Universitat de Girona

REINFORCEMENT LEARNING FOR BOLUS INSULIN DOSING FOR PEOPLE WITH TYPE 1 DIABETES

Sayyar Ahmad

Universitat de Girona

Doctoral Thesis

Reinforcement Learning for Bolus Insulin Dosing for People with Type 1 Diabetes

Sayyar Ahmad, 2023

DOCTORAL THESIS

REINFORCEMENT LEARNING FOR BOLUS INSULIN DOSING FOR PEOPLE WITH TYPE 1 DIABETES

SAYYAR AHMAD

2023

DOCTORAL PROGRAMME IN TECHNOLOGY

Supervised by: Josep Vehí Casellas

Co-supervised by: Aleix Beneyto Tantiña

Presented in partial fulfillment of the requirements for a doctoral degree from the University of Girona

DOCTORAL THESIS

REINFORCEMENT LEARNING FOR BOLUS INSULIN DOSING FOR PEOPLE WITH TYPE 1 DIABETES

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

By:

Sayyar Ahmad

Supervisor:

Josep Vehí Casellas

Co-supervisor:

Aleix Beneyto Tantiña

To my parents, brothers, sisters and wife, for their endless support and encouragement over the years.

ACKNOWLEDGEMENTS

Although completing this doctoral thesis wasn't without its challenges and hardships, it did give me the most gratifying sense of success and pleasure. Without the generous support I received from so many individuals, none of my effort would have been possible. I would want to use this opportunity to express my sincere gratitude to everyone of you.

I am highly indebted to my supervisor Prof. Josep Vehí who gave me the opportunity of pursuing the doctoral degree. Working in his supervision has been a wonderful experience from many aspects. First, his belief in my abilities has kindled my interest in research. Second, his encouragement to come up with my own original ideas has helped me appreciate the value of formulating meaningful research problems and had a profound effect in shaping my vision for future research.

I would like to acknowledge Agency for Management of University and Research Grants (AGAUR), Government of Catalonia for sponsoring my doctoral studies and University of Girona for providing me such a memorable environment.

I am also grateful to my co-supervisor Dr. Aleix Beneyto for all his advice, time and help provided during the development of this research study.

I would like to extend my deepest gratitude to all the members of the MICELAB research group. To Anna Comas for all the technical support and help provided. Special thanks to Dr. Charrise Ramkissoon and Dr. Iván Contreras, with whom I worked hand in hand and published several of the articles presented in this dissertation.

Most importantly, none of this would have been possible without the love and patience of my family. My immediate family to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my family. My extended family has aided and encouraged me throughout this endeavor. I warmly appreciate the generosity and understanding of all of them.

Sayyar Ahmad
Girona, Spain
October, 2023

LIST OF PUBLICATIONS

This thesis is based on a compendium of the following publications:

1. **Sayyar Ahmad**, Charrise M Ramkissoon, Aleix Beneyto, Ignacio Conget, Marga Giménez and Josep Vehí. Generation of virtual patient populations that represent real type 1 diabetes cohorts, *Mathematics*, 9(11):1200, **2021** (Q1).

2. **Sayyar Ahmad**, Aleix Beneyto, Ivan Contreras and Josep Vehí. Bolus insulin calculation without meal information. A reinforcement learning approach, *Artificial Intelligence in Medicine*, 134:102436, **2022** (Q1).

3. **Sayyar Ahmad**, Aleix Beneyto, Ivan Contreras and Josep Vehí. An automatic deep reinforcement learning bolus calculator for automated insulin delivery systems, *Scientific Reports*, Submitted (Q2).

The research work leading to this thesis resulted in additional conference publications, which are listed below and sorted by publication date.

Conferences

1. **Sayyar Ahmad**, Aleix Beneyto and Josep Vehí, A reinforcement learning bolus calculator with no meal information for patients with type 1 diabetes. In 15th Advanced Technologies & Treatments for Diabetes, Barcelona - Spain, page A23, **2022**. (abstract).

2. **Sayyar Ahmad**, Charrise M Ramkissoon, Aleix Beneyto, Josep Vehí, Marga Giménez, Ignacio Conget and Clara Viñals, Creating virtual simulation scenarios that mimic

clinical data acquired from patients with type 1 diabetes, In 20th Annual Virtual Diabetes Technology Meeting, page A2, **2020**. (abstract).

Acronyms and Abbreviations

The following acronyms and abbreviations can be found in this thesis.

Acronyms and abbreviations

ADA	American Diabetes Association
AHCL	Advanced Hybrid Closed-Loop System
AI	Artificial Intelligence
AID	Automatic Insulin Delivery System
AIDs	Automatic Insulin Delivery Systems
AP	Artificial Pancreas
BG	Blood Glucose
CF	Correction Factor
CGM	Continuous Glucose Monitor
CHO	Carbohydrate
CL	Closed-Loop
CLC	Closed-Loop Control
CSII	Continuous Subcutaneous Insulin Infusion
CR	Carbohydrates to Insulin Ratio
DCCT	Diabetes Control and Complication Trail
DL	Deep Learning
DQN	Deep Q Network
DRL	Deep Reinforcement Learning
DSSD	Decision Support Systems for Diabetes
DM	Diabetes Mellitus
EASD	European Association for the Study of Diabetes
FAID	Fully Automatic Insulin Delivery System
FDA	US Food and Drug Administration
HAID	Hybrid Automatic Insulin Delivery System
HCL	Hybrid Closed-Loop System
MDI	Multiple Daily Injections
MDP	Markov Decision Process
ML	Machine Learning
MPC	Model Predictive Control

PD	Proportional-Derivative
PID	Proportional-Integral-Derivative
RL	Reinforcement Learning
SAFE	Safety Auxiliary Feedback Element
SAP	Sensor-Augmented Pump
SBC	Standard Bolus Calculator
SC	Subcutaneous
SL	Supervised Learning
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
UKF	Unscented Kalman Filter
USL	Unsupervised Learning
VPs	Virtual Patients

