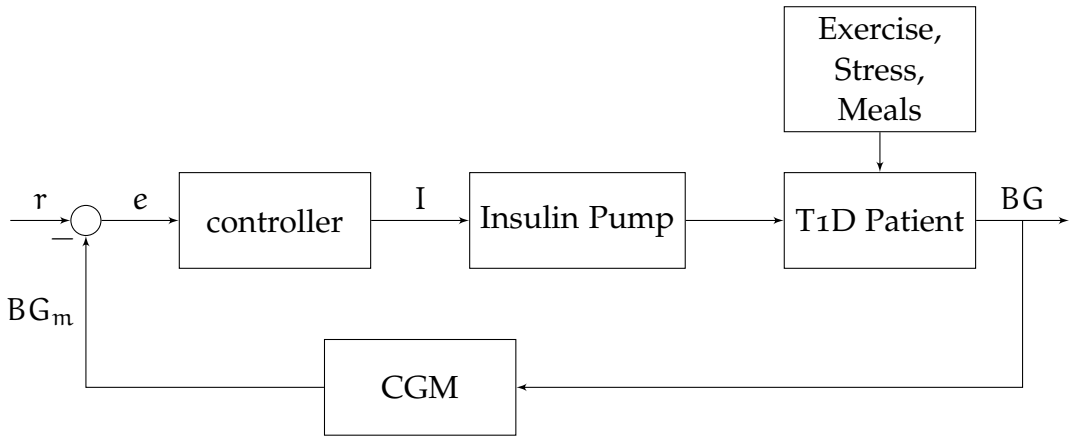

Figure 2: **Block diagram of the closed-loop control for AP**. The main components of the system are the CGM sensor, the controller and the Insulin Pump on the patient.

A fully automated AP system must accurately calculate and administer the right amount of insulin I, minimize hypoglycemic events, generate alerts [8], detect and cope with several types of faults, be able to adapt to changing conditions, such as fasting, food intake and exercise, and should be simple and adjustable for the clinical practice [9]. However, one of the main obstacles for achieving a fully automated AP is the lack of BG prediction models that are reliable enough to model the dynamics of a diabetic patient's physiology. A fully reliable model should not only be able to mimic the patient's physiology, but also cope with external disturbances, such as sensor noise, exercise and unannounced meals [10], among other factors.

Many T1D patients rely on open-loop therapies. Sensor Augmented Pump Therapy (SAP) is a widely known scheme in which the patient and physician are highly involved in the pump dosage management, including the management of pump suspensions and Bolus and Basal insulin, according to the CGM sensor readings. On the other hand MDI therapy is the most common method for insulin treatment [11]. It is an open-loop control strategy where T1D patients regulate their BG levels using several injections of fast-acting insulin (bolus insulin) to act at mealtime and injections of long-acting insulin to cover the fasting conditions. Bolus insulin doses are usually calculated based on an estimation of the carbohydrate intake in grams, Insulin Sensitivity Factor (ISF), Insulin to Carbohydrate Ratio (ICR), the current BG value and the amount of insulin still present in the body from previous injections know as Insulin On Board (IOB). Both ICR and ISF are dependent on ever-changing factors such as physical activity, circadian variations and hormone cycles[12], and therefore, estimating the optimal amount of insulin bolus is still a challenging part of the glycemic control for every T1D patient. It is worth mentioning that MDI therapy, despite being the less automated insulin therapy for T1D, can be complemented with

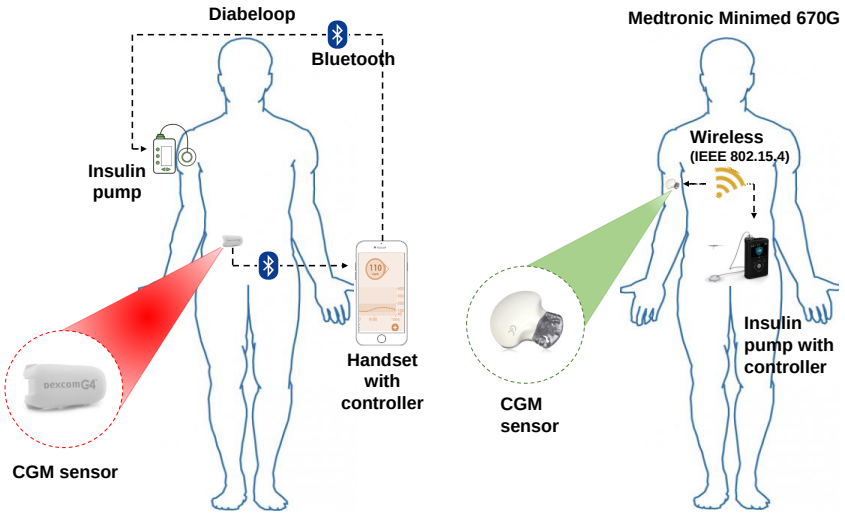

Figure 3: Comparative representation of the two licensed AP systems: Diabeloop and Minimed 670G.

the use of CGM sensors, which benefits sub-optimal BG control in these patients.

In the described context, there are a lot of research opportunities depending on the type of therapy that best suits a given patient. Particularly, this thesis aims to contribute towards the development of new methods for BG forecasting both as a regression and classification problem. A regression solution learns a model that delivers a continuous value prediction, such a BG value. On the other hand, a classification model delivers a class label or discrete value that can be interpreted as a condition such as hypoglycemia or hyperglycemia, for instance. The regression model was developed for SAP therapy and the classification for both SAP and MDI therapy.

1.2 OBJECTIVES

The research work presented in this thesis is devoted to explore methods for forecasting BG values and postprandial hypoglycemic events using ML. More specifically, the objectives pursued in this work are the following:

- To develop a personalized scheme for mid-term BG prediction with a glucose-specific loss function and assess the clinical harmfulness of the deviations from the target values.
- To design a novel method for hypoglycemia prediction as a classification scheme to be used in decision support systems for patients treated using SAP or MDI therapy.
- To validate the methods for the mid-term BG prediction and postprandial hypoglycemia forecasting with retrospective clinical data.

- To design a Bolus advising application and a validation scheme for the postprandial hypoglycemia forecasting models.

1.3 RESEARCH CONTEXT

Blood glucose concentration is sensitive to several variables including the quantity of ingested carbohydrates, insulin administration, physical activity, stress, and the presence of other diseases besides diabetes [13]. Including these variables in models for prediction is challenging because of the inherent complexity of physiological models. In addition, lifestyle and emotion-related variables are difficult to measure and quantify, such as stress or exercise [14]. Variability between patients is another important source of complexity, and it has been addressed by individualizing the forecasting methods for each patient. This is relevant because patients exhibit large variations in their BG signals during the day, especially after a meal or physical activity. Others, for instance, experience a blood sugar increase before, during or after anxious moments [15]. In addition to the daytime food and exercise-related variability, at night is when most of the cases of severe hypoglycemia occur [16]. Therefore, personalized glycemic prediction strategies have become necessary for BG control because it is neither safe nor accurate to use models with generalized parameters that do not reflect the dynamic behavior of the patient during the day.

BG predictive models can be classified into three categories: physiological models, machine learning based models and hybrid models. Physiological models require a previous understanding of insulin and glucose metabolism [17]. They are useful for simulating BG metabolism in the form of compartmental models and for studying the physiological processes that are involved in glucose regulation. These models can also be divided into two types according to their complexity. The first type of models are the minimal models, which are capable of capturing crucial processes of glucose metabolism and insulin action with few equations and identifiable parameters[17]. The second type is maximal or comprehensive models, which comprise all the available knowledge of the physiological system and are capable of simulating a diabetic patient's metabolic response, which allows *in silico* experiments to assess controllers and treatments [18]. In recent decades, several authors have proposed models of insulin action and glucose kinetics using experimental data to measure glucose production, glucose utilization, and insulin and meal absorption. Many of those models are compartmental models[19], which describe the processes that occur in the inaccessible portions of the system because these processes are not directly measurable. Therefore, the inaccessible portion of a system is represented by a number of interconnected compartments. The

most popular proposals regarding physiological models of insulin action and the glucose kinetics system are the Dalla Man Model [20], Hovorka model [21] and Bergman minimal model[22]. The different models allow for the estimation of variables, such as subcutaneous insulin absorption, gastric emptying, carbohydrate digestion and absorption, insulin kinetics, and glucose metabolism. More specifically, the Dalla Man model is composed of one glucose and one insulin subsystem linked by the control of insulin in glucose utilization and endogenous production. In contrast, the Bergman minimal model uses a three-compartment model to represent the concentrations of plasma insulin I (mU mL-1), remote insulin X (min-1), and plasma glucose G (mg dL-1). Finally, the Hovorka model uses two compartments representing the kinetics of glucose and regards each insulin action with its final effect on BG separately. For these models, the input variables include factors from external insulin therapy over time. Physiological models for BG prediction [23],[24],[25] are less popular nowadays because of the advent of ML approaches.

In contrast to physiological models, ML models fully rely on CGM data and, sometimes, additional signals to model a patient's physiological response without involving physiological variables. ML models include time series models [26][27][28][29], GA models, GE models [30], Multi-model approaches [31][32] , Gaussian Mixture Models (GMM) [33] and Artificial Neural Network (ANN) models [34] [35] [36] , among others. For instance, Zarkogianni et al. [37] compared several data-based techniques using as inputs the most recent glucose measurement G(t), the change in glucose level Δ G(t), and the sum of energy expenditures during the last 30 min.

An alternative scheme is to use physiological models for glucose digestion and absorption, insulin absorption and exercise. These models are used in a pre-processing stage and the outputs from this stage are incorporated into a ML model. Models of this type [38] [39] [40] [41] are commonly known as hybrid models because they partially rely on physiological models and require the identification and setting of some physiological parameters.

The scheme for a mixed physiological and ML model is usually a module based on a physiological model followed by a data-driven model that learns the relationship between inputs and future outcomes, which could be expressed either by means of classes (qualitative approach) or by means of the actual BG continuous values (quantitative approach). For instance, Georga et al. [42][43] assessed support vector regression and random forests methods using using CGM(mg/dl), plasma insulin concentration (μU/ml), instantaneous energy expenditure, and meal-derived glucose rate of appearance Ra(mg/min) as inputs. Another example of this approach is the model [44] using GE and physiological models of IOB and glucose rate of appearance Ra(mg/min) for the blood glucose predic-

tion challenge held in 2018 using the Ohio T1D data-set for BG prediction [45].

However, because physiological models are somewhat time-consuming to develop and require previous knowledge to set the physiological constants, scientific efforts are currently concentrated in exploring innovative and less time-consuming models by taking advantage of the always growing machine learning modeling options. Hybrid models make use of the simplest physiological models to process meal information and insulin therapy information and then fit data-driven models to future BG outcomes. Finally, data-driven models completely rely on some non-physiological formulations to characterize the relationship between current and past CGM, insulin and meal carbohydrate content with future BG outcomes.

Future BG concentration is the most popular outcome in predictive models for T1D [46]. Nevertheless, there are other possibilities such as adapting classifiers to detect life-threatening conditions,such as hyper/hypoglycemia, and facilitate decision making for both patients and physicians. For example, if there is a future outcome that lies beyond the established normal ranges, a predefined recommendation could be followed. This approach means that an effective therapy could be established without an explicit estimation of BG concentration, but rather using a class as an outcome, such a BG event. The previous raises the question: Should a model learn to predict future continuous values (regression problem) of BG or should it learn to map inputs to pre-established classes (classification problem)? Those classes could include normal glycemic levels, hypoglycemia and hyperglycemia, for instance. This work deals with both perspectives for prediction and states the practical applications benefited by each one.

As far as prediction horizon is concerned, due to the inherent delays with subcutaneous insulin infusion action and glucose sensing, it is desirable to find a reasonable compromise between the accuracy of the prediction model and its prediction capability. Generally speaking, an increase in the prediction horizon leads to a deterioration in the accuracy of the prediction for a given model. Nevertheless, the inclusion of meal information, physical activity, other input signals and changing the model structure also affect the accuracy of a particular prediction horizon. Therefore, performance metrics should be understood as a function of the selected prediction horizon and the individual, clinician or decision system must select the prediction horizon-accuracy relationship that best meets the patient's needs. In the literature review of the modeling strategies in T1D for BG forecasting [46], presented in Part I of this compendium, a range of 15–120 min is usually explored, and a 30 min prediction horizon is the most common value. On one hand, prediction models for MPC (Model Predictive Control) require shorter predic-

tion windows $< 45min$ whereas alarm systems for low-glucose can range from 15 to 30 min. Applications such as postprandial and nocturnal hypoglycemia prediction demand larger prediction windows $> 45min$.

1.4 THESIS OUTLINE

The body of this thesis consists of a compendium of four peer-reviewed journal articles specified in the *Full list of publications* section. Chapter 1 includes the introduction and the objectives of the research, along with a research context describing the most recent trends for predicting BG levels or BG events. Chapter 2 presents the papers that were produced in the course of this research, including a review paper that expands the research context described in Chapter 1. Each paper is presented as a Part. Figure 4 shows how the papers from parts II to IV are related in the context of T1D therapy applications using machine learning. Each one of the four largest nodes is one of the topics that this research connects. In this manner, Part II presents a CGM forecasting approach using GE for mid-term BG prediction as a regression problem, using an *in-silico* patient cohort as a data source. The machine learning algorithm used for this approach was GE, which is an evolutionary computation algorithm that uses Backus – Naur form to find symbolic expressions that model a specific problem. In addition, minimal physiological models were used to pre-process the carbohydrate and insulin inputs to complement the machine learning model. In Part III, a different point of view on the forecasting problem was adopted by targeting the prediction of postprandial hypoglycemia as a classification problem for SAP therapy. This work added a new challenge to the forecasting task using data from real patients in free-living conditions. This approach was data-driven, using Support Vector Classifier (SVC), a well-known supervised learning algorithm to explore different alternatives of classifiers to predict mild and severe postprandial hypoglycemia. Finally, Part IV kept the classification scheme of Part III as well as the real patient data but this time targeting MDI therapy, given that this type of therapy is still widely used by many T1D patients. This work also used additional machine learning algorithms such as Naive Bayes (NB), AdaBoost and ANN to diversify the underlying prediction principles. Next, Chapter 3 presents a brief discussion on the main results of the articles that form this thesis. Finally, Chapter 4 conveys the main contributions of the research, the conclusions and future works.

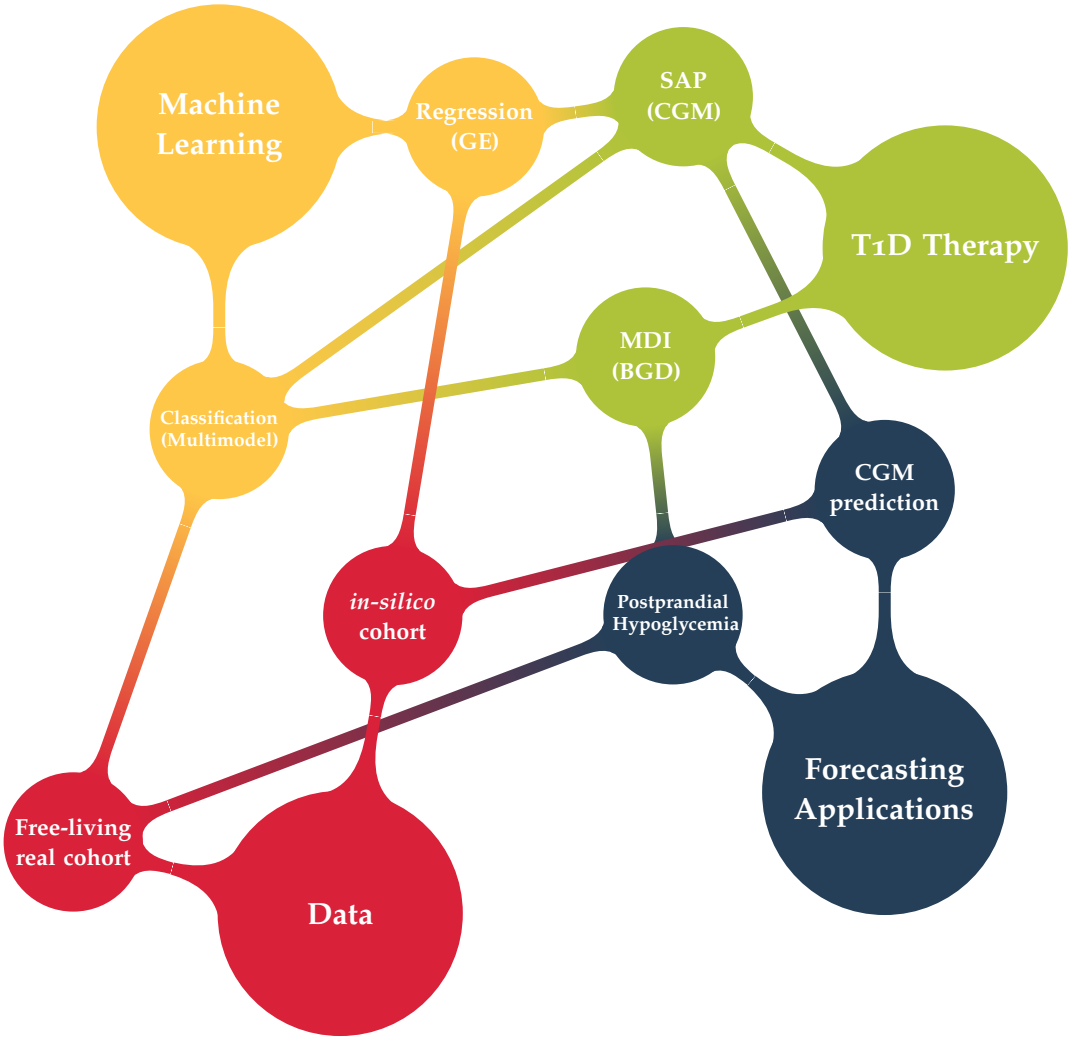

Figure 4: **Relation diagram for the outline of the thesis in the context of T1D therapy using ML tools.** Some of the connections between the smaller circles represent research topics in which this thesis is developed.

2

FORECASTING FOR TYPE-1 DIABETES INSULIN THERAPY USING MACHINE LEARNING

This Chapter consists of four Parts. The first part presents a review article that explains the recent trends in forecasting for T1D. Part II-IV correspond to the submitted or published version of the journal articles in which this thesis is grounded:

- *Part I*: **A review of personalized blood glucose prediction strategies for T1DM patients**
- *Part II*: **Personalized Blood Glucose Prediction: A Hybrid Approach Using Grammatical Evolution and Physiological Models**
- *Part III*: **Risk-based Postprandial Hypoglycemia Forecasting Using Supervised Learning**
- *Part IV*: **Bolus Advisor Application for Machine Learning Based Postprandial Hypoglycemia Forecasting**

Part I

A REVIEW OF PERSONALIZED BLOOD GLUCOSE PREDICTION STRATEGIES FOR T1DM PATIENTS

Published in International Journal for Numerical Methods in Biomedical Engineering. Sept, 2016 (JCR quartile: Q2; JIF: 2.192 in 2016; Ranked 34/77 in Biomedical Engineering).