TABLE OF CONTENTS

List of Publications **v**

Acronyms and Abbreviations **vii**

List of Tables **xi**

List of Figures **xi**

Abstract **xiii**

Resumen **xv**

Resum **xvii**

1 Introduction **1**

1.1 Diabetes Mellitus 1

1.2 Type 1 Diabetes Mellitus 3

1.3 Automatic Insulin Delivery Systems 5

1.4 Artificial Intelligence for Diabetes 9

1.4.1 Reinforcement Learning in Diabetes Technology 10

1.5 Thesis Objectives 14

1.6 Thesis Structure 15

2 Reinforcement Learning for Insulin Bolus Calculation in a Realistic Cohort of Patients with Type 1 Diabetes **17**

2.1 Generation of Virtual Patient Populations that Represent Real Type 1 Diabetes Cohorts 19

2.2 Bolus Insulin Calculation without Meal Information. A Reinforcement Learning Approach 35

2.3 An Automatic Deep Reinforcement Learning Bolus Calculator for Automated Insulin Delivery Systems 45

3 Discussion **73**

3.1 Generation of T1D Virtual Cohort 74

3.2 Reinforcement Learning for Insulin Bolus Dosing 76

3.3 Fully Automatic Insulin Delivery System 77

4 Conclusions **81**

4.1 Contributions 81

4.2 Future Work 83

Bibliography **85**

List of Tables

Table **Page**

1.1 Commercially availabe HCLs. 8

1.2 The state-of-the-art RL algorithms. 13

List of Figures

Figure **Page**

1.1 The blood glucose homeostasis in the human body. 3

1.2 Types of diabetes mellitus. 4

1.3 Block diagram of closed-loop automatic insulin delievery system. 7

ABSTRACT

Type 1 Diabetes (T1D) is a chronic metabolic disorder caused by destruction of the insulin producing beta cells in the islets of Langerhans within the pancreas due to an autoimmune reaction. T1D is distinguished by elevated levels of blood glucose (BG) owing to the deficiency of insulin, a hormone responsible for the regulation of BG within the normal range of 70-180 mg/dL. T1D is associated with various micro-vascular and macro-vascular complications such as nephropathy, neuropathy, retinopathy, coronary heart disease, cerebrovascular disease, peripheral artery disease etc. People with T1D rely on the administration of exogenous insulin to restrict the BG in a healthy range.

The insulin treatment strategies for T1D can be broadly divided into two categories i.e., multiple daily injections (MDI) and continuous subcutaneous insulin infusion (CSII) to avoid the T1D complications. In the past decade, a significant effort has been made by researchers to reproduce the behavior of beta cells and automate the insulin delivery for the management of T1D paving a way for the rapid development of the artificial pancreas (AP) technology. Integration of a continuous glucose monitor (CGM) with closed-loop control (CLC) algorithms to compute the continuous insulin dosing rate constitute an AP system. The preclinical validation and evaluation of the insulin dosing strategies developed by researchers are performed in the simulation environments that represent virtual patients (VPs) with T1D.

The work presented in this thesis provides three contributions. Firstly, a methodology is introduced to generate a cohort of VPs with T1D to replicate the BG metrics of a real cohort of people with T1D from the Hospital Clinic de Barcelona. The clinical data of meals, meal times and insulin (basal and bolus) was utilized to derive realistic scenarios for the generation of VPs. The exercise sessions were introduced as disturbances and were derived from the BG profile of the real patients. The proposed methodology is capable of adopting the daily variations of BG profile from real patients and thus provide a realistic and challenging simulation environment for the validation and evaluation of therapeutic strategies developed for the management of T1D.

Secondly, a Q-Learning based Reinforcement Learning (RL) algorithm is proposed for the bolus insulin calculation in patients with T1D and validated on the generated cohort of VPs with T1D. Usually the bolus insulin calculation is based on carbohydrates (CHO) in meal, CHO to insulin ratio (CR) and the insulin sensitivity based correction factor (CF). On the contrary, the proposed algorithm is independent of the CHO content in meals, CR and CF with an aim to avoid the CHO estimation and counting errors and the management burden on patients with T1D. The results were compared to the standard bolus calculator (SBC) as a baseline. The

proposed algorithm achieved similar performance as compared to the SBC and outperformed SBC in the presence of CHO estimation errors.

Finally, a fully automatic insulin delivery (FAID) system is presented. In hybrid automatic insulin delivery (HAID) system the meal disturbance is compensated by feed-forward control, which requires the announcement of the meal by the patient. To avoid the CHO estimation errors and the patient intervention, the FAID system was realized by integrating a novel Deep Reinforcement Learning (DRL) based bolus insulin dosing system, a proportional-derivative (PD) CLC strategy for the continuous delivery of insulin and an Unscented Kalman Filter (UKF) based meal detector. The PD controller and the UKF meal detector were previously developed by our group. The results of FAID system were compared to HAID system with CHO misestimation error as a baseline. The FAID system showed a satisfactory and comparable glycemic performance as compared to the HAID system.

The outcomes obtained from the algorithms presented in this thesis are encouraging. Firstly, customized cohort of VPs can help in validation of the treatment strategies for people with T1D under challenging conditions in simulation environment. Secondly, the RL algorithms could play a potential role in the future of developing advanced FAID system and AP technology.

RESUMEN

La diabetes tipo 1 (T1D) es un trastorno metabólico crónico causado por la destrucción de las células beta productoras de insulina en los islotes de Langerhans del páncreas debido a una reacción autoinmume. La T1D se caracteriza por niveles elevados de glucosa en sangre (BG) causados por la deficiencia de insulina, una de las hormonas responsables de la regulación de la BG dentro del rango normal de 70-180 mg/dL. La T1D se asocia a diversas complicaciones microvasculares y macrovasculares, como nefropatía, neuropatía, retinopatía, cardiopatía coronaria, enfermedad cerebrovascular, enfermedad de las arterias coronarias, arteriopatía periférica, etc. Las personas con T1D dependen de la administración de insulina exógena para mantener la glucemia en un rango saludable.

Las estrategias de tratamiento con insulina para evitar las complicaciones de la T1D pueden dividirse en dos categorías, la terapia con inyecciones diarias múltiples (MDI) y la infusión subcutánea continua de insulina (CSII). En la última década, los investigadores han realizado un gran esfuerzo para reproducir el comportamiento de las células beta y automatizar la administración de insulina para el tratamiento de la T1D, allanando el camino para el rápido desarrollo de la tecnología del páncreas artificial (AP). Un sistema de AP está integrado por un monitor continuo de glucosa (CGM) con algoritmos de control en bucle cerrado (CLC) para para calcular la tasa de dosificación continua de insulina. La validación preclínica y evaluación de las estrategias de dosificación de insulina desarrolladas por los investigadores se realizan en los entornos de simulación que representan a pacientes virtuales (VP) con T1D.

El trabajo presentado en esta tesis aporta tres contribuciones. En primer lugar, se introduce una metodología para generar una cohorte de VP con T1D que replica las métricas de glucemia de una cohorte real de personas con T1D del Hospital Clínic de Barcelona. Los datos clínicos de cantidad y hora de las comidas, así como la insulina (basal y en bolo), se utilizaron para derivar escenarios realistas para la generación de los VP. Las sesiones de ejercicio se introdujeron como perturbaciones y se derivaron del perfil de glucemia de los pacientes reales. La metodología propuesta es capaz de adoptar las variaciones diarias del perfil de BG de los pacientes reales y proporcionar así un entorno de simulación realista y desafiante para la validación y evaluación de estrategias terapéuticas desarrolladas para el tratamiento de la T1D.

En segundo lugar, se propone un algoritmo de Aprendizaje por Refuerzo (RL) basado en Q-Learning para el cálculo del bolo de insulina prandial en pacientes con T1D y se valida en la cohorte generada de VP con T1D. Tradicionalmente, el cálculo del bolo de insulina se basa en los carbohidratos (CHO) de la comida, la relación CHO/insulina (CR) y el factor de corrección (CF) basado en la sensibilidad a la insulina. A diferencia de esto, el algoritmo propuesto es

independiente del contenido de CHO en las comidas, del CR y del CF con el objetivo de evitar la estimación de CHO, los errores de recuento y la carga de gestión en pacientes con T1D. Los resultados se compararon con la calculadora de bolo estándar (SBC) como referencia. El algoritmo propuesto alcanzó un rendimiento similar al del SBC y superó al SBC en presencia de errores de estimación de CHO.