Silvia Oviedo, Josep Vehi, Remei Calm, and Joaquim Armengol, "A Review of Personalized Blood Glucose Prediction Strategies for T1DM Patients" in *International Journal for Numerical Methods in Biomedical Engineering,* vol. 33, issue 6 (2017) : p. 1-21

http://dx.doi.org/10.1002/cnm.2833

Received: 15 July 2016 / Revised: 15 September 2016 / Accepted: 16 September 2016

Abstract

This paper presents a methodological review of models for predicting blood glucose (BG) concentration, risks and BG events. The surveyed models are classified into three categories, and they are presented in summary tables containing the most relevant data regarding the experimental setup for fitting and testing each model as well as the input signals and the performance metrics. Each category exhibits trends that are presented and discussed. This document aims to be a compact guide to determine the modeling options that are currently being exploited for personalized BG prediction.

Keywords

- artificial pancreas
- blood glucose prediction
- data-driven BG prediction models
- hybrid BG prediction models
- physiological BG prediction models
- predictive models

Part II

PERSONALIZED BLOOD GLUCOSE PREDICTION: A HYBRID APPROACH USING GRAMMATICAL EVOLUTION AND PHYSIOLOGICAL MODELS

Published in PLoS ONE. Nov, 2017 (JCR quartile: Q1; JIF: 2.766 in 2017; Ranked 15/64 in Multidisciplinary Sciences).

PLOS ONE

RESEARCH ARTICLE

Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models

Iván Contreras[1☯]*, Silvia Oviedo[1☯], Martina Vettoretti[2], Roberto Visentin[2], Josep Vehí[1]

1 Institut d'Informàtica i Aplicacions, Parc Científic i Tecnològic de la Universitat de Girona, Girona, Spain,
2 Department of Information Engineering, University of Padova, Padova, Italy

☯ These authors contributed equally to this work.
* ivancontrerasfd@gmail.com

OPEN ACCESS

Citation: Contreras I, Oviedo S, Vettoretti M, Visentin R, Vehí J (2017) Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models. PLoS ONE 12(11): e0187754. https://doi.org/10.1371/journal.pone.0187754

Editor: Petter Bjornstad, University of Colorado Denver School of Medicine, UNITED STATES

Received: July 12, 2017

Accepted: October 25, 2017

Published: November 7, 2017

Data Availability Statement: All software files are available from the public github repository: https://github.com/IvanContrerasFD/MiceLab-Grammatical-Evolution-and-Type-1-Diabetes.

Funding: This work was partly supported by the Spanish Ministry of Science and Innovation (grants DPI 2013-46982-C2-2-R and DPI2016-78831-C2-2-R), the People program (Marie Sklodowska-Curie Actions) of the European Union Seventh Framework Programme (FP7/2007-2013) with agreement (No. 600388) (TECNIOspring

Abstract

The large patient variability in human physiology and the effects of variables such as exercise or meals challenge current prediction modeling techniques. Physiological models are very precise but they are typically complex and specific physiological knowledge is required. In contrast, data-based models allow the incorporation of additional inputs and accurately capture the relationship between these inputs and the outcome, but at the cost of losing the physiological meaning of the model. In this work, we designed a hybrid approach comprising physiological models for insulin and grammatical evolution, taking into account the clinical harm caused by deviations from the target blood glucose by using a penalizing fitness function based on the Clarke error grid. The prediction models were built using data obtained over 14 days for 100 virtual patients generated by the UVA/Padova T1D simulator. Midterm blood glucose was predicted for the 100 virtual patients using personalized models and different scenarios. The results obtained were promising; an average of 98.31% of the predictions fell in zones A and B of the Clarke error grid. Midterm predictions using personalized models are feasible when the configuration of grammatical evolution explored in this study is used. The study of new alternative models is important to move forward in the development of alarm-and-control applications for the management of type 1 diabetes and the customization of the patient's treatments. The hybrid approach can be adapted to predict short-term blood glucose values to detect continuous glucose-monitoring sensor errors and to estimate blood glucose values when the continuous glucose-monitoring system fails to provide them.

Introduction

The human body requires the maintenance of blood glucose (BG) levels in a very narrow range (70–110 mg/dl). Many exogenous factors affect these levels. The pancreas releases insulin and glucagon hormones secreted by β-cells and α-cells, respectively, to regulate the BG levels. Type 1 diabetes mellitus (T1D) is the consequence of an autoimmune attack on β-cells that

programme) of the REA and the Agencia per a la Competitivitat de L'Empresa (ACCIÓ), and by the Spanish Government through contract ES-2014-068289.

Competing interests: The authors have declared that no competing interests exist.

significantly impairs insulin production. Thus, individuals with T1D fully rely on external insulin to manage their BG levels.

Therapies based on continuous glucose monitoring (CGM) devices associated to insulin pump technology (combined systems CGM-CSII) are rapidly becoming more common. A key part of these therapies in order to be truly effective is a decision support system, or an artificial pancreas that is able to predict what is going to happen in a relatively long period of time.

As widely known in clinical practice, achieving tight glycemic control is a complex process for certain patients who exhibit large variations in their BG signals due to several factors that influence the glycemic response and thereby influence glycemic control. Factors such as physical activity, weather conditions, dietary disturbances, age, and the psychological state of the patient [1][2][3] in conjunction with endogenous processes such as circadian rhythms [2], other diseases, and the menstrual period and pregnancy in women [4][5] strongly affect glucose metabolism. Because these factors are varied and often not easily identifiable, the prediction of BG values using personalized models is particularly important. Personalized models can capture specific lifestyle factors that influence the physiological response of each patient to carbohydrate intake and insulin dosage. The great variation in the glycemic response of T1D patients makes predictive modeling a challenging and crucial task.

The treatment of diabetes is conditioned by high intra- and inter-patient variability. Inter-patient variability greatly limits the use of general models because they cannot capture the specific physiological behavior of an individual. Intra-patient variability makes it difficult to apply one model for the glucose dynamics of an individual. Inter- and intra-patient variability is tackled by personalizing and customizing prediction models. This study avoids the limitations of classical modeling by implementing a set of customized models for each patient using an evolutionary approach. This paper targets the midterm (120 min) anticipated BG level while considering the clinical safety of the predictions. The models are based on a machine-learning algorithm that is flexible enough to include innovative features.

Most recent studies on BG prediction used only data-driven models [4][5] or a complementary approach that combined data-driven models and compartmental models [6][7]. Other works focused on control applications for predictions, such as the prevention of nighttime hypoglycemia [8].

BG prediction models are classified into three types: physiological, data-driven, and hybrid. Physiological models require a good understanding of insulin and glucose metabolism and contain parameters that should be set only by those with expert knowledge. These models are commonly used in simulators via compartmental models, as discussed in [9]. Minimal versions of some physiological models exist [10][11]; however, the main challenge of this type of approach is achieving a good model with high generalization capability. Data-driven models completely rely on BG data and possibly other inputs. Data-driven models are typically based on machine learning techniques and use techniques such as genetic algorithms, robust filters, fuzzy logic, rule-based models [12], multi-model approaches [13][14], autoregressive models [15][16], regularized learning, reinforcement learning, random forests, support vector regression, and artificial neural networks models [17]. Finally, an alternative architecture for BG prediction models involves properly setting a physiological model to describe glucose digestion and absorption, a second model for insulin absorption, and possibly other models to account for exercise or other events. These models constitute a preprocessing stage, the output of which enters a data-driven model. This type of model is commonly known as a hybrid model and some recent approaches to them were examined in previous studies [18][19][20]. Oviedo et al. [21] provided a comprehensive review of models for predicting BG.

Recently, BG estimation using grammatical evolution (GE) was included in a study [22] in which a novel customization of BG models for five virtual patients using GE was proposed. GE

is a search algorithm with a modular design that can be used to generate predictive time series models. It uses an evolutionary-like process to achieve expressions or computer programs optimized according to a predefined objective function. GE uses a grammar to implement a linking process between the search algorithm and the actual solution. This is a key component of GE and one of the reasons why GE is attractive. The grammar consists of a set of rules that defines the structure of the expressions generated and thus the final solution by the algorithm. This structure can be modified fairly easily according to the applications needs, meaning that it can be as simple or as complex as the user determines, without altering the search algorithm performance. Authors in [22] incorporated medical knowledge into a grammar aimed to build expression for glucose that considered previous BG values, carbohydrate intake, and insulin administration. This incorporation involved exploring four different grammars and five fitness functions, all of which were evaluated with respect to average error as a performance metric for all patients. The results indicated that it is feasible to evolve useful models for modeling BG values that consider BG readings, meals, and insulin dose information. Another study [23] extended the findings of [22] by including three additional virtual patients and using the root-mean-square error (RMSE) as the fitness function. The authors tested the clinical significance of the results using error grid analysis (EGA) via the Clarke error grid (CEG) and the Parkes error grid.

The present study extends the aforementioned research to investigate a novel and complementary approach that uses symbolic regression through GE to determine an approximation of the underlying glucose dynamics evolving personalized BG predictive models that incorporate physiological models as part of the input. The aim of this approach is to capture the particular lifestyle factors that influence the physiologic response of T1D patients to their insulin doses and carbohydrate intakes. Thus, the study presents a tool meant to assist in T1D management issuing early warnings related to ineffective or poor treatments that could improve overall health, safety, and the quality of life of T1D patients.

The paper is organized as follows: Section 2 describes the type of data sets that are collected and the specific algorithms used to achieve the proposed tasks. Section 3 presents a summary of the experiments and results obtained. Section 4 discusses the results and compares and contrasts this approach with that of other works. Section 5 concludes the paper with a brief summary of the study and a discussion of future challenges.

Materials and methods

Fig 1 shows a schematic representation of the methodology proposed in this study. Initially, insulin, carbohydrate, and CGM data for 100 virtual subjects are generated by simulation over 14 d using the T1D patient decision-making model [24]. The information is preprocessed and divided into files, one file per patient. The information on this files regarding the carbohydrates consumption in each meal is transformed into a continuous signal according to a physiological model that describes the absorption of the carbohydrates. Additionally, the bolus and basal insulin are added and a single continuous insulin signal is generated using an insulin-on-board model. The use of these two physiological models in this stage is beneficial not only because it smooths the original data but also because it provides values proportional to the actual behavior of the diabetic patient. Therefore, the input of the proposed GE-based tool is the output of the aforementioned physiological models in conjunction with a glucose specific fitness function and a customized grammar that represents a flexible structure, so the final solution is able to capture the patient's dynamics. The evolutionary process commences using 10 d of historical time series data for training. The algorithm builds and adjusts prediction models until it reaches a predefined number of generations. Once the final prediction models

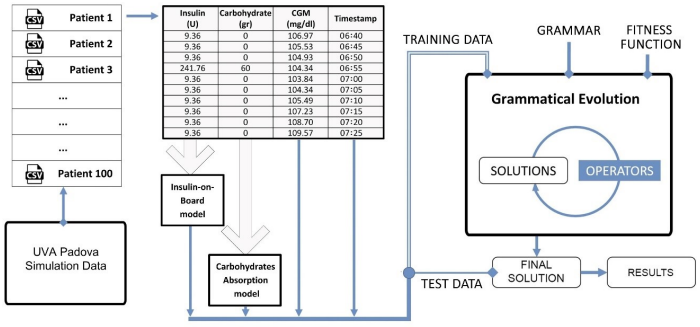

Fig 1. Schematic representation of the method used to generate prediction models for BG values.

https://doi.org/10.1371/journal.pone.0187754.g001

are generated, they are evaluated using data from the remaining historical time series data. The following subsections describe the data, the physiological models and the complete GE setup in detail.

Experimental data set

The personalized BG prediction models were built for a cohort of 100 virtual patients, the data for which were generated using the UVA/Padova T1D simulator [9] that implemented the T1D patient decision-making model [24]. That model consists of four submodels that describe the physiology of the T1D patient, the device used for glucose monitoring, the therapeutic decisions of the patient, and the insulin pump. Specifically, T1D patient physiology is defined by the UVA/Padova T1D simulator, which is a software program approved by U.S. Food and Drug Administration (FDA) as a substitute for preclinical trials for certain insulin treatments. The simulator program is based on a mathematical model of glucose, insulin, and glucagon dynamics in T1D and is equipped with a virtual population that was proven to represent inter-subject glucose variability observed in a clinical trial [25]. Recently, the simulator was updated by incorporating a model of circadian insulin sensitivity variability [26] to extend its domain of validity from single-meal to single-day multiple-meal scenarios, thus enabling more realistic in silico trials [27]. For this study, the T1D patient decision-making model was used to simulate the data sets of 100 subjects. Treatment decisions were based on the self-monitoring of blood glucose (SMBG) measurements, simulated by a model of the One Touch Ultra 2 measurement error [28], and a blinded CGM sensor, the readings of which were simulated by a model of the Dexcom G4 Platinum sensor [29]. Each virtual patient data set comprised the data of a 14-d time series of BG readings collected by the CGM sensor using a 5-min sampling period, and carbohydrate (CHO) intake and insulin delivery via an insulin pump with a 1-min sampling period.

The CHO intake (g) included the carbohydrates from three meals per day with average intakes of 50, 60, and 63.5 g for breakfast, lunch and dinner, respectively, and a coefficient of variation (CV) of 20%, sampled using a Gaussian distribution. The CHO intake time series also included 20-g hypotreatments that were generated every 20 min when the glucose concentration fell below 60 mg/dl, as indicated by the SMBG measurements.

The insulin time series was the sum of the administered basal insulin I_b dose and the bolus insulin I_{bolus} dose (U), expressed at each time step as $I(k) = I_b(k) + I_{bolus}(k)$, where k indexes the current sample. This step differed from the simulations performed in a previous study [24]

because time-varying basal insulin was used to follow the variability pattern of insulin sensitivity. Bolus doses administered at mealtime are calculated as follows:

$$I_{\text{bolus}} = \frac{CHO_{IN}}{CR} + \frac{(G_T - G_B)}{CF} \quad (1)$$

where CHO_{IN} is the estimated amount of CHO in the ingested meal, CR is the CHO-to-insulin ratio, CF is a correction factor, G_T is the glucose target, and G_B is the current preprandial SMBG measurement. CHO_{IN} is calculated by adding a percentage error in the count of CHO (sampled from a Gaussian distribution with zero mean and 20% CV) to the actual CHO content of the meal.

The virtual data used in this study were subject to great variability with respect to the T1D patients. First, the time-varying factors and perturbations implemented in the simulator allowed the use of a set of virtual patients with significant intrapatient variability. Second, the 100 virtual patients used in the simulation had distinct physical characteristics, allowing for the inclusion of interpatient variability in the data set. The modeling algorithm read the data for each patient, taken over 14 consecutive days. The CGM readings (mg/dl) were used in two ways: first, as a source of information to predict future glucose values (i.e., historical data), and second, as a reference value to train or validate the models. We used a piecewise approach in which three models predicted the postprandial BG values and one model predicted the BG value for the overnight period for each patient. As shown in the timeline in Fig 2, each day was divided into four periods of 6 h. Each period contained the same amount of data and were labeled *Nocturnal* from 01:00 to 06:59 h, *Breakfast* from 07:00 to 12:59 h, *Lunch* from 13:00 to 18:59 h, and *Dinner* from 19:00 h to 00:59 h.

The midterm prediction model was developed using simple physiological models to exploit the information contained in the insulin and CHO data series to transform the inputs into continuous signals, and a data-driven model to map past values of the model inputs to future BG values. Each part is discussed in the following two subsections.

Physiological modeling

According to a study by Hovorka et al. [30], glucose excursions are influenced by the glucose absorption process and can be represented by a two-compartment model that delivers the glucose at an absorption rate (mg/min)

$$R_a(t) = \frac{CHO_{IN} * CHO_{BIO} * t * e^{\left(-t/t_{max,G}\right)}}{{t_{max,G}}^2} \quad (2)$$

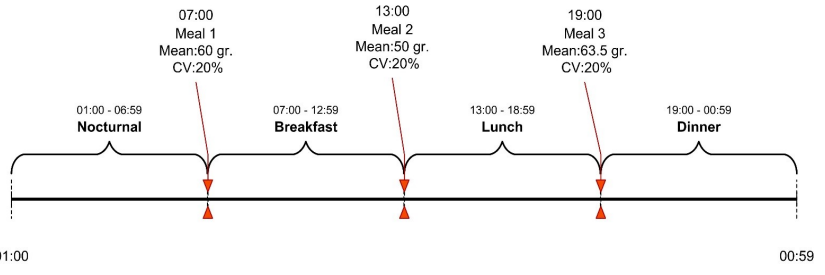

Fig 2. Division of the daily data for piecewise modeling.

https://doi.org/10.1371/journal.pone.0187754.g002

where $t_{\max,G}$ (min) is the time of the maximum appearance rate of glucose in the accessible glucose compartment, CHO_{IN} is the amount of carbohydrates ingested, and CHO_{BIO} (dimensionless) is carbohydrate bioavailability. The glucose absorption rate greatly affects the levels of BG, so the R_a signal is generated using the carbohydrate intake and population values at $t_{\max,G} = 50$ min and $CHO_{BIO} = 0.8$.

For the insulin dose, we used the estimation of insulin on board (IOB) obtained with a two-compartment model based on [31]. Smart insulin pumps can calculate the insulin that remains active within the body by using an estimation of IOB defined as follows:

$$\begin{aligned} \frac{dC_1(t)}{dt} &= u(t) - K_{DIA}C_1(t) \\ \frac{dC_2(t)}{dt} &= K_{DIA}(C_1(t) - C_2(t)) \\ I_{OB}(t) &= C_1(t) + C_2(t) \end{aligned} \tag{3}$$

where DIA is the duration of insulin action (h), which parameterizes the model of $I_{OB}(t)$ and characterizes the dynamics of insulin activity; $C_1(t)$ and $C_2(t)$ denote the compartments, $u(t)$ is the insulin dose, and the constant K replicates its corresponding DIA. For our study, we used a discrete time approximation of the model in Eq (3) with $K_{DIA} = 0.039$, which corresponds to a DIA of 2 h.

Predictive modeling by GE

Grammatical evolution (GE) [32] is a population-based heuristic search algorithm that performs an evolutionary process through selection, recombination, and mutation of a rule-based rewriting sequence on variable-length binary strings. The goal of this evolutionary computational technique is to construct syntactically correct programs that can be assessed in terms of a fitness function. The key to this construction is the grammar that allows the GE to perform a genotype-phenotype mapping process, which decodes bit strings to generate programs in an arbitrary language.

At the genotype level, the underlying genetic algorithm generates and operates the population as binary strings. The genotypes are divided into a variable number of codons, with eight binary alleles representing each codon. The mapping process involves the decoding of the genotype to its phenotype, i.e., the translation of the individual codified information into a problem-specific domain. The GE approach is an attractive method because of its flexibility and because the knowledge related to the problem can be incorporated into the algorithm using a well-structured grammar. In addition, because separate approaches are used for the search and solution spaces, the phenotype can be as complex as necessary, as all the genetic operators are applied to the genotype.

The core of context-free evolutionary grammar, usually defined in Backus normal form (BNF), is a set of derivation rules expressed in the following form:

$$[\text{symbol}] \rightarrow \{\text{production}_1 | \ldots | \text{production}_N\} \tag{4}$$

Each rule has two parts, namely, a non-terminal {symbol} on the left-hand side and a definition of the non-terminal {productions} on the right-hand side. Each definition comprises one or more alternatives separated by the symbol "|". Each alternative is commonly called a production and is composed of a sequence of terminals (bracketless) and non-terminals. Thus, the grammar indicates that a non-terminal can be substituted for any of the defined alternatives. The grammar defines the search space of solutions; thus, the quality of the obtained solutions directly depends on this structure. The framework proposed here combines insulin,

carbohydrates, and BG levels. In addition, other complex and decisive factors such as intradaily insulin sensitivity of T1D patients and the reliance of the generated models on time are considered. An excerpt summarizing the main characteristics of the defined grammar is given in Eq 5:

$$\begin{gathered}
[\mathbf{Body}] \rightarrow \mathrm{Expr}\hat{\mathrm{G}} = ([\mathbf{G}][\mathbf{op}][\mathbf{R_a}][\mathbf{op}][\mathbf{I_{OB}}])[\mathbf{op}][\mathbf{Circadian}];\\
[\mathbf{G}] \rightarrow \mathrm{GetG}([\mathbf{PrevIni}], [\mathbf{PrevFin}], [\mathbf{op}], [\mathbf{preop}])[\mathbf{G}]|\lambda\\
[\mathbf{Ra}] \rightarrow \mathrm{GetRa}([\mathbf{PrevIni}], [\mathbf{PrevFin}], [\mathbf{op}], [\mathbf{preop}])[\mathbf{Ra}]|\lambda)\\
[\mathbf{I_{OB}}] \rightarrow \mathrm{GetIOB}([\mathbf{PrevIni}], [\mathbf{PrevFin}], [\mathbf{op}], [\mathbf{preop}])[\mathbf{I_{OB}}]|\lambda)\\
[\mathbf{preop}] \rightarrow \mathrm{sqrt}|\sin|\log|\mathrm{pow}|\exp|\cos|[\mathbf{preop}][\mathbf{preop}]|\lambda\\
[\mathbf{Circadian}] \rightarrow \mathrm{GetCircadian}([\mathbf{OpB}], [\mathbf{Cte}], [\mathbf{Cte}], [\mathbf{Cte}])|\lambda\\
[\mathbf{Cte}] \rightarrow ([\mathbf{Dgt}][\mathbf{Dgt}].[\mathbf{Dgt}])\\
[\mathbf{op}] \rightarrow [\mathbf{OpA}]|[\mathbf{OpB}]\\
[\mathbf{PrevIni}] \rightarrow 0|1|2|4|6|8|10|12|14|16|18|20|22\\
[\mathbf{PrevFin}] \rightarrow 1|2|4|6|8|10|12|14|16|18|20|22|24\\
[\mathbf{Dgt}] \rightarrow 0|1|2|3|4|5|6|7|8|9\\
[\mathbf{OpA}] \rightarrow +|-\\
[\mathbf{OpB}] \rightarrow /|*
\end{gathered} \tag{5}$$

where λ denotes the empty set that does not contain any terminals. The solutions combine four expressions, namely, (**[G]**, **[Ra]**, **[I_{OB}]**, and **[Circadian]**), with four operators selected from **[Op]**. A more formal definition of these four rules in the continuous time domain can be expressed as follows:

$$\begin{aligned}
G(t) &= \sum_{i=0}^{i=n}[\mathrm{preop}(G^{\alpha}_{t-\beta})op\delta]\\
Ra(t) &= \sum_{i=0}^{i=n}[\mathrm{preop}(Ra^{\alpha}_{t-\beta})op\delta]\\
I_{OB}(t) &= \sum_{i=0}^{i=n}[\mathrm{preop}(IOB^{\alpha}_{t-\beta})op\delta]\\
\mathrm{Circadian}(t) &= A\cos(\omega t+\varphi)
\end{aligned} \tag{6}$$

where **op**{+, -, *, /} and **preop**{$\sqrt{x}$, x^{-1}, log(x), x^{y}, and sin(x) cos(x)} denote the operators selected by the GE methodology, and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{A}$ (maximum elongation), $\boldsymbol{\omega}$ (angular frequency), $\boldsymbol{\varphi}$ (initial phase), and n are constants that are adjusted by the GE methodology in each mathematical expression.

Despite knowing the impact of the inputs on the BG levels (i.e., insulin and carbohydrates have a negative effect on them), this knowledge is not directly incorporated into the initial rule. Instead, four basic operations are allowed to relate the expressions. The grammar is designed to constrain the search space by using functions that operate the previous values of the input signals. In addition to the operation of the three input variables (insulin, CHO, and BG), the sinusoidal function is added to account for the circadian variations in the physiology of patients in the final model (with maximum day-to-day variations having a 20% amplitude).

Fitness function and GE setup

In the identification and predictive approaches of the BG model, the mean squared error (MSE) is the most popular loss function and metric for assessing the performance of the model:

$$\mathrm{MSE}(g(t), \hat{g}(t,\theta)) = \frac{1}{N}\sum_{t=1}^{N}(g(t) - \hat{g}(t,\theta))^2 \tag{7}$$

where the parameter θ is selected to minimize the MSE value. MSE weights all the errors the same, even if they have different impacts in diabetes therapy. In the present study, to obtain the final model, we incorporated a fitness function based on glucose-specific MSE (gMSE), as proposed in [33]. The fitness function weights the clinical impact of errors for hypoglycemia, normoglycemia, and hyperglycemia differently. This is very beneficial to the final model since the fitness function is more sensitive to extremely harmful situations like hypoglycemia, which thereby leads to a safer model.

The fitness function used in the present study is similar to the usual quadratic MSE function. However, it includes a few additional penalties in the zones in which the error represents additional danger from a clinical perspective. For example, it is more dangerous to predict a BG concentration of 75 mg/dl when the target BG concentration actually is 50 mg/dl than it is to predict a BG concentration of 150 mg/dl when the target BG concentration actually is 175 mg/dl because missing hypoglycemic events is considerably more dangerous for the patient. The model of a glucose-specific function is

$$\mathrm{gMSE}(g, \hat{g}) = \mathrm{MSE}(g, \hat{g}) * \mathrm{Pen}(g, \hat{g}) \tag{8}$$

where $\mathrm{Pen}(g, \hat{g})$ is a function that penalizes deviation based on its clinical harmfulness and is expressed as

$$\mathrm{Pen}(g, \hat{g}) = 1 + \alpha_L \bar{\alpha}_{g \le T_L, \beta_L}(g)\alpha_{\hat{g} \ge g, \gamma_L}(\hat{g}, g) + \alpha_H \alpha_{g \ge T_H, \beta_H}(g)\bar{\alpha}_{\hat{g} \le g,\ \gamma_H}(\hat{g}, g) \tag{9}$$

where

$$\alpha_L = 1.5,\ \alpha_H = 1,\ \beta_L = 30, \beta_H = 100, \gamma_L = 10, \gamma_H = 20,\ T_L = 85, T_H = 155$$

As demonstrated in a previous study [33], the standard performance metrics are adapted using the Pen function. Therefore, we modified the fitness function using the penalization factor. This metric was also included in the report on the fitting and test deviation from the target values, as explained in the Results section.

Next, we examined the implementation of the evolutionary algorithm based on a open GE Java implementation [34]. In our approach, we used the following classic operators: elitism, variable crossover by a single point, integer flip mutation, and selection by tournament. The operator parameters are listed in Table 1. Customized genetic operators are not required in GE because it uses the standard operators of genetic algorithms [35]. The individuals are initialized randomly to generate variable-length binary strings.

Results

Performance metrics

The aim of this study was the production and assessment of personalized models for 100 virtual patients using the information contained in CGM readings, insulin dosage, and carbohydrate intake. This section presents the prediction results in terms of the usual performance metrics used to evaluate predictive accuracy and glucose-specific metrics based on the RMSE,

Table 1. General parameters of the implementation of GE and its operators.

Parameters	Value	Parameters	Value
Population	50	Tournament Size	2
Generations	2000	Max. Wraps	2
Crossover prob.	0.90	Mutation prob.	0.005
Elitism	2		

https://doi.org/10.1371/journal.pone.0187754.t001

the mean absolute deviation (MAD), and the mean absolute relative difference (MARD), as proposed in [33]. We present below the equations used to calculate the performance metrics.

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{N}\sum_{t=1}^{N}|g(t)-\hat{g}(t)|^2} \\ gRMSE\left(\frac{mg}{dl}\right) &= \sqrt{\frac{1}{N}\sum_{t=1}^{N}\mathrm{Pen}(g(t),\hat{g}(t))|g(t)-\hat{g}(t)|^2} \\ MAD(\%) &= \frac{1}{N}\sum_{t=1}^{N}|g(t)-\hat{g}(t)| \\ gMAD(\%) &= \frac{1}{N}\sum_{t=1}^{N}\mathrm{Pen}(g(t),\hat{g}(t))|g(t)-\hat{g}(t)| \\ MARD(\%) &= \frac{1}{N}\sum_{t=1}^{N}\frac{|g(t)-\hat{g}(t)|}{g(t)} \\ gMARD(\%) &= \frac{1}{N}\sum_{t=1}^{N}\frac{\mathrm{Pen}(g(t),\hat{g}(t))|g(t)-\hat{g}(t)|}{g(t)} \end{aligned} \tag{10}$$

The CEG [36] was included in the performance metrics to evaluate the clinical significance of the deviation of the estimated BG value from the target value. The CEG uses a Cartesian diagram on which the target and predicted BG values are paired. Each pair is located in one of five regions of the diagram. Region A contains those values within 20% of the reference sensor or pairs in which the predicted values and the reference values are <70 mg/dl. The pairs located in region A represent clinically correct predictions and, therefore, it is highly desirable to have all the results in this zone. Region B contains pairs by which therapy decisions made with an inaccurate estimate of the target value presents little danger. Region C contains pairs that lead to potentially dangerous overtreatment. Region D contains pairs that lead to missed severe episodes of hypoglycemia or hyperglycemia. Finally, region E contains the pairs of values that are the most different and yield the most erroneous predictions. Summarizing, pairs of points within regions A and B are clinically acceptable, while pairs in regions C, D and E are potentially dangerous and are considered significant clinical errors. Most of the CEG results are presented as a percentage of data that falls in each region relative to the total data for each case. The next subsections present and discuss the results of the evolving personalized models obtained using different approaches.

Midterm BG prediction models

A personalized piecewise model was generated for each patient by dividing the days into four 6-h segments, with three segments involving a meal and a segment corresponding to the nocturnal period when no food is ingested. For this purpose, the algorithm incorporates the input

and target values corresponding to a 6-h period and the outcome of the algorithm is set to a constant value for the timestamps that do not correspond to the segment of interest. Eq (11) presents an example of the 6-h breakfast model for Patient 2:

$$\text{Pred}_G(n) = [G(n) * Ra(n) - I_{OB}(n)] - \text{Circadian}(n) \tag{11}$$

where n represents the time step when the prediction is made and glucose, CHO, insulin, and circadian values are obtained via G, I_{OB}, R_a, and Circadian, which are defined in Eq (11) for the same example:

$$\begin{aligned} G(n) &= \frac{3.7\log(G[n-24])}{57.9} + \frac{57.7}{G[n-24]^{2.0}} + \frac{909.1}{G[n-24]^{2.7}} \\ Ra(n) &= -sin(G[n-24]^{2.7}) - 90.0 \\ I_{OB}(n) &= 2.0\ sin\left(G[n]^{4.1}\right) - \frac{4.9\ log\ (G[n-6]^{4.9})}{0.2} \\ \text{Circadian}(n) &= 19.9\ sin((909.7\pi + 5.5n/288)\pi) \end{aligned} \tag{12}$$

Because the four periods of the day are each 6 h, the overall accuracy is the mean of the performance metrics for each portion of the day. Table 2 presents the averaged individual metrics for 100 patients for the four periods and the percentage distribution among the five CEG regions for the training and the test data. Because the four segments have the same amount of information, the overall performance metrics and the percentage distribution for the CEG for a 24-h model can be reported as average values in the last row of Table 2. In addition, to show the accuracy of the results, Fig 3 presents the CEG, i.e., the distribution of the prediction error based on its clinical harmfulness, for the test data of 20 patients in the breakfast scenario.

Our study also considered another perspective with respect to the use of the evolved personalized models. Table 3 presents the metrics results for models trained to perform 4-h postprandial predictions and thereby the deviation from the target values 2 h after the intake of the meal. Therefore, 100 individualized models were evolved to predict BG values using the same virtual cohort, albeit by optimizing the predictions for 4-h periods as specified in Table 3.

Discussion

As observed in Table 2, the models produced adequate predictions for the nocturnal period in terms of RMSE, MAD, MARD, and their corresponding glucose-specific metrics gRMSE, gMAD, and gMARD, which was expected because of the lack of food intake. However, the

Table 2. Mean values of the performance metrics for 100 patients to fit 6-h prediction models.

Segment		RMSE (mg/dl)	gRMSE (mg/dl)	MAD (mg/dl)	gMAD (mg/dl)	MARD (%)	gMARD (%)	ZONE A +B (%)	ZONE A (%)	ZONE B (%)	ZONE C (%)	ZONE D (%)	ZONE E (%)
Nocturnal	Training	10.89	11.36	8.24	8.83	7.01	7.62	99.47	91.93	7.54	0.00	0.53	0.00
	Testing	11.80	12.19	9.10	9.68	7.62	8.25	99.37	90.53	8.84	0.00	0.63	0.00
Breakfast	Training	19.90	21.99	15.62	18.71	10.25	12.13	98.8	87.30	11.50	0.00	1.16	0.00
	Testing	22.09	24.60	17.39	21.11	11.50	13.78	98.68	83.38	15.30	0.00	1.26	0.00
Lunch	Training	18.33	19.90	14.45	16.58	11.12	12.97	98.32	84.50	13.82	0.01	1.66	0.00
	Testing	20.93	23.20	16.51	19.64	12.38	14.81	98.02	81.35	16.67	0.01	1.97	0.00
Dinner	Training	25.14	28.16	19.38	23.40	14.24	17.18	97.41	75.53	21.88	0.04	2.52	0.00
	Testing	29.00	33.00	22.80	27.90	16.00	19.80	97.16	70.66	26.50	0.22	2.65	0.01
24 hour	Training	18.57	20.35	14.42	16.88	10.66	12.48	98.51	84.82	13.69	0.01	1.47	0.00
	Testing	20.96	23.25	16.45	19.58	11.88	14.16	98.31	81.48	16.83	0.06	1.63	0.00

https://doi.org/10.1371/journal.pone.0187754.t002

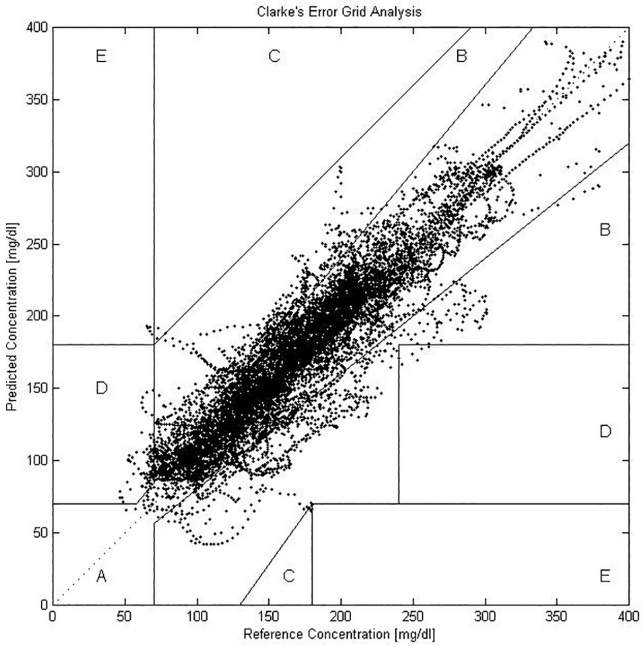

Fig 3. Clarke error grid for test data of 20 patients in the breakfast period.

https://doi.org/10.1371/journal.pone.0187754.g003

results obtained for postprandial periods were not adequate. The dinner period models had the highest mean scores for the standard and glucose-specific metrics. Therefore, these models, on average, deviated the most from the reference values. However, even for this period, more than 97% of the prediction results fell inside regions A and B for the test data, which implies that the prediction was safe from a therapeutic point of view. Regarding the general distribution of the clinical harmfulness of the deviations, it is to be noted that most of the errors out of the zones A+B are concentrated in zone D. As previously stated in the performance metrics section, having pairs located in region D is highly undesirable because it means that the prediction missed a severe hypoglycemia or hyperglycemia state. Despite always being under 4% for all the scenarios, the use of the CEG led to the identification of a specific flaw in the personalized models, especially in the model including the last meal of the day, which is the possible underestimation of the BG concentrations during hyperglycemic events and overestimation of

Table 3. Mean values of the performance metrics for 100 patients fitting 4-h prediction models.

Segment		RMSE (mg/dl)	gRMSE (mg/dl)	MAD (mg/dl)	gMAD (mg/dl)	MARD (%)	gMARD (%)	ZONE A +B (%)	ZONE A (%)	ZONE B (%)	ZONE C (%)	ZONE D (%)	ZONE E (%)
09:00–13:00	Training	16.14	17.56	12.67	14.8	8.99	10.59	98.5	89.33	9.17	0.00	1.50	0.00
	Testing	18.46	20.37	14.67	17.55	10.5	12.57	98.34	85.42	12.92	0.01	1.63	0.03
15:00–19:00	Training	14.99	16.02	11.86	13.34	9.93	11.63	97.89	86.54	11.35	0.00	2.11	0.02
	Testing	18.41	20.13	14.75	17.26	11.96	14.45	97.41	81.58	15.83	0.00	2.58	0.00
19:00–01:00	Training	19.53	21.45	15.26	17.96	12.33	14.98	96.87	81.09	15.78	0.04	3.09	0.00
	Testing	26.62	29.76	20.85	24.81	16.22	19.71	96.39	70.44	25.95	0.09	3.50	0.01

https://doi.org/10.1371/journal.pone.0187754.t003

the BG levels during hypoglycemic events. This finding highlights the importance of the use of clinical harmfulness evaluation systems like CEG to identify poor performance in terms of relevant predictions like hypo/hyperglycemic events.