Por último, se presenta un sistema de administración de insulina totalmente automático (FAID). En el sistema híbrido de administración automática de insulina (HAID), la perturbación de la comida se compensa mediante un control feed-forward, que requiere el anuncio de la comida por parte del paciente. Para evitar los errores de estimación de CHO y la intervención del paciente, el sistema FAID se ha realizado integrando un novedoso sistema de aprendizaje profundo por refuerzo (DRL) para la dosificación de insulina en bolo, una estrategia de CLC proporcional-derivativo (PD) para la administración continua de insulina y un detector de comidas basado en el filtro "unscented" de Kalman (UKF). El controlador PD y el detector de comidas UKF fueron desarrollados previamente por nuestro grupo. Los resultados del sistema FAID se compararon con los del sistema HAID con errores de estimación de CHO como referencia. El sistema FAID mostró un rendimiento glucémico satisfactorio y comparable al del sistema HAID.

Los resultados obtenidos con los algoritmos presentados en esta tesis son alentadores. En primer lugar, la cohorte personalizada de VP puede ayudar en la validación de las estrategias de tratamiento para las personas con T1D en condiciones difíciles en un entorno de simulación. En segundo lugar, los algoritmos de RL podrían desempeñar un papel potencial en el futuro desarrollo de sistemas FAID avanzados y tecnología AP.

RESUM

La diabetis tipus 1 (DT1) és un trastorn metabòlic crònic causat per la destrucció de les cèl·lules beta productores d'insulina als illots de Langerhans dins del pàncrees a causa de una reacció autoimmune. La T1D es caracteritza per nivells elevats de glucosa en sang (BG) a causa de la deficiència d'insulina, una de les hormones responsables de la regulació de la glucosa dins del rang normal de 70-180 mg/dL. La T1D s'associa a diverses complicacions microvasculars i macrovasculars, com ara nefropatia, neuropatia, retinopatia, cardiopatia coronària, malaltia cerebrovascular, malaltia arterial perifèrica, etc. Les persones amb T1D depenen de l'administració d'insulina exògena per mantenir la glucèmia en un rang saludable.

Les estratègies de tractament amb insulina per evitar les complicacions de la T1D es poden dividir en dues categories, és a dir, múltiples injeccions diàries (MDI) o infusió contínua subcutània d'insulina (CSII). Durant l'última dècada, els investigadors han fet un esforç significatiu per reproduir el comportament de les cèl·lules beta i automatitzar el lliurament d'insulina per a la gestió de la T1D, obrint un camí per al desenvolupament ràpid de la tecnologia del pàncrees artificial (AP). L'integració d'un monitor continu de glucosa (CGM) amb algorismes de control de llaç tancat (CLC) per calcular la taxa de dosificació contínua d'insulina constitueix un sistema AP. La validació i avaluació preclínica de les estratègies de dosificació d'insulina desenvolupades pels investigadors es realitzen en els entorns de simulació que representen pacients virtuals (VP) amb T1D.

El treball presentat en aquesta tesi aporta tres contribucions. En primer lloc, una metodologia és introduïda per a la generació d'una cohort de VP amb T1D per replicar les mètriques de BG d'una cohort real de persones amb T1D de l'Hospital Clínic de Barcelona. Les dades clíniques dels àpats, els horaris dels àpats i la insulina (basal i bolus) es van utilitzar per obtenir escenaris realistes per a la generació de VP. Les sessions d'exercici es van introduir com pertorbacions i es van derivar del perfil de glucosa dels pacients reals. La metodologia proposada és capaç d'adoptar les variacions diàries del perfil de glucosa en pacients reals i, per tant, proporcionar un entorn de simulació realista i desafiant per a la validació i avaluació de les estratègies terapèutiques desenvolupades per al maneig de la T1D.

En segon lloc, es proposa un algorisme d'aprenentatge per reforç (RL) basat en Q-Learning per al càlcul de bolus d'insulina en pacients amb T1D i es valida en la cohort generada de VP amb T1D. Normalment, el càlcul de la insulina en bolus es basa en els hidrats de carboni (CHO) dels àpats, la proporció CHO a insulina (CR) i el factor de correcció basat en la sensibilitat a la insulina (CF). A diferència d'això, l'algoritme proposat és independent del contingut de CHO

en els àpats, del CR i del CF amb l'objectiu d'evitar els errors d'estimació i recompte de CHO i la càrrega de gestió dels pacients amb T1D. Els resultats es van comparar amb la calculadora de bolus estàndard (SBC) com a punt de referència. L'algorisme proposat va aconseguir un rendiment similar en comparació amb el SBC i va superar-lo en presència d'errors d'estimació de CHO.

Finalment, es presenta un sistema d'administració d'insulina totalment automàtic (FAID). En els sistemes híbrids de lliurament d'insulina (HAID), la pertorbació dels àpats es compensa mitjançant un control feed-forward, que requereix l'anunci de l'àpat per part del pacient. Per evitar els errors d'estimació de CHO i la intervenció del pacient, el sistema FAID es va dissenyar a partir de la integració d'un nou sistema de dosificació d'insulina en bolus basat en aprenentatge profund per reforç (DRL), una estratègia CLC derivada-proporcional (PD) per al lliurament continu d'insulina i un sistema detector d'àpats basat en el filtre "unscented" de Kalman (UKF). El controlador PD i el detector d'àpats UKF van ser desenvolupats prèviament pel nostre grup. Els resultats del sistema FAID es van comparar amb el sistema HAID amb un amb errors d'estimació de CHO com a referència. El sistema FAID va mostrar un rendiment glucèmic satisfactori i comparable en al del sistema HAID.

Els resultats obtinguts dels algorismes presentats en aquesta tesi són encoratjadors. En primer lloc, una cohort personalitzada de VP pot ajudar a validar les estratègies de tractament per a persones amb T1D en condicions difícils en un entorn de simulació. En segon lloc, els algorismes RL podrien tenir un paper potencial en el futur del desenvolupament de sistema avançats FAID i la tecnologia AP.

CHAPTER 1

INTRODUCTION

This chapter describes and provides an introduction to diabetes mellitus (DM) in Section 1.1; Section 1.2 particularly describes type 1 diabetes (T1D) with its associated problems and early treatment strategies. Then, Section 1.3 emphasizes the need of automatic insulin delivery (AID) system and provides an overview of the currently available AID systems for people with T1D. Section 1.4 introduces the role of artificial intelligence (AI) in diabetes technology and specifically highlights the related work done by researchers in the area of AI and reinforcement learning (RL) in an AID framework. Finally, the main objectives of the work are presented in Section 1.5, and Section 1.6 concludes with the organization of this document.

1.1 Diabetes Mellitus

DM is a metabolic disease distinguished by concurrent hyperglycemia as a result of a disorder in insulin secretion, insulin action, or both. The condition of elevated blood glucose (BG) levels above the normal range in DM is termed as hyperglycemia (Association, 2022). The

consequences of chronic hyperglycemia include damage, failure and dysfunction of several organs, e.g., kidney, retina, heart, blood vessels and nervous system (Association, 2014). In healthy individuals, the BG levels are maintained in a normal range of 70 mg/dL to 180 mg/dL.