Since the dinner segment is, on average, the most challenging with respect to predictions, it likely would benefit from strategies designed specifically to improve the accuracy of the fitting process. For instance, Table 3 shows the results obtained after removing the first 2 h after meals from the fitting process because that period is unpredictable (at least a 120-min prediction horizon). The averaged performance metrics and the CEG percentages in Table 3 improved for the three meals compared with the results in Table 2. The relevance of this perspective concerns the applications in risk-based advisory systems or alarm systems that allow model training in dynamic response once blood glucose is expected to decay. Conversely, the full-segments perspective is justified by the empirical evidence, which shows that predicting the dynamics of the 2-h period immediately after meal intake is considerably more complex than predicting the BG level when the meal effect vanishes. To make decisions with respect to insulin therapy or the rescue of carbohydrate ingestion, the accuracy of the algorithm is favored by the evolution of specific prediction models for the postprandial period.

Antecedents of the approach presented in this paper are a study that assessed the feasibility of GE prediction systems based on a time series of historical prices [37], and the first study that adopted an approach toward personalized BG predictions using GE [38]. In contrast with previous GE approaches [22][23] that were limited to short-term predictions (>60 min) and were tested with five and eight virtual patients, respectively, obtained from the AIDA simulator, the present study was built and tested for midterm predictions (120 min) and in a more robust manner by using the UVA/Padova T1D Simulator [9], which described both intra- and interpatient glucose variability [26] implemented in the T1D patient decision-making model [39].

In general, recent studies usually predicted T1D glucose for the next 30 min [21]. However, some studies have reported on the evaluation of a 120-min prediction horizon. For instance, Georga et al. [40][41] assessed support vector regression and random forests methods using data from 15 patients. In both cases, the best performance metrics were obtained when using CGM (mg/dl), plasma insulin concentration (μU/ml), instantaneous energy expenditure, meal-derived glucose rate of appearance, and R_a (mg/min) as inputs. This resulted in an RMSE of 7.62 mg/dl using the SMV approach and 10.83 mg/dl using random forests. Likewise, Zarkogianni et al. [5] compared several data-based techniques using as inputs the most recent glucose measurement $G(t)$, the change in glucose level $\Delta G(t)$, and the sum of energy expenditures during the last 30 min. This resulted in a technique based on the self-organizing map that yielded the most favorable results in terms of RMSE (31.00 ± 6.07) and MARD (14.56 ± 3.46) for a cohort of ten patients. Aside from using physical activity as an additional input signal, these approaches generated models that were trained and evaluated using the same data set and an independent data set was not used for validation.

Leal et al. [42] used support vector regression to predict nocturnal glucose, using CGM and insulin delivery information to produce individual models for the same 100 in silico T1D adults used in this study [9]. The results for four simulated night periods between 1 and 7 am were measured in terms of glucose-specific metrics. According to the means of those test metrics (RMSE = 15.0, gRMSE = 15.7, MAD = 10.9, gMAD = 11.7, MARD = 9.0, gMARD = 9.6), the predictions achieved by the proposed approach presented in this paper were more accurate. Moreover, in this paper, the clinical reliability of the predictions was evaluated in both the modeling process and the outcomes.

Conclusion and future work

The results of this simulated study of both experimental approaches described in the previous section are promising. They confirm the assumption that the use of a glucose-specific cost function that takes into account the clinical harmfulness of deviations makes the prediction models more reliable in terms of clinical usefulness. In addition, these results suggest that dividing the day into different segments that can be studied separately and generating piecewise models improve the accuracy and clinical reliability of the overall model. These results also show that the performance and safety of the predictions can be improved further by generating a set of interchangeable models that predict useful BG values for control and therapy purposes based on the determination of individual specific dynamics, lifestyle, and other factors.

The ability to provide an early warning of ineffective or poor insulin treatment, which usually leads to hyperglycemic or hypoglycemic episodes, is of great interest. Useful real-time predictions of future CGM measurements are possible but challenging owing to various factors, including variability and the associated delays of food and insulin absorption. In addition, there is a 10–15-min lag between the actual blood plasma values and the sensor measurements, resulting in an approximate mean absolute relative difference (MARD) of 9% [3] for the best sensors.

Despite being limited by these delays, accurate forecasts can provide enough time to act in anticipation of the CGM measurements to prevent hyperglycemia or hypoglycemia. In this study, we developed a hybrid model that uses GE, insulin on board, and glucose rate of absorption models to predict BG values with a prediction horizon of 120 min. The algorithm relies on the construction of a set of rules that determine the search space for an optimization algorithm based on GE. A glucose-specific fitness function leads the evolution of the solution while penalizing deviations based on their clinical harmfulness and a tailored evolutionary grammar.

Our study proposed a hybrid GE and physiological model-based methodology for determining personalized midterm predictions of CGM readings. To the best of our knowledge, this is the first methodology to include physiological models in the overall GE model.

Future work will address the following points:

- The FDA-approved simulator used in this work is a valid substitute for the preclinical testing of novel technologies in diabetes care (see, e.g., [43][44][45]). However, the use of in silico data is not meant as substitute to human trial; rather, it can be considered a good starting point for evaluating our GE method: in fact simulated data are complete, i.e. there is no missing information related to meals, boluses, hypotreatments, or any other unexpected event. The natural extension of this work will be testing personalization of BG prediction models in a more challenging situation involving real subjects.
- Estimation of BG values can be automatically processed ahead of time to generate risk-based predictions. In addition, the risk of life-threatening events can be incorporated directly into the fitness function.
- There are few conclusive reports on exercise management with respect to the prediction model. We will examine the manner in which input signals related to physical exercise can improve or deteriorate the accuracy of personalized models.
- The grammar shapes the solution. For our approach, other grammar architectures can be explored to improve the accuracy and flexibility of the patient model. Future studies will include a comparative analysis that explores the decrease and increase in grammar complexity.

Author Contributions

Conceptualization: Iván Contreras, Josep Vehí.

Data curation: Martina Vettoretti, Roberto Visentin.

Formal analysis: Iván Contreras, Silvia Oviedo, Josep Vehí.

Funding acquisition: Iván Contreras, Silvia Oviedo, Josep Vehí.

Investigation: Iván Contreras, Silvia Oviedo.

Methodology: Iván Contreras.

Project administration: Josep Vehí.

Resources: Martina Vettoretti, Roberto Visentin.

Software: Iván Contreras, Silvia Oviedo.

Supervision: Iván Contreras, Silvia Oviedo, Josep Vehí.

Validation: Iván Contreras, Silvia Oviedo.

Writing – original draft: Silvia Oviedo.

Writing – review & editing: Iván Contreras, Silvia Oviedo, Martina Vettoretti, Roberto Visentin, Josep Vehí.

References

1. Cobelli C, Renard E, Kovatchev B. Artificial pancreas: Past, present, future. Diabetes. 2011; 60: 2672–2682. https://doi.org/10.2337/db11-0654 PMID: 22025773
2. Hinshaw L, Dalla Man C, Nandy DK, Saad A, Bharucha AE, Levine JA, et al. Diurnal pattern of insulin action in type 1 diabetes: implications for a closed-loop system. Diabetes. American Diabetes Association; 2013; 62: 2223–9. https://doi.org/10.2337/db12-1759 PMID: 23447123
3. Bailey TS, Chang A, Christiansen M. Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm. J Diabetes Sci Technol. Diabetes Technology Society; 2015; 9: 209–14. https://doi.org/10.1177/1932296814559746 PMID: 25370149
4. Henry R, Ieee B, Member S, Daskalaki E, Member I, Diem P, et al. Multi-model data fusion to improve an early warning system for hypo-/hyperglycemic events. Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2014; 2014: 4843–6. https://doi.org/10.1109/EMBC.2014.6944708 PMID: 25571076
5. Zarkogianni K, Mitsis K, Litsa E, Arredondo M-T, Fico G, Fioravanti A, et al. Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring. Med Biol Eng Comput. 2015; 53: 1333–43. https://doi.org/10.1007/s11517-015-1320-9 PMID: 26049412
6. Cescon M, Johansson R, Renard E. Subspace-based linear multi-step predictors in type 1 diabetes mellitus. Biomed Signal Process Control. 2015; 22: 99–110. https://doi.org/10.1016/j.bspc.2014.09.012
7. Zarkogianni K, Litsa E, Vazeou A, Nikita KS. Personalized glucose-insulin metabolism model based on self-organizing maps for patients with Type 1 Diabetes Mellitus. 13th IEEE International Conference on BioInformatics and BioEngineering. IEEE; 2013. pp. 1–4. https://doi.org/10.1109/BIBE.2013.6701604
8. Dassau E, Cameron F, Bequette BW, Zisser H, Jovanovič L, Chase HP, et al. Real-Time Hypoglycemia Prediction Suite Using Continuous Glucose Monitoring. Diabetes Care. 2010; 33: 1249–1254. https://doi.org/10.2337/dc09-1487 PMID: 20508231
9. Man CD, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C. The UVA/PADOVA Type 1 Diabetes Simulator: New Features. J Diabetes Sci Technol. 2014; 8: 26–34. https://doi.org/10.1177/1932296813514502 PMID: 24876534
10. Fernandez M, Villasana M, Streja D. Glucose dynamics in Type I diabetes: Insights from the classic and linear minimal models. Comput Biol Med. 2007; 37: 611–627. https://doi.org/10.1016/j.compbiomed.2006.05.008 PMID: 16867301

11. Ghosh S. A differential evolution based approach for estimating minimal model parameters from IVGTT data. Comput Biol Med. 2014; 46: 51–60. https://doi.org/10.1016/j.compbiomed.2013.12.014 PMID: 24529205
12. Fong S, Mohammed S, Fiaidhi J, Kwoh CK. Using causality modeling and Fuzzy Lattice Reasoning algorithm for predicting blood glucose. Expert Syst Appl. 2013; 40: 7354–7366. https://doi.org/10.1016/j.eswa.2013.07.035
13. Buckingham B, Chase HP, Dassau E, Cobry E, Clinton P, Gage V, et al. Prevention of Nocturnal Hypoglycemia. Diabetes Care. 2010; 33: 1013–1017. https://doi.org/10.2337/dc09-2303 PMID: 20200307
14. Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. Diabetes Technol Ther. 2013; 15: 792–801. https://doi.org/10.1089/dia.2013.0104 PMID: 23883406
15. Lu Y, Gribok A V., Ward WK, Reifman J. The importance of different frequency bands in predicting subcutaneous glucose concentration in type 1 diabetic patients. IEEE Trans Biomed Eng. 2010; 57: 1839–1846. https://doi.org/10.1109/TBME.2010.2047504 PMID: 20403780
16. Novara C, Pour NM, Vincent T, Grassi G. A Nonlinear Blind Identification Approach to Modeling of Diabetic Patients. Proc 19th World Congr Int Fed Autom Control. 2015; 1–9.
17. Fernandez de Canete J, Gonzalez-Perez S, Ramos-Diaz JC. Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes. Comput Methods Programs Biomed. 2012; 106: 55–66. https://doi.org/10.1016/j.cmpb.2011.11.006 PMID: 22178070
18. Balakrishnan NP, Rangaiah GP, Samavedham L. Personalized blood glucose models for exercise, meal and insulin interventions in type 1 diabetic children. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2012. pp. 1250–1253. 10.1109/EMBC.2012.6346164
19. Estrada GC, Kirchsteiger H, Eric R. Innovative Approach for Online Prediction of Blood Glucose Profile in Type 1 Diabetes Patients. Am Control Conf (ACC), 2010. 2010; 2015–2020.
20. Zecchin C, Facchinetti A, Sparacino G, Cobelli C. Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information. Comput Methods Programs Biomed. Elsevier Ireland Ltd; 2014; 113: 144–152. https://doi.org/10.1016/j.cmpb.2013.09.016 PMID: 24192453
21. Oviedo S, Vehi J, Calm R, Armengol J. A REVIEW OF PERSONALIZED BLOOD GLUCOSE PREDICTION STRATEGIES FOR T1DM PATIENTS. Int j numer method biomed eng. 2016; https://doi.org/10.1002/cnm.2833 PMID: 27644067
22. Hidalgo JI, Colmenar JM, Risco-Martin JL, Cuesta-Infante A, Maqueda E, Botella M, et al. Modeling glycemia in humans by means of Grammatical Evolution. Appl Soft Comput. 2014; 20: 40–53. https://doi.org/10.1016/j.asoc.2013.11.006
23. Hidalgo JI, Maqueda E, Colmenar JM, Botella M, Risco-martín JL, Cuesta-infante A, et al. Clarke and Parkes Error Grid Analysis of Diabetic Glucose Models obtained with Evolutionary Computation. Proc 2014 Conf companion Genet Evol Comput companion—GECCO Comp '14. 2014; 1305–1312.
24. Vettoretti M, Facchinetti A, Sparacino G, Cobelli C. Patient decision-making of CGM sensor driven insulin therapies in type 1 diabetes: In silico assessment. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2015. pp. 2363–2366. 10.1109/EMBC.2015.7318868
25. Visentin R, Dalla Man C, Kovatchev B, Cobelli C. The university of Virginia/Padova type 1 diabetes simulator matches the glucose traces of a clinical trial. Diabetes Technol Ther. Mary Ann Liebert, Inc.; 2014; 16: 428–34. https://doi.org/10.1089/dia.2013.0377 PMID: 24571584
26. Visentin R, Dalla Man C, Kudva YC, Basu A, Cobelli C. Circadian variability of insulin sensitivity: physiological input for in silico artificial pancreas. Diabetes Technol Ther. Mary Ann Liebert, Inc.; 2015; 17: 1–7. https://doi.org/10.1089/dia.2014.0192 PMID: 25531427
27. Visentin R, Man CD, Cobelli C. One-Day Bayesian Cloning of Type 1 Diabetes Subjects: Toward a Single-Day UVA/Padova Type 1 Diabetes Simulator. IEEE Trans Biomed Eng. 2016; 63: 2416–2424. https://doi.org/10.1109/TBME.2016.2535241 PMID: 26930671
28. Vettoretti M, Facchinetti A, Sparacino G, Cobelli C. A Model of Self-Monitoring Blood Glucose Measurement Error. J Diabetes Sci Technol. 2017;Epub ahead. https://doi.org/10.1177/1932296817698498 PMID: 28299958
29. Facchinetti A, Del Favero S, Sparacino G, Cobelli C. Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices. Med Biol Eng Comput. 2015; 53: 1259–1269. https://doi.org/10.1007/s11517-014-1226-y PMID: 25416850
30. Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Federici MO, et al. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas. 2004; 25: 905–920. https://doi.org/10.1088/0967-3334/25/4/010 PMID: 15382830

31. Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R. Insulin kinetics in type-I diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Trans Biomed Eng. 2005; 52: 3–12. https://doi.org/10.1109/TBME.2004.839639 PMID: 15651559

32. O'Neill M, Ryan C. Grammatical evolution. IEEE Trans Evol Comput. 2003; 5: 349–358. https://doi.org/10.1109/4235.942529

33. Favero S Del, Facchinetti A, Cobelli C. A glucose-specific metric to assess predictors and identify models. IEEE Trans Biomed Eng. 2012; 59: 1281–1290. https://doi.org/10.1109/TBME.2012.2185234 PMID: 22275716

34. GE4T1D: Grammatical Evolution for T1D [Internet]. Modeling, Identification and Control Engineering (MICELab); 2017. Available: https://github.com/IvanContrerasFD/MiceLab-Grammatical-Evolution-and-Type-1-Diabetes

35. Ryan C, Collins J, O'Neil M. Grammatical Evolution: Envolving Programs for an Arbitrary Language. Springer Berlin Heidelb. 1998; 83–96. https://doi.org/10.10007/BFb0055930

36. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. Diabetes Care. American Diabetes Association; 1987; 10: 622–628. https://doi.org/10.2337/diacare.10.5.622 PMID: 3677983

37. Contreras I., Hidalgo J. Ignacio N ez Z. a hybrid automated trading system based on multi-objective grammatical evolution. J Intell Fuzzy Syst. 2016;

38. Contreras, I., Vehi J. Mid-term prediction of blood glucose from continuous glucose sensors, meal information and administered insulin. XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016: MEDICON 2016. Springer; 2016.

39. Vettoretti M, Facchinetti A, Sparacino G, Cobelli C. Type 1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments. IEEE Trans Biomed Eng. 2017; 1–1. https://doi.org/10.1109/TBME.2017.2746340 PMID: 28866479

40. Georga EI, Protopappas VC, Ardigo D, Marina M, Zavaroni I, Polyzos D, et al. Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression. Biomed Heal Informatics, IEEE J. 2013; 17: 71–81. https://doi.org/10.1109/titb.2012.2219876 PMID: 23008265

41. Georga EI, Protopappas VC, Polyzos D, Fotiadis DI. A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests. Conf Proc IEEE Eng Med Biol Soc. 2012; 2012: 2889–92. https://doi.org/10.1109/EMBC.2012.6346567 PMID: 23366528

42. Y. Leal, L. Gonzalez-Abril, R. Visentin, S. Del Favero, M. Vettoretti, A. Facchinetti, G. Sparacino CC. Support Vector Regression for Mid-Term Nocturnal Glucose Prediction from Continuous Glucose Monitoring and Insulin Delivery Information. Poster Present 9th Int Conf Adv Technol Treat Diabetes (ATTD), Diabetes Technol Ther. Poster presented at: 9th International Conference on Advanced Technologies & Treatments for Diabetes (ATTD), At Diabetes Technology and Therapeutics; 2016; 10.1089/dia.2016.2526

43. Visentin R, Giegerich C, Jäger R, Dahmen R, Boss A, Grant M, et al. Improving Efficacy of Inhaled Technosphere Insulin (Afrezza) by Postmeal Dosing: In-silico Clinical Trial with the University of Virginia/Padova Type 1 Diabetes Simulator. Diabetes Technol Ther. Mary Ann Liebert, Inc.; 2016; 18: 574–85. https://doi.org/10.1089/dia.2016.0128 PMID: 27333446

44. Toffanin C, Visentin R, Messori M, Di Palma F, Magni L, Cobelli C. Towards a Run-to-Run Adaptive Artificial Pancreas: In Silico Results. IEEE Trans Biomed Eng. 2017;PP: 1–1. https://doi.org/10.1109/TBME.2017.2652062 PMID: 28092515

45. Edelman S V. Regulation Catches Up to Reality. J Diabetes Sci Technol. 2017; 11: 160–164. https://doi.org/10.1177/1932296816667749 PMID: 27630249

Part III

RISK-BASED POSTPRANDIAL HYPOGLYCEMIA FORECASTING USING SUPERVISED LEARNING

Submitted to International Journal of Medical Informatics. Sept,2018 (JCR quartile: Q1; JIF: 2.957 in 2017; Ranked 6/25 in Medical Informatics).