The blood glucose homeostasis in the human body is accomplished by a complex network of hormones mainly secreted by the brain, pancreas, liver, intestine and adipose and muscle tissue. Pancreas play a vital role within this network by secreting the glucose lowering hormone insulin and together with glucagon, which increases the glucose in blood plasma (Röder et al., 2016). The alpha cells in the pancreas are responsible for the secretion of glucagon whenever the BG drops below the normal range and the beta cells in pancreas are responsible for the secretion of insulin in case of elevated BG to achieve the normoglycemia in healthy individuals (Göke, 2008). The two key biological processes involved are the glycogenolysis and glycogenesis. Glycogenolysis is a mechanism that converts glycogen, the major carbohydrate stored in liver and muscle cells, into glucose in order to provide energy quickly and keep BG levels stable during fasting. Glycogenesis is the procedure through which glucose molecules are incorporated into chains of glycogen for storage. This mechanism is triggered in the liver during periods of rest and is also triggered by insulin in response to high BG. The homeostasis of BG concentration in the human body is presented in figure 1.1.

DM can be broadly classified into two categories namely, T1D and type 2 diabetes (T2D). T1D is characterized by the absence of insulin due to the destruction of the insulin secreting beta cells in pancreas as a consequence of an autoimmune reaction. The chronic T1D is associated with micro-vascular and macro-vascular complications (Katsarou et al., 2017). T2D is distinguished by a metabolic disorder in carbohydrate (CHO), lipid and protein and caused by the insufficient insulin secretion, insulin resistance or a combination of both (DeFronzo et al., 2015). The types of DM are presented in figure 1.2.

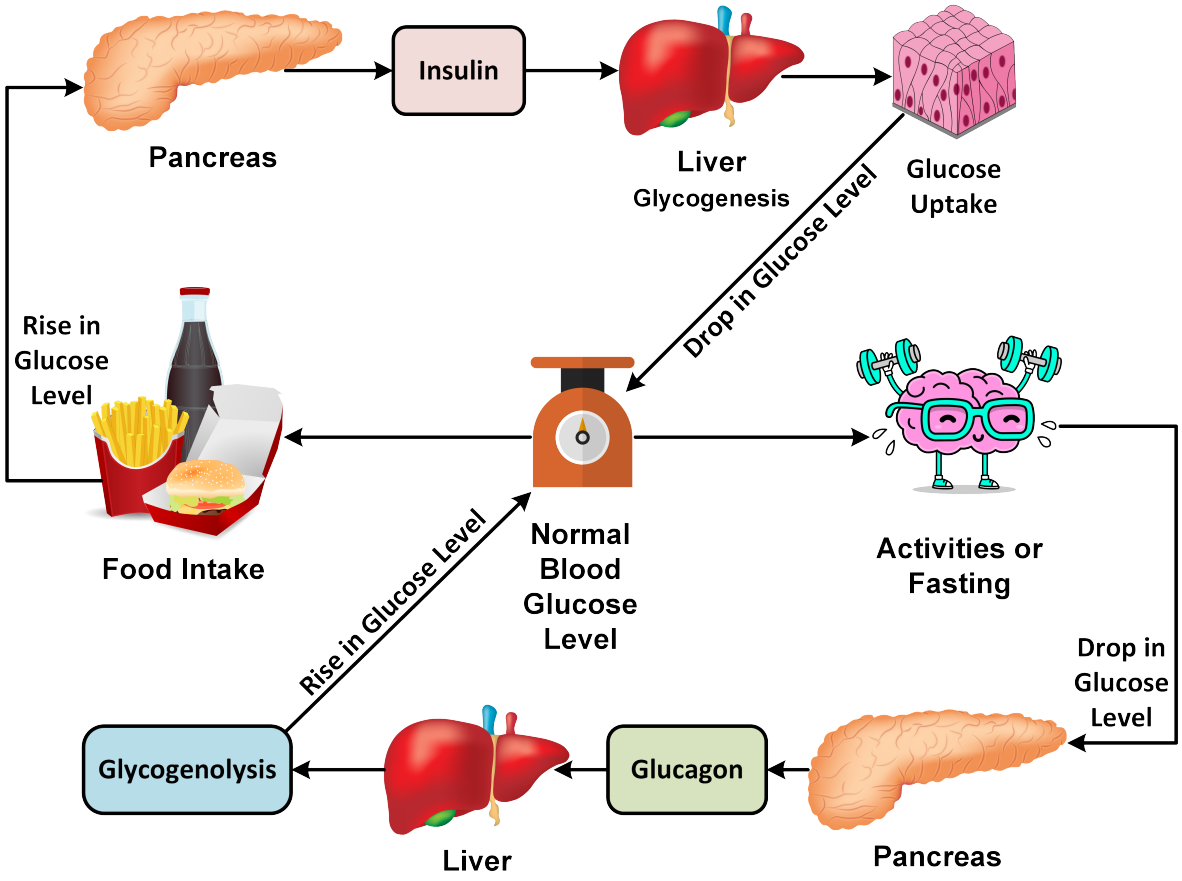

Figure 1.1: The blood glucose homeostasis in the human body.

1.2 Type 1 Diabetes Mellitus

A century ago T1D turned form an acutely fatal disease into a chronic condition when insulin was discovered by Banting and Charles Best (Banting et al., 1922). T1D if not properly treated is associated with several complications such as retinopathy, glaucoma, cataracts, neuropathy, nephropathy, and cardiovascular diseases (Melendez-Ramirez et al., 2010). People with T1D rely on exogenous insulin administration and the Diabetes Control and Complication Trail (DCCT) revealed that proper insulin therapy results in decreasing the likelihood of long-term complications associated with T1D (Control and Group, 1993).

The majority of individuals with T1D receive insulin therapy either as multiple daily injections (MDI) or as continuous subcutaneous insulin infusion (CSII). The MDI therapy