Risk-based Postprandial Hypoglycemia Forecasting Using Supervised Learning

Silvia Oviedo[a], Ivan Contreras[a], Carmen Quirós[b], Marga Giménez[c,d], Ignacio Conget[c,d], Josep Vehi[a,d,*]

[a]*Institut d'Informatica i Aplicacions. Universitat de Girona, Spain*
[b]*Servicio de Endocrinología y Nutrición. Hospital Universitari Mutua de Terrassa, Terrassa, Spain*
[c]*Diabetes Unit. Endocrinology and Nutrition Dpt. IDIBAPS (Institut d'investigacions biomdiques August Pi I Sunyer), Barcelona, Spain*
[d]*Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Spain*

Abstract

Background Predicting insulin-induced postprandial hypoglycemic events is critical for the safety of type 1 diabetes patients because an early warning of hypoglycemia facilitates correction of the insulin bolus before its administration. The postprandial hypoglycemic event counts can be lowered by reducing the size of the bolus based on a reliable prediction but at the cost of increasing the average blood glucose.

Methods We developed a method for predicting postprandial hypoglycemia using machine learning techniques personalized to each patient. The proposed system enables on-line therapeutic decision making for patients using a sensor augmented pump therapy. Two risk-based approaches were developed for a window of 240 min after the meal/bolus, and they were tested based on real retrospective data from 10 patients using 70 mg/dL and 54 mg/dL as thresholds according to the consensus for Level 1 and Level 2 hypoglycemia, respectively. Due to the small size of the patient cohort, we trained personalized models for each patient.

Results The median specificity and sensitivity were 79% and 71% for Level 1 hypoglycemia, respectively, and 81% and 77% for Level 2.

Conclusions The results demonstrated that it is feasible to anticipate hypoglycemic events with a reasonable false-positive rate. The accuracy of the results and the trade-off between performance metrics allow its use in decision support systems for patients who wear insulin pumps.

Keywords: blood glucose, bolus calculation, hypoglycemia prediction, machine learning, postprandial hypoglycemia, Type 1 diabetes

1. Introduction

Type 1 diabetes (T1D) is a chronic condition that affects the pancreas, which is characterized by an autoimmune response where the insulin-producing cells are destroyed, thereby resulting in insufficient or a total absence of insulin production. Iatrogenic hypoglycemia causes recurrent acute and chronic morbidity in most people with T1D and it is a huge burden for people with this disease. Frequent and repeated episodes of hypoglycemia generally result in a reduced ability or failure to recognize the symptoms and signs of hypoglycemia. In addition, hypoglycemia is a major barrier to achieving normoglycemia over a lifetime of using intensive insulin therapy, thereby precluding the long-term benefits of euglycemia [1]. Recently, a consensus was reached regarding clinically meaningful outcomes in the development and evaluation of T1D therapies, where the following three levels of hypoglycemia were defined according to the plasma glucose levels[2].

*Josep Vehí, Universitat de Girona, Modeling, Identification and Control Engineering (MICELab), Institut d'Informàtica i Aplicacions, Girona, Girona, 17003, Spain

Email addresses: silvia.oviedo@udg.edu (Silvia Oviedo), ivancontrerasfd@gmail.com (Ivan Contreras), cmquiros@clinic.cat (Carmen Quirós), gimenez@clinic.cat (Marga Giménez), iconget@clinic.cat (Ignacio Conget), josep.vehi@udg.edu (Josep Vehi)

- Level 1: $Glucose < 70$ mg/dL (3.9 mmol/L) and $Glucose \geq 54$ (3.0 mmol/L).
- Level 2: $Glucose < 54$ mg/dL (3.0 mmol/L).
- Level 3: A severe event characterized by altered mental and/or physical status requiring assistance.

Tightening the glycemic normal range for T1D patients can lead to an increased risk of hypoglycemia[3], and thus it is essential to anticipate this risk. Therefore, anticipating the risk of hypoglycemia at every meal is a helpful feature of a CGM prediction system based on data obtained by continuous glucose monitoring (CGM) for patients who wear insulin pumps or those receiving therapy with multiple doses of insulin (MDI). Consequently, CGM prediction is a major focus of the T1D research community where there are two main application areas. The first application area is closed-loop control, where a model predicts the CGM values and a controller generates a control action based on the predictions. The second application area is in decision-support systems or open-loop applications, where a given model is set to predict the future CGM levels or associated events, and the predicted outcome is then transformed into an alarm, a dosage correction, or some other interpretable outcome for decision making, such as insulin suspension or bolus re-estimation. Newer models of insulin pumps can predict hypoglycemia 30 min in advance and suspend the delivery of insulin in order to reduce the frequency of hypoglycemic episodes and the hypoglycemic intensity[4].

A previous comprehensive review of models for predicting CGM using CGM data[5] showed that less than 15% of previous studies aimed to predict glycemic events or risks, despite their critical roles in therapeutic decisions. Interestingly, more than 50% of the risk/event prediction models were developed as direct classification problems, and thus their predictions of events such as hypoglycemia and hyperglycemia did not employ a prior estimation of the CGM level. Models that focus strictly on predicting particular outcomes such as hypoglycemic events are meaningful because they can be employed in decision-support applications. However, these approaches are mostly limited to nocturnal hypoglycemia [6] [7] [8] [9], where the effects of disturbances such as meals or exercise are minimal, and thus prediction is less challenging.

In the present study, we focused on the patient-specific prediction of hypoglycemic events when a meal is announced, which allows the evaluation of the impact of a given insulin bolus on the postprandial response and optimization to achieve safer dosages.

This study makes two main contributions. First, we approach the problem of predicting hypoglycemia from a classification perspective, which has previously been restricted mainly to nocturnal hypoglycemia and shorter prediction windows. Second, the prediction scheme is specialized for postprandial hypoglycemia, which allows predictions of the effect of the insulin dosage for a given meal, thereby facilitating on-line therapy decision making, such as bolus re-estimation for insulin pump users.

2. Methods

Next, we present the process employed to generate a risk-based prediction model of hypoglycemia. The core of the model is a bi-class support vector classifier (SVC) trained and tested using scikit-learn[10][11]. The SVC can deal with unbalanced classes using a class weight (CW) parameter that places more emphasis on a class by penalizing the mislabeled classes. This is done by setting the regularization parameter C of each class to C^*CW. Therefore, no synthetic data are used and oversampling is not performed in the training set. Thus, the training and testing sets belong to the same distribution, and the cost function penalizes the misclassification of the minority class in proportion to the class imbalance.

2.1. Patient Cohort and Data Preprocessing

We collected retrospective data from T1D patients at Hospital Clínic i Universitari in Barcelona who had used insulin pumps and CGM for several months. In total, 10 patients comprising eight males and two females were included in the analysis. The study was conducted under free living conditions. The patients were not asked to deliver physical activity, diet or other specific information. The study was restricted to the adult population with a mean (±SD) age of 41 ± 10 years. The average duration of diabetes in the population was 27 ± 10 years and they had been insulin pump therapy users for 10 ± 5 years. The average body weight and HbA_{1c} were 65 ± 13 kg and 7.3 ± 0.5%, respectively. The mean (±SD) monitoring period was 786 ± 263 days. Patients wore Paradigm Veo or 640G insulin pumps (Medtronic MiniMed, Northridge, CA, USA). Some of them upgraded from one to the other during the study. We adapted a particular routine for each pump model to extract the critical data into a csv file containing the dates, time-stamps, delivered insulin, carbohydrate consumption, and CGM signal for each patient. CGM was conducted with Enlite-2 sensors (Medtronic MiniMed, Northridge, CA, USA). After extracting the

data into a csv file, the next step aimed to assure the data integrity by cleaning and organizing the data in a dataframe with the relevant features as columns. A script Java was used to extract the data from the pump output files, and we employed Python to generate the prediction models and performance metrics, specifically the Scikit-learn[10], Pandas[12], and Matplotlib[13] packages.

2.2. Support Vector Machines for Bi-class Classification

Support vector machines are widely used supervised machine learning algorithms that model a separating hyperplane in a multidimensional space to solve a given classification task. Given a training set of N instances of features $\boldsymbol{x}$ and their corresponding labels $y \in -1, 1$, the parameters ω, b are used to model the classifier as[14]:

$$h_{\omega,b}(x) = g(\omega^T x + b), \tag{1}$$

where $g(z) = 1$ if $z \geq 0$ and $g(z) = -1$ otherwise. The geometric margin can be understood as the distance from a certain training example $(x^{(i)}, y^{(i)})$ to the decision boundary, and it can be defined by equation 2, as follows.

$$\gamma^{(i)} = \left(\frac{\omega}{||\omega||}^T x^{(i)} + \frac{b}{||\omega||} \right) \tag{2}$$

The optimal margin classifier is determined by solving equation 3:

$$min_{\omega,b} \frac{1}{2}(\omega \bullet \omega), \tag{3}$$

Where $\bullet$ represents the scalar product of two vectors. Equation 3 is subject to the restriction that $y_i(\omega.x_i + b) \geq 1, i = 1, \ldots, m$. When the classes are not separable, the loss function becomes 4 [15]:

$$J(\omega, b, \xi) = \frac{1}{2}||\omega||^2 + C \sum_{i=1}^{N} \xi_i \tag{4}$$

subject to

$$y_i[(\omega.x_i + b) \geq 1 - \xi_i, i = 1, 2 \ldots, N \tag{5}$$

and

$$\xi_i \geq 0, i = 1, 2 \ldots, N \tag{6}$$

Where ξ_i are called *slack variables* and $C > 0$ is the *regularization parameter* that controls the trade-off between the generalization capability and the training error.

If the training data are not linearly separable, then it is possible to transform the data set into a new dimensional space, where the data can be linearly separable [17]. For this purpose, a mapping function $\phi(\bullet)$ is defined in terms of the scalar product in the original space. Instead of defining the transformation function, a kernel function $K(u \bullet v)$ is specified. A kernel function performs the space transformation and the scalar product in one step. This application uses one of the most common kernels, i.e., the radial basis function, which defines the scalar product between two feature vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ according to equation 7.

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2}, \gamma > 0 \tag{7}$$

The parameter γ must be selected in an appropriate manner according to the application. Tuning of this parameter and the other hyper-parameters is conducted by grid-search with cross-validation.

2.3. Model for Prediction of Hypoglycemia

We built a personalized model for each patient by randomly dividing the total instances into training (80%) and test (20%). The number of instances available for training and testing are listed in Table 1.

Table 1: Training and testing instances for 10 patients.

	Train (80%)	Test (20%)	Total instances
P1	359	90	449
P2	1832	458	2290
P3	556	140	696
P4	712	179	891
P5	565	142	707
P6	198	50	248
P7	448	113	561
P8	416	105	521
P9	207	52	259
P10	472	118	590

An instance was defined by a meal intake provided that no other meal was consumed during the subsequent 4 h. This was set as a requirement because the prediction window of interest in this study was 4 h after every meal. In addition, the instances with missing sensor data were excluded.

The labeling of the outcome for each instance was set differently using two thresholds according to the definitions of Level 1 and Level 2 hypoglycemia[2]. In the first case, C1, a hypoglycemia event was predicted whenever three consecutive CGM measurements (15 min) were below 70 mg/dL. In the second case, C2, a

hypoglycemia event was predicted whenever three consecutive CGM measurements (15 min) were below 54 mg/dL. Postprandial hypoglycemia under 70 mg/dL and 54 mg/dL occurred with a mean of 29% and 17% of the meals, respectively.

Alternatively, we proposed another labeling scheme based on the hypoglycemia risk as the area below 70 mg/dL and the CGM curve. After the risk was calculated for every instance, six additional cases were defined. As summarized in Table 2, a hypoglycemic event was predicted if the risk was above a specific value. Finally, if the duration of pump suspension (PS) was longer than 70 min, the label was also set as hypoglycemic event in all cases. For model evaluation purposes, we will use Level 1 and Level 2 as labels throughout the paper, bearing in mind that in this approach, they are not mutually exclusive, as they correspond to C1 and C2, respectively.

Table 2: Labeling scheme for postprandial hypoglycemic event prediction.

	Postprandial hypoglycemia
C1	Three consecutive $CGM < 70$ mg/dL or PS > 70 min
C2	Three consecutive $CGM < 54$ mg/dL or PS > 70 min
C3	Risk > 200 or PS > 70 min
C4	Risk > 500 or PS > 70 min
C5	Risk > 700 or PS > 70 min
C6	Risk > 1000 or PS > 70 min
C7	Risk > 1200 or PS > 70 min
C8	Risk> 1500 or PS > 70 min

Using the CGM data, several time-domain features were extracted and analyzed by training various SVC configurations. The following nine features were selected for the prediction architecture, where k represents the sample when the meal was recorded.

1. *Glucose difference (GR)*: Absolute value of the difference between the last and first CGM values in the last hour, *i.e.*, $GR = abs(CGM[k-12] - CGM[k])$.
2. *CGM[k]*: CGM glucose value at k.
3. AUC_{1h}: Area under the curve for a CGM threshold of 70 mg/dL and the CGM signal over the last hour.
4. *Glucose Rate of Change*: Glucose rate of change over the last 30 min.
5. *Mean Glucose*: Mean value of glucose signal in the last hour.
6. *Cumulated Basal (CB)*: Cumulative sum of the basal insulin over the last 2 h *i.e.*, *CB = B(k)+...+B(k-24)*.
7. *Forecast Basal (FB)*: Cumulative sum of the programed basal insulin for the next 4 h *CB = B(k)+...+B(k+48)*.
8. *Bolus*: Cumulative sum of the bolus insulin up to 1 h after its administration (to consider square and dual wave boluses).
9. *Carbohydrate*: Carbohydrate intake.

The general flow of the method is presented in Figure 1. After data preprocessing, each SVC prediction model required the optimization of the hyperparameter C and the kernel parameter γ by via grid search using stratified fivefold cross-validation and MCC as a scorer. After fixing the optimized hyperparameters for each of the eight approaches (C1–C8) in Table 2, the models were tested for every patient. Stratification based on cross-validation ensured that each fold represented the overall distribution of the data. Therefore, both classes were considered to be similarly represented in every fold. In addition, the following two tasks were performed in order to test the robustness of the models. First, we randomly shuffled the data five times. Second, we trained and tested five different models with 80% and 20%. The calculated average performance metrics are reported in the Results section. It should be noted that the cross-validation and testing data subsets were identical across cases (C1–C8).

2.4. Performance Metrics

The domain considered in this study is highly imbalanced because hypoglycemia must be avoided by patients by ingesting carbohydrates, suspending the insulin dose, or lowering the basal dose. Therefore, accuracy is a poor metric choice in this case. In contrast, the sensitivity (SE) and specificity (SP) are not sensitive to the class distribution and they have been used extensively in previous hypoglycemia prediction studies[16][17][18]. The Matthews correlation coefficient (MCC) is used to merge the confusion matrix into a single metric that correlates the target and the predicted binary outcomes, which is desirable because it allows a unique metric to be optimized when training each model. The MCC index returns a value in the range of $[-1, 1]$, where 1 is a perfect prediction and –1 is an erroneous classification. These metrics are defined in Table 3. The MCC was used in the hyper-parameter tuning grid search and the selection of the best model from C1-C8.

3. Results

The average results including SP (%), SE (%), and their corresponding standard deviations as well as the

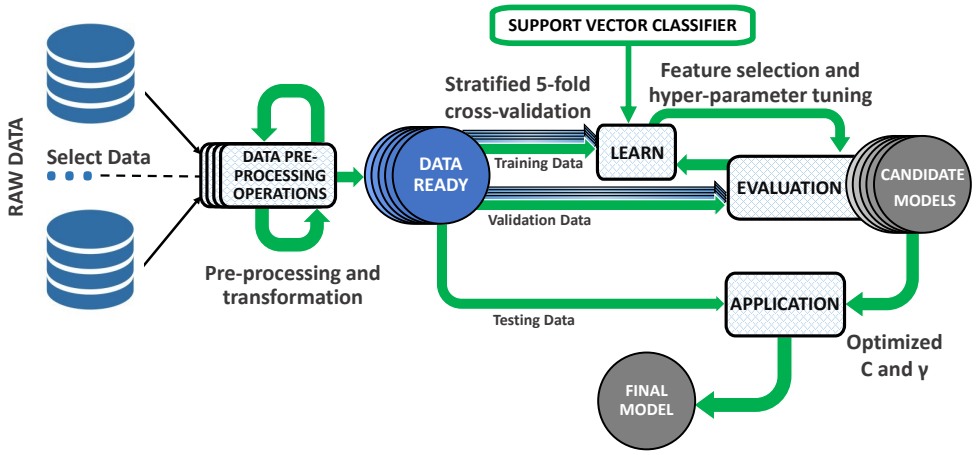

Figure 1: General diagram of the proposed method.

Table 3: Metrics used for candidate model evaluation.

Metric	Definition
TP	Positive example classified as positive
TN	Negative example classified as negative
FP	Negative example classified as positive
FN	Positive example classified as negative
SE	Ratio of positives that are correctly classified: $TP/(TP+FN)$
SP	Ratio of negatives that are correctly classified: $TN/(TN+FP)$
MCC	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

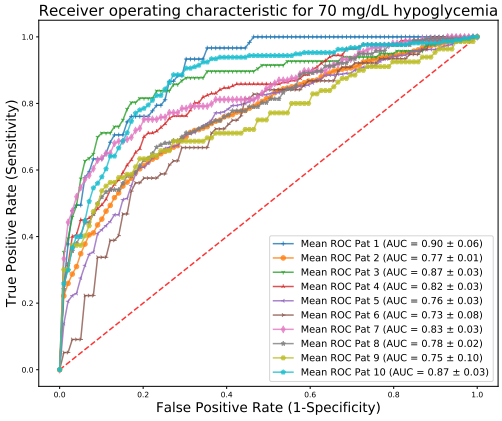

Figure 2: ROC curve for prediction results for Level 1 hypoglycemia

CMM for the 10 subjects are presented in Table 4 to Table 7. The distribution of the averaged test results in the Receiver Operating Characteristic (ROC) for Level 1 and Level 2 is presented in Fig2 and Fig3, respectively. Table 4 and Table 5 show the average testing results obtained with the labels generated according to the Level 1 condition (Test A), while Table 6 and Table 7 present the average results for the individual models in cases C1–C8 tested with the labels generated according to the Level 2 condition (Test B). For the best models, the median SP and SE were 79% and 71% for Level 1 hypoglycemia, respectively, and the median SP and SE were 81% and 77% for Level 2, as listed in Table 8.