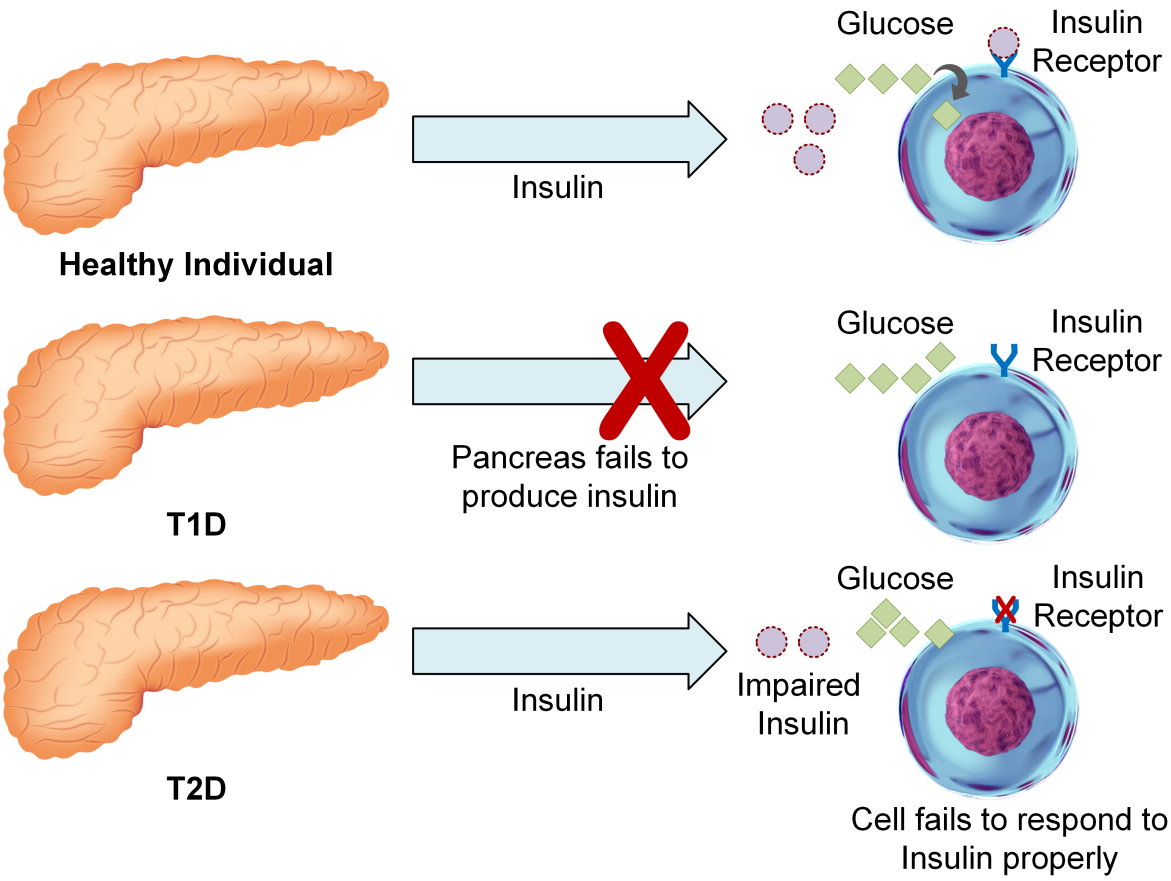

Figure 1.2: Types of diabetes mellitus.

involves using long-acting basal insulin, usually delivered as a bolus once or twice a day, together with rapid-acting insulin delivered to compensate for the CHO intake during meals. CSII or insulin-pump therapy provides several advancements in terms of programmable basal rates for insulin infusion during the day and night as well as various patterns for the delivery of bolus insulin at meals (Pickup, 2012).

There exist a debate among researchers after the introduction of CSII therapy for more than 40 years about the effectiveness of MDI as compared to CSII therapies. It has been reported (Pickup, 2019) that in certain people with T1D, MDI and CSII can achieve tight BG control without hypoglycemia, especially in those who are motivated, have received structured diabetes education, and have a high level of ongoing support from medical professionals. CSII therapy is

particularly effective in case of individuals associated with continued higher HbA_{1c}, significant dawn phenomenon, higher glycemic variability and/or exhibiting frequent hypoglycemia. In terms of cost-effectiveness a systematic review in eight countries (Australia, Canada, Denmark, Spain, Poland, Italy, UK and USA) revealed that CSII is 1.4 times more costly as compared to MDI. Although, this cost is partially neutralized by savings from the improved metabolic control resulting in reduced T1D complications (Roze et al., 2015).

Achieving optimal insulin delivery to maintain normoglycemia in people with T1D using either MDI or CSII is difficult without a high commitment from the people and a proper support from the diabetes health care staff. The difficulty in maintaining normal BG in free living conditions is associated with the variations in insulin requirements owing to many factors (stress, physical activity, uncertain dynamics, etc) leading to a mismatch in the timely insulin delivery as per the individual's requirement (Renard, 2022). The failure to achieve optimal glycemic control internationally was emphasized by an observational study on the management of T1D in adults, which also included recommendations for improvements in self-management of insulin therapy, the use of technology like CGM, and the provision of healthcare support (Renard et al., 2021).

1.3 Automatic Insulin Delivery Systems

The AID also referred to as artificial pancreas (AP) for patients with T1D date back to the 1970s. A typical AID system is an integration of three major components i.e., insulin infusion system, a control algorithm that modulates the insulin delivery according to the patients's requirement and a continuous glucose monitoring (CGM) device. The early AID systems used the intravenous route for the infusion of insulin and sensing of BG such as Biostator (Clemens, 1977). The portable AID systems were introduced in the 1980s, mostly utilizing the subcutaneous (SC) route for insulin infusion. Still, the availability of accurate and reliable CGM sensors was a problem to allow the outpatient trials. The development of relatively accurate

and safe SC CGM sensors from 1999 pave the way for the technological advancement of AID systems (Mastrototaro, 1999).

The modern CGM device consists of a disposal sensor to measure the glucose in the interstitial fluid (approximation of BG value) and a transmitter to trasmit the BG value (usually every 5 minutes) to the controller, smartphone or the cloud (Shah et al., 2018). The three major CGM devices (Minimed, Glucowatch and Medtronic) were developed in 2000s and were used in clinical trials to demonstrate the benefits of using CGM technology in T1D management (Group, 2008). In recent developments, the Dexcom G5 and G6 and the Abott Freestyle Libre sensors have been approved by the US Food and Drug administration (FDA) to be used as a CGM device in advanced AID systems. The data from modern CGM devices now allows to visualize the glycemic patterns and standardized reports i.e., the ambulatory glucose profile, providing the core CGM metrics (Battelino et al., 2019).

The commonly used CLC algorithms in commercially available AID systems are either based on proportional integral derivative (PID) control, model predictive control (MPC) or fuzzy logic control (Nwokolo and Hovorka, 2023). PID controllers compute the insulin infusion rate to deliver by considering the glucose error term (difference in real time BG value and target BG value) from three perspectives: the magnitude of the error (proportional part), the area under the curve of the error (integral part) and the rate of change of the error (derivative part). The MPC algorithms adjust the insulin delivery by minimizing the difference between predicted BG value and the target BG value over a pre-designed prediction window. In fuzzy logic control the insulin is calculated based on a set of rules exploiting the empirical knowledge of diabetes clinicians (Boughton and Hovorka, 2021). The block diagram of a typical CL AID system is shown in the figure 1.3.