According to the results summarized in Table 4, as expected, the models trained specifically for predicting hypoglycemic events according to the Level 1 hypoglycemia condition (C1) obtained satisfactory performance in terms of the SE and SP metrics for the majority of the cohort. Moreover, the results in Table 5 allowed us to determine the feasibility of using other criteria to

Table 4: Average sensitivity (SE) and specificity (SP) with the standard deviation in parentheses obtained by the hypoglycemia prediction models for 10 patients using Level 1 labeling (Test A).

		P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*
C1	*SE*	*60(0.1)*	*63(0.05)*	*82(0.04)*	*74(0.04)*	*84(0.05)*	*75(0.1)*	*76(0.06)*	*63(0.1)*	*71(0.1)*	*82(0.05)*
	SP	*89(0.02)*	*77(0.04)*	*79(0.04)*	*71(0.04)*	*43(0.08)*	*61(0.1)*	*78(0.06)*	*78(0.09)*	*74(0.1)*	*78(0.02)*
C2	*SE*	*63(0.04)*	*53(0.03)*	*58(0.08)*	*61(0.04)*	*70(0.05)*	*59(0.2)*	*34(0.05)*	*50(0.1)*	*45(0.09)*	*70(0.07)*
	SP	*85(0.07)*	*81(0.01)*	*83(0.04)*	*81(0.01)*	*73(0.05)*	*74(0.09)*	*91(0.02)*	*87(0.06)*	*92(0.06)*	*82(0.04)*
C3	*SE*	*73(0.01)*	*60(0.04)*	*64(0.1)*	*76(0.06)*	*74(0.06)*	*71(0.2)*	*66(0.08)*	*62(0.08)*	*51(0.1)*	*82(0.04)*
	SP	*83(0.06)*	*81(0.02)*	*82(0.03)*	*73(0.03)*	*62(0.08)*	*66(0.1)*	*83(0.05)*	*84(0.07)*	*86(0.05)*	*79(0.03)*
C4	*SE*	*72(0.1)*	*56(0.04)*	*61(0.09)*	*62(0.04)*	*69(0.1)*	*63(0.1)*	*47(0.1)*	*55(0.1)*	*42(0.2)*	*84(0.03)*
	SP	*85(0.07)*	*77(0.02)*	*83(0.03)*	*79(0.02)*	*71(0.06)*	*70(0.1)*	*82(0.04)*	*86(0.09)*	*97(0.03)*	*76(0.04)*
C5	*SE*	*44(0.1)*	*54(0.06)*	*56(0.08)*	*57(0.03)*	*69(0.06)*	*64(0.2)*	*44(0.1)*	*49(0.09)*	*38(0.03)*	*77(0.06)*
	SP	*94(0.02)*	*77(0.01)*	*87(0.01)*	*81(0.02)*	*71(0.06)*	*75(0.01)*	*83(0.03)*	*86(0.07)*	*96(0.03)*	*75(0.02)*
C6	*SE*	*43(0.1)*	*52(0.04)*	*50(0.04)*	*63(0.04)*	*66(0.06)*	*50(0.2)*	*23(0.05)*	*44(0.09)*	*33(0.1)*	*76(0.07)*
	SP	*94(0.02)*	*77(0.02)*	*87(0.02)*	*78(0.02)*	*74(0.04)*	*81(0.1)*	*92(0.2)*	*90(0.06)*	*94(0.07)*	*75(0.03)*
C7	*SE*	*43(0.1)*	*50(0.03)*	*50(0.04)*	*62(0.04)*	*62(0.05)*	*46(0.1)*	*22(0.05)*	*41(0.06)*	*33(0.2)*	*53(0.07)*
	SP	*94(0.02)*	*77(0.03)*	*87(0.02)*	*78(0.02)*	*76(0.03)*	*80(0.09)*	*93(0.02)*	*90(0.06)*	*96(0.05)*	*85(0.05)*
C8	*SE*	*43(0.1)*	*49(0.04)*	*49(0.05)*	*68(0.09)*	*64(0.04)*	*40(0.08)*	*20(0.06)*	*41(0.08)*	*34(0.1)*	*52(0.05)*
	SP	*94(0.02)*	*76(0.02)*	*87(0.02)*	*74(0.04)*	*75(0.04)*	*85(0.01)*	*93(0.03)*	*88(0.06)*	*98(0.03)*	*84(0.04)*

Table 5: Average Matthew's coefficient with the standard deviation in parentheses obtained by the hypoglycemia prediction models for 10 patients using Level 1 labeling.

	P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*
C1	*0.44(0.08)*	*0.40(0.02)*	*0.47(0.07)*	*0.36(0.06)*	*0.3(0.07)*	*0.34(0.1)*	*0.49(0.08)*	*0.4(0.07)*	*0.42(0.2)*	*0.58(0.7)*
C2	*0.41(0.1)*	*0.35(0.04)*	*0.35(0.07)*	*0.37(0.03)*	*0.42(0.05)*	*0.33(0.1)*	*0.3(0.05)*	*0.4(0.06)*	*0.44(0.1)*	*0.52(0.08)*
C3	*0.44(0.2)*	*0.42(0.03)*	*0.37(0.8)*	*0.39(0.06)*	*0.37(0.05)*	*0.36(0.1)*	*0.47(0.08)*	*0.46(0.07)*	*0.4(0.1)*	*0.59(0.06)*
C4	*0.47(0.2)*	*0.32(0.03)*	*0.37(0.08)*	*0.36(0.03)*	*0.4(0.05)*	*0.32(0.1)*	*0.28(0.1)*	*0.43(0.05)*	*0.49(0.2)*	*0.57(0.04)*
C5	*0.42(0.1)*	*0.32(0.04)*	*0.37(0.05)*	*0.33(0.02)*	*0.4(0.07)*	*0.38(0.1)*	*0.28(0.1)*	*0.38(0.07)*	*0.44(0.2)*	*0.5(0.06)*
C6	*0.4(0.11)*	*0.30(0.05)*	*0.33(0.01)*	*0.34(0.04)*	*0.4(0.04)*	*0.32(0.2)*	*0.21(0.04)*	*0.4(0.05)*	*0.37(0.2)*	*0.49(0.07)*
C7	*0.4(0.11)*	*0.28(0.04)*	*0.33(0.02)*	*0.34(0.05)*	*0.39(0.03)*	*0.28(0.01)*	*0.2(0.05)*	*0.36(0.05)*	*0.38(0.2)*	*0.41(0.05)*
C8	*0.4(0.11)*	*0.25(0.04)*	*0.33(0.01)*	*0.34(0.08)*	*0.39(0.02)*	*0.28(0.1)*	*0.18(0.08)*	*0.33(0.04)*	*0.44(0.2)*	*0.38(0.07)*

train the personalized models in terms of CMM, which was optimized for every case. For instance, C4 represented a better model according to MCC for P1. As listed in Table 4, C4 was slightly less specific than C1, but much more sensitive than C1. In addition, C5 was the best modeling option for P6. Similarly, the C3 models for P2, P4, P8, and P10 had less FPs with slightly lower true detection of hypoglycemic events compared with C1, and thus they are suitable models.

According to the results obtained from Test B (Table 6 and Table 7), as expected, most of the C2 models produced satisfactory results. In particular, the C2 models were the best options for P3, P4, P5, P7, and P10 in terms of CMM. However, the results obtained for P1 showed that C5–C8 had the best CMM. Models C3 and C5 were more suitable for P6 compared with C2 and the remaining cases. Based on the results for C5, the performance of this model was satisfactory for P3 and P6, with the same CMM values as C2. The SE results for C3 based on P6 showed that this model would outperform C2 in insulin bolus dosage applications. Therefore, it would be less likely to miss hypoglycemic events but at the cost of a higher FP rate. For P9, C8 performed better than C2 and the other models. According to the results, C5 performed better than C2 for P1 and P6, and their performance was comparable for P3. Model C6 was a better choice than C2 for P8. Finally, models C4 and C7 did not perform better than C1 or C2 for any patient.

Table 8 presents the median SP and SE obtained for the patient cohort according to both the Level 1 and Level 2 definitions using the most favorable results for each patient based on the CMM criterion. Table 8 presents the overall results but it is important to note that the selection of the best model for each patient should

Table 6: Average sensitivity (SE) and specificity (SP) with the standard deviation in parentheses obtained by the hypoglycemia prediction models for 10 patients using Level 2 labeling (Test A).

		P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*
C1	*SE*	*90(0.01)*	*75(0.03)*	*93(0.01)*	*82(0.06)*	*88(0.05)*	*81(0.13)*	*84(0.1)*	*77(0.1)*	*82(0.2)*	*90(0.04)*
	SP	*87(0.02)*	*68(0.03)*	*74(0.03)*	*68(0.04)*	*38(0.06)*	*55(0.06)*	*70(0.06)*	*73(0.08)*	*69(0.09)*	*72(0.03)*
C2	*SE*	*93(0.08)*	*69(0.05)*	*85(0.2)*	*77(0.06)*	*76(0.08)*	*66(0.1)*	*60(0.1)*	*72(0.1)*	*78(0.1)*	*85(0.08)*
	SP	*83(0.07)*	*74(0.01)*	*82(0.04)*	*79(0.02)*	*66(0.04)*	*68(0.1)*	*90(0.03)*	*85(0.06)*	*92(0.06)*	*79(0.04)*
C3	*SE*	*97(0.07)*	*74(0.03)*	*85(0.2)*	*81(0.05)*	*79(0.07)*	*80(0.1)*	*83(0.1)*	*80(0.06)*	*75(0.2)*	*90(0.03)*
	SP	*80(0.05)*	*72(0.02)*	*73(0.03)*	*69(0.03)*	*55(0.05)*	*60(0.1)*	*77(0.05)*	*79(0.06)*	*84(0.04)*	*73(0.04)*
C4	*SE*	*97(0.07)*	*73(0.05)*	*85(0.2)*	*78(0.05)*	*75(0.1)*	*65(0.1)*	*76(0.2)*	*76(0.1)*	*69(0.2)*	*92(0.01)*
	SP	*82(0.05)*	*70(0.02)*	*80(0.03)*	*78(0.03)*	*64(0.05)*	*63(0.1)*	*80(0.04)*	*82(0.08)*	*95(0.03)*	*69(0.05)*
C5	*SE*	*80(0.2)*	*72(0.06)*	*78(0.02)*	*72(0.04)*	*76(0.08)*	*72(0.09)*	*68(0.2)*	*72(0.1)*	*67(0.2)*	*87(0.04)*
	SP	*93(0.02)*	*71(0.01*	*84(0.02)*	*80(0.02)*	*64(0.04)*	*68(0.1)*	*81(0.04)*	*84(0.07)*	*95(0.02)*	*70(0.03)*
C6	*SE*	*80(0.2)*	*67(0.07)*	*68(0.2)*	*79(0.08)*	*72(0.08)*	*58(0.2)*	*52(0.2)*	*66(0.1)*	*61(0.2)*	*84(0.05)*
	SP	*93(0.01)*	*71(0.01)*	*85(0.02)*	*76(0.02)*	*67(0.03)*	*76(0.1)*	*93(0.02)*	*88(0.06)*	*95(0.05)*	*70(0.04)*
C7	*SE*	*80(0.02)*	*65(0.04*	*68(0.2)*	*77(0.08)*	*70(0.08)*	*49(0.1)*	*50(0.2)*	*61(0.1)*	*61(0.2)*	*63(0.1)*
	SP	*93(0.02)*	*72(0.02)*	*85(0.02)*	*76(0.02)*	*70(0.03)*	*76(0.1)*	*93(0.02)*	*88(0.06)*	*96(0.04)*	*83(0.05)*
C8	*SE*	*80(0.1)*	*61(0.04)*	*66(0.05)*	*80(0.09)*	*71(0.04)*	*48(0.08)*	*43(0.06)*	*62(0.08)*	*65(0.1)*	*63(0.05)*
	SP	*93(0.2)*	*71(0.01)*	*85(0.02)*	*72(0.04)*	*69(0.03)*	*81(0.08)*	*93(0.02)*	*86(0.06)*	*98(0.02)*	*82(0.04)*

Table 7: Average Matthew's coefficient with the standard deviation in parentheses obtained by the hypoglycemia prediction models for 10 patients using Level 2 labeling.

	P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*
C1	*0.45(0.1)*	*0.3(0.08)*	*0.31(0.08)*	*0.33(0.07)*	*0.26(0.03)*	*0.24(0.08)*	*0.35(0.08)*	*0.38(0.07)*	*0.39(0.2)*	*0.53(0.09)*
C2	*0.43(0.1)*	*0.32(0.05)*	*0.35(0.2)*	*0.4(0.05)*	*0.38(0.06)*	*0.24(0.03)*	*0.42(0.07)*	*0.47(0.08)*	*0.66(0.2)*	*0.57(0.1)*
C3	*0.4(0.09)*	*0.33(0.03)*	*0.32(0.1)*	*0.33(0.04)*	*0.32(0.05)*	*0.28(0.08)*	*0.4(0.08)*	*0.46(0.05)*	*0.5(0.1)*	*0.55(0.1)*
C4	*0.42(0.1)*	*0.3(0.05)*	*0.34(0.1)*	*0.39(0.04)*	*0.36(0.07)*	*0.2(0.1)*	*0.37(0.07)*	*0.47(0.07)*	*0.64(0.1)*	*0.53(0.08)*
C5	*0.51(0.08)*	*0.30(0.05)*	*0.35(0.1)*	*0.38(0.05)*	*0.37(0.06)*	*0.28(0.06)*	*0.34(0.08)*	*0.47(0.07)*	*0.63(0.2)*	*0.49(0.08)*
C6	*0.51(0.08)*	*0.28(0.05)*	*0.3(0.2)*	*0.38(0.07)*	*0.37(0.06)*	*0.26(0.1)*	*0.4(0.04)*	*0.49(0.07)*	*0.59(0.2)*	*0.47(0.07)*
C7	*0.51(0.08)*	*0.26(0.03)*	*0.3(0.2)*	*0.37(0.07)*	*0.37(0.06)*	*0.2(0.09)*	*0.39(0.07)*	*0.44(0.1)*	*0.6(0.3)*	*0.43(0.05)*
C8	*0.51(0.08)*	*0.23(0.03)*	*0.3(0.1)*	*0.35(0.07)*	*0.37(0.03)*	*0.23(0.06)*	*0.34(0.09)*	*0.42(0.09)*	*0.68(0.2)*	*0.43(0.05)*

be determined by the treating physician.

4. Discussion

In the present study, we determined the feasibility of developing a predictive model for postprandial hypoglycemia using a classification approach, given that most of the previous studies obtained predictive models based on regression. Moreover, we introduced a new feature to consider the risk of hypoglycemia defined as the area between a given hypoglycemia threshold and the CGM curve. Furthermore, we used the same risk concept to determine different degrees of hypoglycemia and applied it to assign the labels in several cases, which were evaluated under the consensus definition of hypoglycemia[2]. The generalization capability of the proposed method was demonstrated according to SE, SP, and MCC metrics.

The effect of using the area under the curve as a risk definition for training models C3–C8 was also tested under the objective definitions of Level 1 and Level 2 hypoglycemia according to the consensus standards[2], which might seem counterintuitive because models C3–C8 were trained using the targets generated by the risk criterion, i.e., positive class when the area under the threshold was above a certain limit but negative class otherwise. However, the practical capacity of this approach was demonstrated given that in some cases, the performance metrics obtained were similar to those obtained for C2 and C1, but suitable results were also produced for P1, P2, P4, P6, P8, P9, and P10 in Level 1, and for P1, P2, P6, P8, and P9 in Level 2. The main limitation of this risk approach is that there are several possible ways of defining a given risk threshold. For in-

Table 8: Median specificity (SP) and sensitivity (SE) results obtained for the most favorable model of each patient according to hypoglycemia Level 1 and Level 2.

		P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*	*Median*
Level 1	*SE*	72	60	82	76	70	64	76	62	42	82	71
	SP	85	81	79	73	73	75	78	84	97	79	79
Level 2	*SE*	80	74	85	77	76	80	60	66	65	85	77
	SP	93	72	82	79	66	60	90	88	98	79	81

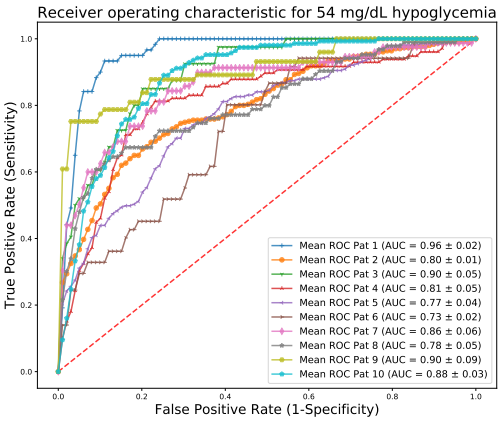

Figure 3: ROC curve for prediction results for Level 2 hypoglycemia

stance, in this particular study, we defined the thresholds based on the amount of hypoglycemic events that each represented and how these numbers compared with the targets in C2 and C1, as well as the overall distribution of risk for each patient. Clearly, this definition is quite flexible but it requires many experiments to achieve adequate results, although we demonstrated that it might be the fairest choice for some individuals.

Regarding the scope of the predictions, the models were designed to employ meals to activate the predictions. We used a window of 240 min after a meal/bolus to consider a general scenario assuming a fairly regular meal intake and a correct estimation of the carbohydrate content. An important disadvantage of this approach is that during the day, some patients have shorter times between meals and the nocturnal prediction is limited to only 4 h after dinner. Thus, a crucial improvement would involve personalizing the prediction window according to the behavioral patterns of each individual and demonstrating the feasibility of adapting a model for nocturnal hypoglycemia using a 6 h prediction window.

The results in Table 4 and Table 6 demonstrate the acceptable performance obtained for all patients in both scenarios, with alternative options in every case. The results obtained in this study are significant for at least two major reasons. First, we demonstrated the feasibility of predicting postprandial hypoglycemic events from a classification perspective using the data from SAP comprising the insulin dosage, CGM measurements, and carbohydrate intake information. This model has the potential to become a support tool for decision making in free-living conditions. Second, the proposed method uses the risk as both a feature for training predictive models and as a labeling scheme, where a given area threshold represents the severity of a hypoglycemic event. As listed in Table 1, the number of training examples varied significantly across patients. Despite the expected advantage of employing more training instances due to greater generalization capability, the performance metrics for the patients with more instances and those with less did not agree with the previous expectations. Moreover, we suggest that this is a reassuring indication of the need for personalized models and that fair comparisons cannot be made across individuals in this application.

The ROC curves for each patient displayed in Figure 2 and Figure 3 represent the mean test for the most favorable model according to MCC. These curves show an interesting contribution. Beyond the personalization of models by the data-driven techniques, these curves provide series of pairs [sensitivity, specificity] that can lead to a better customization of the models, depending on the physiological response of patients to the insulin therapy. Thus, for example, it would be possible to choose more sensitive models (at the cost of losing specificity), for those patients who experience non-significant post-pandrial glucose increase due to unnecesary actions from false positives.

In addition, patient-specific bolus insulin dosage optimization demands assessment of the pre-prandial calculation of the postprandial glucose levels based on an accurate prediction. An on-line dosage optimization tool is a target application that could be developed based on the results of this study. For instance, an on-line per-

sonalized hypoglycemia predictive model can compute the outcome from the insulin bolus at mealtime. If the prediction is positive, then this will be interpreted as an excessive amount of insulin and the bolus should be reduced by a fraction. This new bolus amount will be part of a new computation in the predictive model together with the other unaffected inputs. This process should be repeated until the prediction is a negative class, i.e., a glucose level not below 70 mg/dL (Level 1) or 54 mg/dL (Level 2). Other approaches could use the risk assessment to simultaneously adjust the bolus and basal insulin for the next few hours. Using this method, many hypoglycemic events could be avoided but without increasing the postprandial peak.

The effects of this approach are measurable. For instance, the average metrics for P2 using the C2 model evaluated for Level 2 hypoglycemia were TP = 46, TN = 289, FP = 102, and FN = 21. The number of hypoglycemic events (HE) was 67, and thus 69% (TP/HE) of these events could be avoided by using a bolus reduction/suspension strategy. Reducing the bolus in 148 ($TP + FP$) of the total meals would also lead to an increase in the area under the postprandial CGM curve in 22% of the cases ($FP/458$) on average due to unnecessary treatment.

5. Conclusions

In this study, we developed a method for predicting hypoglycemic events in the 4 h after a meal and a bolus. Our feasibility study using retrospective data from 10 patients determined a median SE higher than 70% for both hypoglycemic Level 1 and Level 2, with a low percentage of FPs.

Our model for predicting hypoglycemia obtained satisfactory performance (in terms of SE and SP) and it allows the identification or rejection of forthcoming hypoglycemia in the postprandial period. The accuracy of the results and the trade-off in terms of the performances of the metrics allow this method to be used in decision support systems for patients treated using an insulin pump.

6. Summary Points

What was already known on the topic?

- Hypoglycemia is a major burden within daily life of patients with Type 1 Diabetes. In addition, it is a barrier to achieving normoglycemia over a lifetime of using intensive insulin therapy, thereby precluding the long-term benefits of euglycemia. Repeated episodes of hypoglycemia reduce the ability of the patient to recognize its symptoms and signs.
- Previous studies employing diverse machine learning techniques showed that the field of blood glucose prediction is actively growing. Nevertheless, most of the works to date have been mainly restricted to nocturnal hypoglycemia and shorter prediction windows.
- Models that focus strictly on predicting particular outcomes such as hypoglycemic events are meaningful because they can be employed in decision-support applications.

What this study added to our knowledge?

- We introduce a novel method for model training using the hypoglycemia risk as a feature and as class-labeling factor.
- A method for predicting postprandial hypoglycemia using machine learning techniques has been developed.
- The proposed methodology provides a satisfactory performance, which would allow it to be used in decision support systems for patients treated using an insulin pump or MDI therapy.
- The prediction system allows predictions of the effect of the insulin dosage for a given meal, thereby facilitating on-line therapy decision making, such as bolus re-estimation for insulin-dependent patients.

7. Acknowledgments

This study was partly supported by the Spanish Ministry of Science and Innovation (grant DPI2016-78831-C2-2-R), the Catalan Government through grant 2017SGR1551, and by the Spanish Government through contract ES-2014-068289.

References

[1] G. B. Bolli, Hypoglycaemia unawareness., Diabetes & metabolism 23 Suppl 3 (1997) 29–35.

[2] G. Agiostratidou, H. Anhalt, D. Ball, L. Blonde, E. Gourgari, K. N. Harriman, A. J. Kowalski, P. Madden, A. H. McAuliffe-Fogarty, M. McElwee-Malloy, et al., Standardizing clinically meaningful outcome measures beyond hba1c for type 1 diabetes: A consensus report of the american association of clinical endocrinologists, the american association of diabetes educators, the american diabetes association, the endocrine society, jdrf international, the leona m. and harry b. helmsley charitable trust, the pediatric endocrine society, and the t1d exchange, Diabetes care 40 (12) (2017) 1622–1630.

[3] R. J. McCrimmon, R. S. Sherwin, Hypoglycemia in type 1 diabetes., Diabetes 59 (10) (2010) 2333–9. doi:10.2337/db10-0103.

[4] T. Biester, O. Kordonouri, M. Holder, K. Remus, D. Kieninger-Baum, T. Wadien, T. Danne, "Let the Algorithm Do the Work": Reduction of Hypoglycemia Using Sensor-Augmented Pump Therapy with Predictive Insulin Suspension (SmartGuard) in Pediatric Type 1 Diabetes Patients., Diabetes technology & therapeutics 19 (3) (2017) 173–182. doi:10.1089/dia.2016.0349.

[5] S. Oviedo, J. Vehí, R. Calm, J. Armengol, A review of personalized blood glucose prediction strategies for T1DM patients, International Journal for Numerical Methods in Biomedical Engineering 33 (6) (2017) e2833. doi:10.1002/cnm.2833.

[6] D. Elleri, J. M. Allen, M. Biagioni, K. Kumareswaran, L. Leelarathna, K. Caldwell, M. Nodale, M. E. Wilinska, C. L. Acerini, D. B. Dunger, R. Hovorka, Evaluation of a portable ambulatory prototype for automated overnight closed-loop insulin delivery in young people with type 1 diabetes, Pediatric Diabetes 13 (6) (2012) 449–453. doi:10.1111/j.1399-5448.2012.00903.x.

[7] S. Schmidt, D. Boiroux, A. K. Duun-Henriksen, L. Frøssing, O. Skyggebjerg, J. B. Jørgensen, N. K. Poulsen, H. Madsen, S. Madsbad, K. Nørgaard, Model-based closed-loop glucose control in type 1 diabetes: The DiaCon experience, Journal of Diabetes Science and Technology 7 (5) (2013) 1255–1264. doi:10.1177/193229681300700515.

[8] D. Boiroux, A. K. Duun-Henriksen, S. Schmidt, K. Nørgaard, S. Madsbad, O. Skyggebjerg, P. R. Jensen, N. K. Poulsen, H. Madsen, J. B. Jørgensen, Overnight Control of Blood Glucose in People with Type 1 Diabetes, Proceedings of the 8th IFAC Symposium on Biological and Medical Systems (2012) 73–78doi:10.3182/20120829-3-HU-2029.00106.

[9] B. Buckingham, H. P. Chase, E. Dassau, E. Cobry, P. Clinton, V. Gage, K. Caswell, J. Wilkinson, F. Cameron, H. Lee, B. W. Bequette, F. J. Doyle, Prevention of nocturnal hypoglycemia using predictive alarm algorithms and insulin pump suspension., Diabetes care 33 (5) (2010) 1013–7. doi:10.2337/dc09-2303.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[11] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing 14 (3) (2004) 199–222. arXiv:arXiv:1011.1669v3, doi:10.1023/B:STCO.0000035301.49549.88.

[12] W. McKinney, Data structures for statistical computing in python, in: S. van der Walt, J. Millman (Eds.), Proceedings of the 9th Python in Science Conference, 2010, pp. 51 – 56.

[13] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing In Science & Engineering 9 (3) (2007) 90–95. doi:10.1109/MCSE.2007.55.

[14] B. Scholkopf, A. J. Smola, Learning with kernels : support vector machines, regularization, optimization, and beyond, MIT Press, 2002.

[15] S. Theodoridis, K. Koutroumbas, Chapter 3 - Linear Classifiers, in: S. Theodoridis, K. Koutroumbas (Eds.), Pattern Recognition (Fourth Edition), fourth edi Edition, Academic Press, Boston, 2009, pp. 91–150. doi:https://doi.org/10.1016/B978-1-59749-272-0.50005-0.

[16] E. I. Georga, V. C. Protopappas, D. Ardigò, D. Polyzos, D. I. Fotiadis, A Glucose Model Based on Support Vector Regression for the Prediction of Hypoglycemic Events Under Free-Living Conditions, Diabetes Technology & Therapeutics 15 (8) (2013) 634–643. doi:10.1089/dia.2012.0285.

[17] R. H. Botwey, E. Daskalaki, P. Diem, S. G. Mougiakakou, Multi-model data fusion to improve an early warning system for hypo-/hyperglycemic events, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2014, IEEE, 2014, pp. 4843–4846. doi:10.1109/EMBC.2014.6944708.

[18] E. Dassau, F. Cameron, H. Lee, B. W. Bequette, H. Zisser, L. Jovanovic, H. P. Chase, D. M. Wilson, B. A. Buckingham, F. J. Doyle, Real-Time Hypoglycemia Prediction Suite Using Continuous Glucose Monitoring: A safety net for the artificial pancreas, Diabetes Care 33 (6) (2010) 1249–1254. doi:10.2337/dc09-1487.

Part IV

BOLUS ADVISOR APPLICATION FOR MACHINE LEARNING BASED POSTPRANDIAL HYPOGLYCEMIA FORECASTING

Submitted to IEEE Journal of Biomedical and Health Informatics. Sept, 2018 (JCR quartile: Q1; JIF: 3.850 in 2017; Ranked 4/25 in Medical Informatics).

Silvia Oviedo, Iván Contreras, Arthur Bertachi, Carmen Quirós, Marga Giménez, Ignacio Conget and Josep Vehi, "Bolus Advisor Application for Machine Learning Based Postprandial Hypoglycemia Forecasting". Manuscript submitted for publication.

Abstract

Diabetic patients treated with intensive insulin therapies require a tight glycemic control and may benefit from advanced tools to predict blood glucose (BG) concentration levels and ypo/hyperglycemia events. Prediction systems using Machine learning techniques have mainly focused on applications for sensor augmented pump (SAP) therapy. In contrast, insulin bolus calculators that rely on BG prediction for multiple daily insulin (MDI) injections for patients are scarce because of insuficient data sources and limited prediction capability of forecasting models. In this work, we trained individualized models that can predict postprandial hypoglycemia via machine learning algorithms using retrospective data from 10 real patients. The median [IQR] sensitivity and specificity for hypoglycemia cases where the BG level was below 70 mg/dL were 0.49 [0.2–0.5] and 0.74 [0.7–0.9], respectively. For hypoglycemia cases where the BG level was below 54 mg/dL, the median [IQR] sensitivity and specificity were 0.51 [0.4–0.6] and 0.74 [0.7–0.8], respectively. In addition, we designed and tested a bolus advising strategy for a similar in silico population. The advisor generates a bolus reduction suggestion as the scaled weighted sum of the predictions. We evaluated the general and postprandial glycemic outcomes of the in silico population to assess the systems capability of avoiding hypoglycemias. The results indicated a decrease of 37% in the median number of postprandial hypoglycemias median decrease of 44% for hypoglycemias of 70 mg/dL and 54 mg/dL, respectively. This dramatic reduction makes this method a good candidate to be integrated into any Decision Support System for diabetes management.

Keywords

bolus calculation, hypoglycemia prediction, machine learning, postprandial hypoglycemia, SAP, type-1 diabetes

3

DISCUSSION

The results obtained in this dissertation showed that training personalized models is an efficient approach to cope with intra-patient variability. In addition, Machine Learning (ML) models alone or in combination with minimal physiological models make the models more flexible and simpler to adjust. The works presented in the previous chapter were developed aiming to extend the prediction horizon from the typical 30 min to 1 h [46], in order to determine the technical viability of forecasting glucose values/events beyond these prediction windows. Furthermore, the predictions were demonstrated to be valuable both, as regression and classification outcomes for different therapeutic options. Below, a brief discussion for each forecasting application is presented:

3.1 PERSONALIZED BLOOD GLUCOSE PREDICTION: A HYBRID APPROACH USING GRAMMATICAL EVOLUTION AND PHYSIOLOGICAL MODELS

Personalized models were evolved using GE for a virtual cohort of 100 patients using the glucose-specific loss function Glucose-specific Root Mean Square Error (gRMSE) and tested using an independent set of data for a 120 min prediction window. The approach of this work was hybrid, using an IOB model and a glucose rate of appearance model as a pre-processing stage before the GE model. Additionally, the training and test data were split into 6h segments every day. This allowed building specific models for every meal and the overnight period. Another approach used in this work proposed the training and testing of personalized models for the time period from 2h to 6h after a meal, targeting specifically this period of time where the chances of hypoglycemia are higher. The overall averaged gRMSE for 24 h was 23.25 (mg/dL). In general, models for the dinner segment showed the highest deviation from the true values, with an averaged gRMSE of 33 (mg/dL) for the first approach and 29.8 (mg/dL) for the second, making this period the most challenging under both approaches. However, the evaluation of the deviations under Clarke Error Grid (CEG) allowed quantifying the clinical harmfulness of the deviations. Despite having the largest deviation of the segments, CEG for $A + B$ zones retained more than 98% of the data pairs, thus, making the predictions under both approaches clinically reliable.

Removing the first 2h after the meals from the models allowed improving the performance metrics after meals as well as increasing the number of pairs in zone A of CEG. Applications for both approaches include the optimization of the insulin Bolus and the prevention of overnight hypoglycemia.

3.2 RISK-BASED POSTPRANDIAL HYPOGLYCEMIA FORECASTING USING SUPERVISED LEARNING

This work demonstrated the feasibility of predicting postprandial hypoglycemia for a prediction window of 240 min. Most of the previous works in prediction of hypoglycemic events have been designed as short-term regression problems, where a ML algorithm is trained to predict CGM values and then, those values are evaluated according to defined thresholds for hypoglycemia. This approach proposed the forecasting of two severity levels of postprandial hypoglycemia, 70mg/dL and 54mg/dL, directly as a classification problem. The use of Matthews Correlation Coefficient (MCC) as a target metric allowed merging the Sensitivity (SE) and Specificity (SP) into a single metric that is sensible to both performance metrics. This means that, in order to have a high MCC, both SE and SP must be high. The target applications for this prediction scheme are SAP and MDI decision support systems. Patients estimate the insulin bolus dose based on the estimated carbohydrate content of the food at mealtime. Carbohydrates are often overestimated and thus the consequent overestimation of the bolus dose that can lead to postprandial hypoglycemia. A classifier that is able to reliably predict hypoglycemia after a meal is a useful feature for therapy purposes.

This work proposes a classifier training scheme based on a set of 8 labeling criteria. The labeling criteria included the formal definitions of hypoglycemic events using the 70mg/dL and 54mg/dL thresholds and additionally, included a risk-based definition of hypoglycemia. The median specicity and sensitivity were 79% and 71% for Level 1 hypoglycemia, respectively, and 81% and 77% for Level 2 hypoglycemia.

The risk concept was proved to be useful as some of the models demonstrated better test MCC under the risk-based training approach. Nevertheless, this approach used a specific threshold for each case that was defined empirically after many experiments and it adds another parameter that must be optimized along with the hyper-parameters of the ML forecasting algorithms.

This work validated a method for training personalized classifiers that forecast hypoglycemic episodes after meals, using the information from the insulin pump and CGM of real patients in free-living conditions. The results motivated the adaptation of the method for a application in the context of MDI therapy.

3.3 BOLUS ADVISOR APPLICATION FOR MACHINE LEARNING BASED POSTPRANDIAL HYPOGLYCEMIA FORECASTING

In this work, individualized models were produced to predict postprandial hypoglycemia using retrospective data from 10 real patients in free-living conditions. The median[IQR] SE and SP for $70mg/dL$ hypoglycemia were $0.49[0.2-0.5]$ and $0.74[0.7-0.9]$ respectively. For $54mg/dL$ hypoglycemia, the median [IQR] SE and SP were $0.51[0.4-0.6]$ and $0.74[0.7-0.8]$, respectively. Additionally, a bolus advising strategy was designed. It generates a bolus reduction as the scaled weighted sum of the predictions for an *in-silico* population that mimics the real population. The general and postprandial glycemic outcomes were assessed to determine the system capability to avoid hypoglycemia. The median value of postprandial episodes of hypoglycemia decreased 37% and 44% for $70mg/dL$ and $54mg/dL$ hypoglycemia, respectively. From a glycemic control perspective, the bolus intervention lead to a 9% increase in the postprandial peak, a 10% increase in the mean CGM and a 35% increase in the percentage of time spent in hyperglycemia. Although these results are far from ideal, the bolus reduction strategy was specifically designed to reduce the number of postprandial hypoglycemia events, which is a major concern for the patient cohort of the study and therefore justifies its use. A much more sophisticated strategy would be required to reduce the number of hypoglycemic events without worsening the time above $180mg/dL$ in MDI applications. This study provided a methodology for bolus correction based on the prediction of postprandial hypoglycemia, according to the risk-based method described in the previous section. Nevertheless, this approach targets patients under MDI therapy and therefore, the inputs for this method are scarce compared to those for SAP. This is because MDI therapy is the simplest form of control in T1D and having additional inputs in the model would make it less practical.
This work consisted in the design and test of a set of forecasting models based on ML for a real patient cohort of 10 patients under SAP therapy. The data from the CGM were used for the labeling scheme. As expected, the test with a hold-out set gives weak results in terms of the MCC metric. This could be attributed to the limited information contained in the inputs of the models. Nevertheless, the target application for this methodology is the bolus dose adjustment for MDI patients. In this scenario, the SE and SP values achieved a low false positive rate and a moderate to low true positive rate. The median SE values, although are far from ideal, do not represent a naive classifier, given the SP median values that reflect a satisfactory false positive rate. This allows using a forecast from this method as a bolus advising strategy, targeting the true positive cases. The test of the robustness of this strategy was performed in a virtual co-

hort that reproduced the real cohort and showed satisfactory results in terms of the glycemic outcomes and specially, the prevention of postprandial hypoglycemia.

4

CONCLUSIONS

4.1 SUMMARY OF COMPLETED WORK

This thesis addressed the problem of developing personalized models for mid-term BG and postprandial hypoglycemia predictions using machine learning algorithms.

The use of a glucose-specific cost function that takes into account the clinical harmfulness of deviations makes the predictions more reliable in terms of clinical usefulness for the GE model. In addition, results suggest that generating piece-wise models improve the accuracy and clinical reliability of the overall model.