The CLC technology has improved the quality of life in people with T1D. The improvements experienced by the patients include reduction in anxiety, less restriction in eating habits, improved sleep, increased confidence owing to the reduction in nocturnal hypoglycemia and

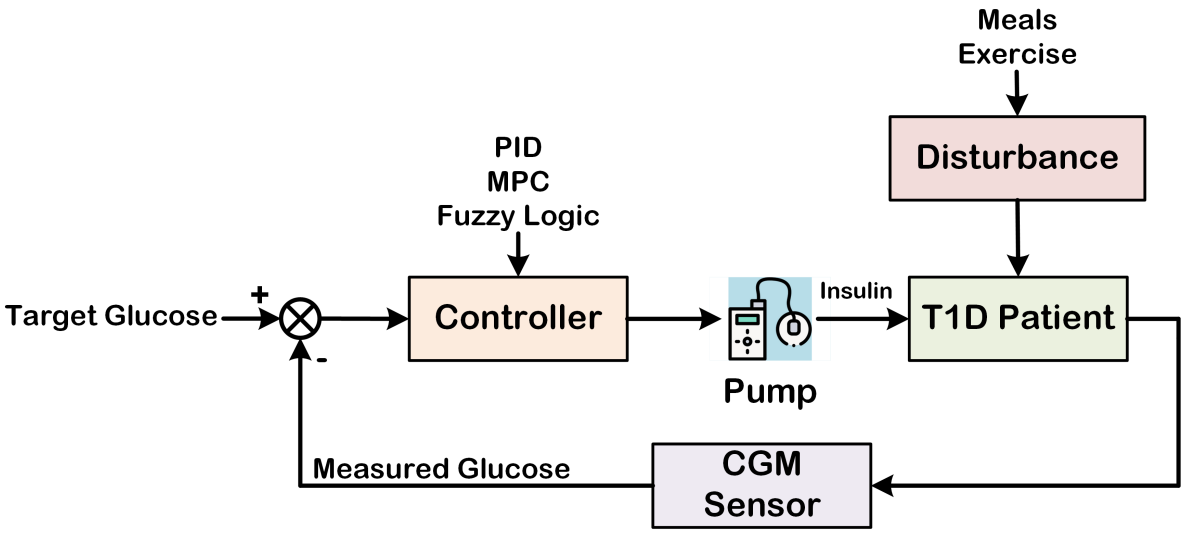

Figure 1.3: Block diagram of closed-loop automatic insulin delievery system.

a partial release from the burden of diabetes management (Farrington, 2018). However, the currently available AID technology is not fully automated. It requires the intervention of patients to announce disturbances such as meals and exercise to the system and therefore, is termed as hybrid closed-loop system (HCL). A positive user experience with improved glycemic control and reduction in diabetes management burden using the HCL is reported based on the use of HCL for six months (Roberts et al., 2022).

The first commercially approved HCL was the Medtronic 670G whose safety and efficacy was investigated in 30 adolescents and 94 adults with T1D over a period of 3 months (Garg et al., 2017). The time spent in normoglycemia increased from 60% to 67% in adolescents and from 69% to 74% in adults comparing baseline to the CL systems respectively. Medtronic 780G advanced HCL (AHCL) has been developed with additional features such as adjustable glucose targets and automated correction boluses. The glycemic outcomes of the Medtronic 780G reported in a recent study concluded that the AHCL progresses towards a significant improvement of the overall glucose control in patients with T1D (Pintaudi et al., 2022). The Control IQ system (t:slim X2 pump with Dexcom G6 sensor [Tandem, San Diego, CA, USA]) was compared to the sensor-augmented pump therapy (SAP) in a randomized controlled study

HCL System	Medtronic 670G/770G	Medtronic 780G	Tandem Control-IQ	CamAPS FX	Diabeloop DBLG 1/Hu	Omnipod 5
Main Control Algorithm	PID	PID	Linear MPC (control-to-range)	MPC	MPC Machine Learning Reinforcement Learning	MPC
Insulin Pump	670G/770G	780G	Tandem t:slim X2	DANA Diabecare RS DANA-i	Kaliedo patch AccuChek Insight Dana-i pump EOFlow	Omnipod 5 patch pumps
CGM System	Guardian Link 3	Guardian Link 3/4	Dexcom G6	Dexcom G6	Dexcom G6	Dexcom G6
User Interface	Android Apple phone (view only)	Android Apple phone (view only)	t:connect Mobile for view and mobile bolus in US	Android phone	Separate handheld device or Android or Apple phone app	System "Controller" Android-based device
FDA Approval	Age ≥ 7		Age ≥ 6			Age ≥ 2

Table 1.1: Commercially availabe HCLs.

over a period of 6 months. The patients were divided in a 2:1 ratio with the CL therapy (CL group) and the SAP therapy (control group). The time spent in normoglycemia increased from 61% to 71% in the CL group and remained unchanged at 59% in the control group (Brown et al., 2019). The details of commercially available AID systems with the associated performance are presented in recent review articles (Rodríguez-Sarmiento et al., 2022; Peacock et al., 2023) and table 1.1 presents a summary of the system components characteristics.

The AID systems also use pramlintide, glucagon, and insulin formulations that are currently on the market. Although, there has been research into developing more effective glucagon formulations (Steiner et al., 2010), glucagon is frequently challenging to work with since it cannot remain stable in solution. The controller must be implemented utilizing readily available computers, tablets, smartphones, or other small computer devices. All of these options should take battery life into account, especially when Bluetooth or other wireless connectivity is required. As existing commercial devices are used in current AID system designs, it is crucial that these designs be resilient to communication issues that may arise between various hardware components. They also need to include failure modes in case of issues like signal interruptions from transmission or intrinsic sensor loss. Prior to clinical application, it is crucial for current systems to show they can handle these issues during their validation phases (Doyle III et al., 2014).

1.4 Artificial Intelligence for Diabetes

Artificial intelligence (AI) can be defined in several ways but one of the widely accepted definitions is to make computers able to perform tasks that require human intelligence (Boden, 1980). Intelligence is a combination of capabilities such as understanding, learning and reasoning for solving problems and making decisions. AI utilizes various tools and methodologies to emulate such aspects of human intelligence. As a result of a rapid increase in the computational resources and computer performance, the mainstream technologies of AI and machine learning (ML) in 2023 i.e., supervised ML (SL), unsupervised ML (USL) and reinforcement learning (RL) have made a significant progress. ML is broadly defined as a computational methodology that learns from its past experience to improve the performance of making predictions accurately (Schapire and Freund, 2012). SL uses a set of well defined data and is focused on the classification of data and USL uses unlabeled data and is focused on identifying patterns in the data. RL is a trial and error based mechanism relying on a reward methodology to learn and optimize the performance of the learning agent to perform certain task (Sutton and Barto, 2018). Moreover, the term deep learning (DL) in AI and ML refers to multiple processing layers of ML models to allow learning of data representations with multiple layers of abstraction (LeCun et al., 2015).

As the AI software and hardware capabilities are growing rapidly, the researchers in the biomedical field are actively proposing AI algorithms to improve the diagnostic and treatment abilities and efficacy of the healthcare systems. AI algorithms have been extensively used by the researchers for the diagnosis and detection of many diseases such as Alzheimer's disease, cancer, diabetes, chronic diseases, heart diseases, tuberculosis, cerebrovascular diseases, hypertension, skin diseases, liver diseases etc. For example a ML algorithm was proposed for the diagnosis of Alzheimer's disease (Khan and Zubair, 2022), a recent review presents the AI algorithms proposed by the researches for the diagnosis and prognosis of cancer along with highlighting the future challenges (Huang et al., 2020), the applications of ML prediction models for the

diagnosis of the chronic diseases can be seen in a recent review (Battineni et al., 2020), a recent systematic review presents the developments in ML algorithms for the diagnosis of heart diseases (Ahsan and Siddique, 2022) and an AI based diagnostic system for kidney diseases such as cyst, stone and tumor was recently proposed (Islam et al., 2022).