A method for predicting hypoglycemic events in the time window from 2h to 6h after a meal was proposed and validated using retrospective data from 10 patients. The classification model for predicting hypoglycemia showed satisfactory performance as it allows identifying forthcoming hypoglycemia in the postprandial period. The accuracy of the results and the trade-off in terms of the performances of the metrics allow this method to be used in decision support systems for patients treated using SAP or MDI therapy.

Finally, a prediction-based bolus advisor was developed using a minimal input set. Initial *in silico* results for a cohort of 10 patients mimicking the retrospective data from 10 real patients demonstrate the feasibility of a potential therapeutic use in patients under MDI therapy. The performance of the prediction is a strong limitation for a forecast system relying on MDI data. Nevertheless, the results obtained in the proof of concept using *in silico* data allow this method to be considered for decision support systems for real patients.

4.2 CONTRIBUTIONS

As a result of this thesis the following contributions have been made:

- Literature review and analysis of the current trends in physiological, data-driven and hybrid models for blood glucose forecasting and control.
- Adaptation of a hybrid model using grammatical evolution and physiological models for mid-term prediction of BG. The use of a glucose-specific loss function in the evolution of the models is the first reported in the literature and allowed clinically safe predictions.

- Development and validation of a personalized model training scheme for postprandial hypoglycemia prediction for SAP and MDI users. The scheme relies on the estimation of a defined risk parameter, along with the predefined 70mg/dL and 54mg/dL hypoglycemia thresholds.
- Proposal of a bolus reduction strategy based on postprandial hypoglycemia prediction models for MDI users. The benefits of the proposed method were assessed using a virtual cohort mimicking a real patient cohort.

4.3 FUTURE WORK

The results of the research conducted in this thesis are promising for a wide range of applications in T1D therapy. Nevertheless, the performance and safety of the predictions can be improved further by generating a set of interchangeable models that predict useful BG values for control and therapy purposes based on the determination of individual specific dynamics, lifestyle, and other factors. An extension of this work will include testing personalized BG prediction models in a more challenging situation involving real subjects.
In the case of GE models, other grammar architectures can be explored to improve the accuracy and flexibility of the patient model. Future studies shall include a comparative analysis that explores the effect of grammar complexity in practical applications. In addition, there is a lack of conclusive reports on exercise models. Further research might explore how to incorporate input signals related to physical exercise can improve or deteriorate the accuracy of forecasting models.
A natural progression of the work in prediction using classifiers is to adapt the risk-based training method to data from real patients under MDI therapy that also wear CGM sensors. CGM data allows the models to have much more predictive capability and therefore, a more reliable output.

BIBLIOGRAPHY

[1] N H Cho, J E Shaw, S Karuranga, Y Huang, J D da Rocha Fernandes, A W Ohlrogge, and B Malanda. "IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045." In: *Diabetes Research and Clinical Practice* 138 (2018), pp. 271–281. ISSN: 0168-8227. DOI: https://doi.org/10.1016/j.diabres.2018.02.023. URL: http://www.sciencedirect.com/science/article/pii/S0168822718302031.

[2] International Diabetes Federation. *IDF DIABETES ATLAS*. Seventh Ed. 2015. ISBN: 9782930229812. URL: www.diabetesatlas.org.

[3] Vijayaganapathy Vaithilingam, Sumeet Bal, and Bernard E. Tuch. "Encapsulated islet transplantation: Where do we stand?" In: *Review of Diabetic Studies* 14.1 (2017), pp. 51–78. ISSN: 16130575. DOI: 10.1900/RDS.2017.14.51.

[4] Xiangwei Xiao et al. "Endogenous Reprogramming of Alpha Cells into Beta Cells, Induced by Viral Gene Therapy, Reverses Autoimmune Diabetes." In: *Cell Stem Cell* 22.1 (2018), 78–90.e4. ISSN: 18759777. DOI: 10.1016/j.stem.2017.11.020. URL: https://doi.org/10.1016/j.stem.2017.11.020.

[5] Satish K. Garg et al. "Glucose Outcomes with the In-Home Use of a Hybrid Closed-Loop Insulin Delivery System in Adolescents and Adults with Type 1 Diabetes." In: *Diabetes Technol. Ther.* 19.3 (2017), pp. 155–163. ISSN: 1520-9156. DOI: 10.1089/dia.2016.0421. URL: http://online.liebertpub.com/doi/10.1089/dia.2016.0421.

[6] Yuxiang Zhong, Siddharth Arunachalam, Pratik Agrawal, Huzefa Neemuchwala, Toni L Cordero, and Francine R Kaufman. *Real-World Assessment of Sugar. IQ with Watson—A Cognitive Computing-Based Diabetes Management Solution*. 2018.

[7] Lia Bally et al. "Day-and-night glycaemic control with closed-loop insulin delivery versus conventional insulin pump therapy in free-living adults with well controlled type 1 diabetes: an open-label, randomised, crossover study." In: *The Lancet Diabetes & Endocrinology* 5.4 (2017), pp. 261–270. DOI: 10.1016/s2213-8587(17)30001-3. URL: https://doi.org/10.1016/s2213-8587(17)30001-3.

[8] Giovanni Sparacino, Andrea Facchinetti, and Claudio Cobelli. ""Smart" continuous glucose monitoring sensors: on-line signal processing issues." In: *Sensors (Basel).* 10.7 (2010), pp. 6751–72. ISSN: 1424-8220. DOI: 10.3390/s100706751. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3231130{\&}tool=pmcentrez{\&}rendertype=abstract.

[9] J. Bondia, J. Vehí, C.C. Palerm, and P. Herrero. "El Páncreas Artificial: Control Automático de Infusión de Insulina en Diabetes Mellitus Tipo 1." In: *Rev. Iberoam. Automática e Informática Ind. RIAI* 7.2 (2010), pp. 5–20. ISSN: 16977912. DOI: 10.1016/S1697-7912(10)70021-2. URL: http://www.sciencedirect.com/science/article/pii/S1697791210700212.

[10] Rebecca A Harvey et al. "Quest for the artificial pancreas: combining technology with treatment." In: *IEEE Eng. Med. Biol. Mag.* 29.2 (2010), pp. 53–62. ISSN: 1937-4186. DOI: 10.1109/MEMB.2009.935711. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3823381{\&}tool=pmcentrez{\&}rendertype=abstract.

[11] J. B. McGill and A. Ahmann. "Continuous Glucose Monitoring with Multiple Daily Insulin Treatment: Outcome Studies." In: *Diabetes Technol. Ther.* 19.S3 (June 2017), S3–S12.

[12] Pau Herrero, Peter Pesl, Monika Reddy, Nick Oliver, Pantelis Georgiou, and Christofer Toumazou. "Advanced insulin bolus advisor based on run-to-run control and case-based reasoning." In: *IEEE J. Biomed. Heal. Informatics* 19.3 (2015), pp. 1087–1096. ISSN: 21682194. DOI: 10.1109/JBHI.2014.2331896. arXiv: arXiv:1208.5721.

[13] Roberto Visentin, Chiara Dalla Man, Boris Kovatchev, and Claudio Cobelli. "The university of Virginia/Padova type 1 diabetes simulator matches the glucose traces of a clinical trial." In: *Diabetes technology & therapeutics* 16.7 (2014), pp. 428–34. ISSN: 1557-8593. DOI: 10.1089/dia.2013.0377. URL: http://www.ncbi.nlm.nih.gov/pubmed/24571584http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4074748.

[14] Erin J. Mansell, Paul D. Docherty, Liam M. Fisk, and J. Geoffrey Chase. "Estimation of secondary effect parameters in glycaemic dynamics using accumulating data from a virtual type 1 diabetic patient." In: *Math. Biosci.* 266 (2015), pp. 108–117. ISSN: 00255564. DOI: 10.1016/j.mbs.2015.06.002. URL: http://www.sciencedirect.com/science/article/pii/S0025556415001200.

[15] Gary Scheiner. *Think Like a Pancreas: A Practical Guide to Managing Diabetes with Insulin–Completely Revised and Updated*. Da Capo Lifelong Books, 2012.

[16] Moshe Phillip, Tadej Battelino, Eran Atlas, Olga Kordonouri, Natasa Bratina, Shahar Miller, Torben Biester, Magdalena Avbelj Stefanija, Ido Muller, Revital Nimri, et al. "Nocturnal glucose control with an artificial pancreas at a diabetes camp." In: *New England Journal of Medicine* 368.9 (2013), pp. 824–833.

[17] Margarita Fernandez, Minaya Villasana, and Dan Streja. "Glucose dynamics in Type I diabetes: Insights from the classic and linear minimal models." In: *Computers in Biology and Medicine* 37 (2007), pp. 611–627. DOI: 10.1016/j.compbiomed.2006.05.008. URL: www.intl.elsevierhealth.com/journals/cobm.

[18] Claudio Cobelli, Chiara Dalla Man, Giovanni Sparacino, Lalo Magni, Giuseppe De Nicolao, and Boris P Kovatchev. "Diabetes : Models , Signals , and Control." In: *IEEE Trans Biomed Eng* 2 (2009), pp. 54–96.

[19] E.D Lehmann and T Deutsch. "Compartmental models for glycaemic prediction and decision-support in clinical diabetes care: promise and reality." In: *Computer Methods and Programs in Biomedicine* 56.2 (1998), pp. 193–204. ISSN: 01692607. DOI: 10.1016/S0169-2607(98)00025-X. URL: http://www.sciencedirect.com/science/article/pii/S016926079800025X.

[20] Chiara Dalla Man, Robert a. Rizza, and Claudio Cobelli. "Meal simulation model of the glucose-insulin system." In: *IEEE Transactions on Biomedical Engineering* 54.10 (2007), pp. 1740–1749. ISSN: 00189294. DOI: 10.1109/TBME.2007.893506.

[21] Roman Hovorka et al. "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes." In: *Physiol. Meas* 25 (2004), pp. 905–920. DOI: 10.1088/0967-3334/25/4/010.

[22] R N Bergman, Y Z Ider, C R Bowden, and C Cobelli. "Quantitative estimation of insulin sensitivity." In: *The American journal of physiology* 236.6 (1979), E667–77. ISSN: 0002-9513. URL: http://www.ncbi.nlm.nih.gov/pubmed/443421.

[23] Naviyn Prabhu Balakrishnan, Lakshminarayanan Samavedham, and Gade Pandu Rangaiah. "Personalized mechanistic models for exercise, meal and insulin interventions in children and adolescents with type 1 diabetes." In: *Journal of theoretical biology* 357 (2014), pp. 62–73. ISSN: 1095-8541. DOI: 10.1016/j.jtbi.2014.04.038. URL: http://www.sciencedirect.com/science/article/pii/S0022519314002720.

[24] M García-Jaramillo, R Calm, J Bondia, and J Vehí. "Prediction of postprandial blood glucose under uncertainty and intrapatient variability in type 1 diabetes: a comparative study of three interval models." In: *Computer methods and programs in biomedicine* 108.1 (2012), pp. 224–33. ISSN: 1872-7565. DOI: 10.

1016/j.cmpb.2012.04.003. URL: http://www.sciencedirect.com/science/article/pii/S0169260712001058.

[25] Qiang Fang, Lei Yu, and Peng Li. "A new insulin-glucose metabolic model of type 1 diabetes mellitus: An in silico study." In: *Comput. Methods Programs Biomed.* 120.1 (2015), pp. 16–26. ISSN: 01692607. DOI: 10.1016/j.cmpb.2015.03.009. URL: http://www.sciencedirect.com/science/article/pii/S0169260715000759.

[26] C Novara, N Mohammad Pour, T Vincent, and G Grassi. "A Nonlinear Blind Identification Approach to Modeling of Diabetic Patients." In: *Proc. 19th World Congr. Int. Fed. Autom. Control* (2015), pp. 1–9.

[27] Kamuran Turksoy, Elif S Bayrak, Laurie Quinn, Elizabeth Littlejohn, and Ali Cinar. "Adaptive Multivariable Closed-Loop Control of Blood Glucose Concentration in Patients with Type 1 Diabetes." In: *Am. Control Conf.* (2013), pp. 2905–2910. ISSN: 07431619.

[28] Peng Li, Lei Yu, Jiping Wang, Liquan Guo, and Qiang Fang. "Effect of meal intake on the quality of empirical dynamic models for Type 1 Diabetes." In: *2014 IEEE Int. Symp. Bioelectron. Bioinforma. (IEEE ISBB 2014)* 1 (2014), pp. 1–4. DOI: 10.1109/ISBB.2014.6820942. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6820942.

[29] Chunhui Zhao and Yongji Fu. "Statistical analysis based online sensor failure detection for continuous glucose monitoring in type I diabetes." In: *Chemom. Intell. Lab. Syst.* 144 (2015), pp. 128–137. ISSN: 01697439. DOI: 10.1016/j.chemolab.2015.04.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0169743915000830.

[30] J. Ignacio Hidalgo, J. Manuel Colmenar, José L. Risco-Martin, Alfredo Cuesta-Infante, Esther Maqueda, Marta Botella, and José Antonio Rubio. "Modeling glycemia in humans by means of Grammatical Evolution." In: *Appl. Soft Comput.* 20 (2014), pp. 40–53. ISSN: 15684946. DOI: 10.1016/j.asoc.2013.11.006. URL: http://www.sciencedirect.com/science/article/pii/S156849461300402X.

[31] Simon Fong, Sabah Mohammed, Jinan Fiaidhi, and Chee Keong Kwoh. "Using causality modeling and Fuzzy Lattice Reasoning algorithm for predicting blood glucose." In: *Expert Syst. Appl.* 40.18 (2013), pp. 7354–7366. ISSN: 09574174. DOI: 10.1016/j.eswa.2013.07.035. URL: http://www.sciencedirect.com/science/article/pii/S0957417413005149.

[32] F. Stahl, R. Johansson, and Eric Renard. "Bayesian combination of multiple plasma glucose predictors." In: *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* (2012), pp. 2839–2844. ISSN: 1557170X. DOI: 10.1109/EMBC.2012.6346555.

[33] Hajrudin Efendic, Harald Kirchsteiger, Guido Freckmann, and Luigi del Re. "Short-term prediction of blood glucose concentration using interval probabilistic models." In: *22nd Mediterr. Conf. Control Autom.* IEEE, 2014, pp. 1494–1499. ISBN: 978-1-4799-5901-3. DOI: 10.1109/MED.2014.6961587. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6961587.

[34] Chiara Zecchin, Andrea Facchinetti, Giovanni Sparacino, and Claudio Cobelli. "How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to CGM data? A proof-of-concept study." In: *Journal of Diabetes Science and Technology* 10.5 (2016), pp. 1149–1160. ISSN: 19322968. DOI: 10.1177/1932296816654161.

[35] J Fernandez de Canete, S Gonzalez-Perez, and J C Ramos-Diaz. "Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes." In: *Comput. Methods Programs Biomed.* 106.1 (2012), pp. 55–66. ISSN: 1872-7565. DOI: 10.1016/j.cmpb.2011.11.006. URL: http://www.sciencedirect.com/science/article/pii/S0169260711003117.

[36] S. Shanthi, P. Balamurugan, and D. Kumar. "Performance comparison of featured neural network with gradient descent and levenberg-marquart algorithm trained neural networks for prediction of blood glucose values with continuous glucose monitoring sensor data." In: *Emerg. Trends Sci. Eng. Technol. (INCOSET), 2012 Int. Conf.* (2012), pp. 385–391. DOI: 10.1109/incoset.2012.6513938. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6513938.

[37] K. Zarkogianni, K. Mitsis, E. Litsa, M.-T. Arredondo, G. Fic, A. Fioravanti, and K. S. Nikita. "Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring." In: *Med. Biol. Eng. Comput.* 53.12 (2015), pp. 1333–43. ISSN: 0140-0118. DOI: 10.1007/s11517-015-1320-9. URL: http://link.springer.com/10.1007/s11517-015-1320-9.

[38] Marzia Cescon, Rolf Johansson, and Eric Renard. "Subspace-based linear multi-step predictors in type 1 diabetes mellitus." In: *Biomed. Signal Process. Control* 22 (2015), pp. 99–110. ISSN: 17468094. DOI: 10.1016/j.bspc.2014.09.012. URL: http://www.sciencedirect.com/science/article/pii/S1746809414001475.

[39] Chiara Zecchin. "Online Glucose Prediction in Type 1 Diabetes by Neural Network Models." In: *Univ. Degli Stud. Di Padova. Sch. Inf. Eng. Sect. Bioeng. XXVI Ser.* January (2014).

[40] K. Zarkogianni, E. Litsa, A. Vazeou, and K. S. Nikita. "Personalized glucose-insulin metabolism model based on self-organizing maps for patients with Type 1 Diabetes Mellitus." English. In: *13th IEEE Int. Conf. Bioinforma. Bioeng.* IEEE, 2013, pp. 1–4. ISBN: 978-1-4799-3163-7. DOI: 10.1109/BIBE.2013.6701604. URL: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6701604.

[41] Yenny Leal, Winston Garcia-Gabin, Jorge Bondia, Eduardo Esteve, Wifredo Ricart, Jose-Manuel Fernández-Real, and Josep Vehí. "Real-time glucose estimation algorithm for continuous glucose monitoring using autoregressive models." In: *J. Diabetes Sci. Technol.* 4.2 (2010), pp. 391–403. ISSN: 1932-2968. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2864176{\&}tool=pmcentrez{\&}rendertype=abstract.

[42] E. I. Georga, V. C. Protopappas, Diego Ardigo, M. Marina, I. Zavaroni, D. Polyzos, and D. I. Fotiadis. "Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression." In: *Biomed. Heal. Informatics, IEEE J.* 17.1 (2013), pp. 71–81. ISSN: 2168-2208. DOI: 10.1109/titb.2012.2219876. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6307869.

[43] Eleni I Georga, Vasilios C Protopappas, Demosthenes Polyzos, and Dimitrios I Fotiadis. "A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests." In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2012 (2012), pp. 2889–92. ISSN: 1557-170X. DOI: 10.1109/EMBC.2012.6346567. URL: http://www.ncbi.nlm.nih.gov/pubmed/23366528.

[44] Biagi L Oviedo S Vehi J Contreras I Bertachi A. In: *Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data*. 2018.

[45] Cynthia R. Marling and Razvan C. Bunescu. "The OhioT1DM Dataset For Blood Glucose Level Prediction." In: *KHD@IJCAI*. 2018.

[46] S. Oviedo, J. Vehi, R. Calm, and J. Armengol. "A review of personalized blood glucose prediction strategies for T1DM patients." In: *Int J Numer Method Biomed Eng* 33.6 (June 2017). DOI: {10.1002/cnm.2833}. URL: {https://www.ncbi.nlm.nih.gov/pubmed/27644067}.

Universitat de Girona