Researchers in the field of diabetes technology actively proposed AI and ML based algorithms in various domains of diabetes such as diabetes education, prediction of diabetes, diagnosis of diabetic complications, self-management of diabetes, decision support systems for diabetes (DSSD), forecasting glucose values, detecting glycemic patterns, detecting hypoglycemia, optimizing insulin therapy in T1D and bolus insulin calculation in T1D. AI based learning methodologies are presented in a recent systematic review for the prediction of diabetes mentioning the algorithms and datasets used along with the accuracy achieved (Kaul and Kumar, 2020). Another review features the applications of AI and ML in the detection, diagnosis and self-management of diabetes from the perspective of six different aspects i.e., datasets, methods of data pre-processing, feature extraction methodology, ML algorithm, classification phenomenon and performance measures (Chaki et al., 2022). A review explains the proposed AI algorithms for DSSD (T1D) regarding the personalized recommendations considering insulin doses and daily behaviors (Tyler and Jacobs, 2020). Our research group had also reviewed the proposed ML algorithms for the prediction of hypoglycemia along with latest trends and challenges that the researchers face in this area (Mujahid et al., 2021). The above discussion show that the research in the field of applications of AI and ML algorithms in diabetes technology have been grown rapidly.

1.4.1 Reinforcement Learning in Diabetes Technology

RL is a way of learning that maps situations to actions with an aim of maximizing a numerical reward signal. The optimized actions are not fed to the learning agent but must be discovered by hit-and-trial mechanism. The actions taken not only affect the immediate reward values,

but also the next situation and thus all subsequent rewards. The trial and error search and the net reward characteristics are the most distinguished features of RL. RL is different from SL because the learning is not based on a labeled set of training data. RL is also different from USL because in USL the learner identify hidden patterns in a set of unlabeled data (Sutton and Barto, 2018).

In RL the main elements are an agent, an environment, a policy, a reward signal, a value function and a model of the environment, which is optional and depends on the specific RL algorithm. The agent is the learner to achieve the goal, the environment is the entity with which the agent interacts, the policy defines the way an agent interacts with the environment at a given time, a reward signal translates the immediate actions of the agent towards the goal of the RL problem, a value function quantifies the optimal behavior towards achieving the goal in a long run and the model is a mathematical representation of the environment. For a problem to be solved by RL, it first needs to be mathematically formalized as a Markov decision process (MDP). A MDP is a stochastic decision-making process that makes sequential decisions over time and employs a mathematical framework to simulate the decision-making of a dynamic system in situations where the outcomes are either random or controlled by a decision-maker. The decision-maker or learner in this context is the RL agent and the entity with which the agent is interacting is the environment. The agent interacts with the environment by taking an action ($A_t \in A$) at discrete time step (t). As a result, the change in environment is observed in terms of a state ($S_{t+1} \in S$) and the agent receives an immediate numerical reward ($R_{t+1} \in R$), quantifying the suitability of the taken action. Thus, in an MDP the interaction of the agent with an environment give rise to a sequence: $S_0,\ A_0,\ R_1,\ S_1,\ A_1,\ R_2,\ S_2, ...$ In a finite MDP the A, S, and R are all finite sets (Sutton and Barto, 2018).

In the last decade, RL has increasingly been adopted by the researchers and scientists in the field of diabetes technology. As RL is rapidly emerging as a control technique (Wang and Hong, 2020), it is used by most researchers as a feedback controller in an AID framework. In early

studies, The Diabetes Technology Research Group from the University of Bern presented an actor-critic form of RL for the control of BG in people with T1D. Both basal and bolus insulin dosage was managed and the initialization was based on clinical guidelines (Daskalaki et al., 2013). The algorithm was validated using the educational version of the UVa/Padova T1D simulator (Kovatchev et al., 2009). Later, a feasibility report was presented to pave the way for the testing of the proposed algorithm in clinical settings (Daskalaki et al., 2016). In 2018, the group proposed an adaptive basal insulin rate and insulin to carbohydrates (CHO) ratio advisor based on the RL algorithm and evaluation was performed on a cohort of 100 patients from UVa/Padova simulator (Sun et al., 2018). (De Paula, Avila and Martinez, 2015) combined the RL with Gaussian processes to cope with the uncertainties and variability in glucose dynamics utilizing one of the minimal models of glucose-insulin dynamics (Lehmann and Deutsch, 1992). They also presented another study focusing on personalizing and adding the online policy learning feature to the developed algorithm (De Paula, Acosta and Martínez, 2015). (Tejedor et al., 2020) presented a detailed systematic review of the RL algorithms developed by the researchers in the field of diabetes technology highlighting the key components of RL (states, actions, rewards) and the simulation environment used for the validation of the developed RL algorithms.

In the past few years, various RL based approaches for BG regulation for people with T1D have been proposed. (Zhu, Li, Herrero and Georgiou, 2020) proposed a deep RL algorithm for the single (insulin only) and dual hormone (insulin and glucagon) delivery to regulate the BG in people with T1D and considering the UVa/Padova simulator for validation. They also presented a bolus insulin advisor based on deep RL (actor-critic model) utilizing the relation of standard bolus calculator (SBC) for patients with T1D and using a dual learning methodology (general and personalized) (Zhu, Li, Kuang, Herrero and Georgiou, 2020). (Nordhaug Myhre et al., 2020) proposed a model-free deep policy gradient version of RL for the BG control problem in people with T1D. The Hovorka model was used for the purpose of validation of the

proposed algorithm (Wilinska and Hovorka, 2008). (Lee et al., 2020) proposed a bio-inspired reward function driven RL algorithm for AID with unannounced meals and the experiments were performed on UVa/Padova T1D simulator. (Jafar et al., 2021) proposed RL algorithm to adjust the programmable basal rates and CHO-to-insulin ratio (CR) for insulin therapy in a hybrid AP setting while the overall insulin delivery was managed by MPC controller. (Lim et al., 2021) proposed an actor-critic version of RL guided by PID control in the early stage of learning process for the purpose of safety in a BG control framework and results comparable to PID control were presented. (Viroonluecha et al., 2022) presented a DRL algorithm for CLC of BG in people with T1D and considering the basal insulin as a control variable. The states were based on CGM samples only and CHO information was left for the user to be announced. A simplified reward function strategy was presented and the results were compared to PID control. (Emerson et al., 2023) recently presented the application of offline RL algorithms for insulin dosing policies to eliminate potentially risky patient interaction during training. The state-of-the-art algorithms are summarized in the table 1.2.

Algorithm	**Therapy**	**Training Platfrom**	**Citation**
Deep Reinforcement Learning (double Q-learning with dilated recurrent neural networks)	Basal insulin only and both insulin and glucagon	UVa/Padova T1D Simulator	(Zhu, Li, Herrero and Georgiou, 2020)
Deep Reinforcement Learning (actor-critic model based on deep deterministic policy gradient)	Bolus insulin	UVa/Padova T1D Simulator	(Zhu, Li, Kuang, Herrero and Georgiou, 2020)
Policy Gradient Reinforcement Learning (trust-region policy optimization algorithm)	Basal insulin	Hovorka T1D Simulator	(Nordhaug Myhre et al., 2020)
Bioinspired Reinforcement Learning (proximal policy optimization)	Continuous insulin delievery	UVa/Padova T1D Simulator	(Lee et al., 2020)
Reinforcement Learning (Q Learning)	Programmable basal rates and carbohydrate-to-insulin ratio	Hovorka T1D Simulator	(Jafar et al., 2021)
Reinforcement Learning (soft actor-critic)	Continuous insulin delievery	UVa/Padova T1D Simulator	(Lim et al., 2021)
Deep Reinforcement Learning (soft actor-critic model and deep deterministic policy gradient)	Basal insulin	UVa/Padova T1D Simulator	(Viroonluecha et al., 2022)
Offline Reinforcement Learning (batch constrained deep Q-learning, conservative Q-learning and) (twin delayed deep deterministic policy gradient)	Basal insulin	UVa/Padova T1D Simulator	(Emerson et al., 2023)

Table 1.2: The state-of-the-art RL algorithms.

1.5 Thesis Objectives

The goal of this thesis is to investigate and develop RL algorithms for the calculation of optimal insulin bolus doses without patient intervention in AID systems. Additionally, to develop a realistic cohort of of virtual patients (VPs) with T1D to provide a challenging simulation environment for in-silico validation of the developed algorithms.

The research objectives in this thesis can be further split into the following specific goals:

- To develop a realistic cohort of people with T1D. To achieve this objective clinical data of a cohort of 14 individuals with T1D form the Hospital Clínic de Barcelona has been used. A virtual cohort is generated with the goal to achieve statistically similar glycemic behaviour as compared to the original cohort. The core glycemic metrics of the real and the generated cohort are presented.

- To design a RL algorithm for bolus insulin calculation with no information about CHO content in meals and patient's specific parameters (CR and insulin sensitivity based correction factor (CF)) for people with T1D. The calculation of CHO content in meals is associated with estimation error and adds extra burden on patients. The goal of the proposed algorithm is to overcome these problems.

- Implementation of a fully CL AID system for people with T1D. This is achieved by clinically validated CL controller, an advanced version of the RL algorithm and a meal detector. The unannounced meals are firstly detected by utilizing an algorithm previously developed by our research group and secondly the bolus insulin is delivered through the proposed RL algorithm. A CL controller previously developed by our research group provides the continuous delivery of insulin.

1.6 Thesis Structure

This document is organized as follows: A copy of the articles allowing presentation of this thesis as a compendium of publications makes up Chapter 2. The primary contributions of the publications that make up this thesis are briefly discussed in Chapter 3. Finally, the conclusions and future work are presented in Chapter 4.

CHAPTER 2

REINFORCEMENT LEARNING FOR INSULIN BOLUS CALCULATION IN A REALISTIC COHORT OF PATIENTS WITH TYPE 1 DIABETES

This chapter consists of three sections. Section 2.1 presents a paper which develops a methodology to generate a cohort of VPs with T1D that represents a cohort from the Hospital Clinic de Barcelona. The methodology utilizes clinical data of real patients from the Hospital Clinic de Barcelona. Section 2.2 consists of a paper in which RL based insulin bolus calculator is presented. The proposed algorithm for the calculation of insulin bolus is validated on the VPs described in the section 2.1. Section 2.3 presents a paper that proposes a FAID system for people with T1D.

- **2.1 Generation of Virtual Patient Populations that Represent Real Type 1 Diabetes Cohorts**

- **2.2 Bolus Insulin Calculation without Meal Information. A Reinforcement Learning Approach**

- **2.3 An Automatic Deep Reinforcement Learning Bolus Calculator for Automated Insulin Delivery Systems**

2.1 Generation of Virtual Patient Populations that Represent Real Type 1 Diabetes Cohorts

In this publication, we propose a methodology to develop a cohort of VPs with T1D that exhibits the glycemic features of a cohort from Hospital Clinic de Barcelona. The main focus is to develop a methodology of generating a realistic cohort of VPs with T1D to provide a challenging simulation environment for the validation and development of treatment strategies for patients with T1D. As part of the candidate's contribution to this publication, the VPs generation approach was developed, designed, and implemented. The candidate also participated in the discussion and edited it throughout the review process. Dr. Charrise Ramkissoon and Dr. Aleix Beneyto assisted the candidate while the work was being developed. Dr. Ignacio Conget and Dr. Marga Giménez managed and provided the clinical data. Dr. Josep Vehí supervised the candidate during the development of the work.

Title: Generation of Virtual Patient Populations that Represent Real Type 1 Diabetes Cohorts

Authors: **Sayyar Ahmad**, Charrise M. Ramkissoon, Aleix Beneyto, Ignacio Conget, Marga Giménez and Josep Vehí

Journal: Mathematics

Volume: 9, Issue: 11, Pages: 1200–1214, Published: May 2021

DOI: https://doi.org/10.3390/math9111200

Quality index: SCIE 2021 Mathematics, Impact factor: 2.592, Q1 (21/332)

mathematics

MDPI

Article

Generation of Virtual Patient Populations That Represent Real Type 1 Diabetes Cohorts

Sayyar Ahmad [1], Charrise M. Ramkissoon [1], Aleix Beneyto [1], Ignacio Conget [2,3], Marga Giménez [2,3] and Josep Vehi [1,*]

[1] Institute of Informatics and Applications, University of Girona, 17003 Girona, Spain; sayyar.ahmad@udg.edu (S.A.); charrise.ramkissoon@udg.edu (C.M.R.); aleix.beneyto@udg.edu (A.B.)
[2] Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 28001 Madrid, Spain; iconget@clinic.cat (I.C.); gimenez@clinic.cat (M.G.)
[3] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08023 Barcelona, Spain
* Correspondence: josep.vehi@udg.edu; Tel.: +34-620131826

Abstract: Preclinical testing and validation of therapeutic strategies developed for patients with type 1 diabetes (T1D) require a cohort of virtual patients (VPs). However, current simulators provide a limited number of VPs, lack real-life scenarios, and inadequately represent intra- and inter-day variability in insulin sensitivity and blood glucose (BG) profile. The generation of a realistic scenario was achieved by using the meal patterns, insulin profiles (basal and bolus), and exercise sessions estimated as disturbances using clinical data from a cohort of 14 T1D patients using the Medtronic 640G insulin pump provided by the Hospital Clínic de Barcelona. The UVa/Padova's cohort of adult patients was used for the generation of a new cohort of VPs. Insulin model parameters were optimized and adjusted in a day-by-day fashion to replicate the clinical data to create a cohort of 75 VPs. All primary and secondary outcomes reflecting the BG profile of a T1D patient were analyzed and compared to the clinical data. The mean BG 166.3 versus 162.2 mg/dL ($p = 0.19$), coefficient of variation 32% versus 33% ($p = 0.54$), and percent of time in range (70 to 180 mg/dL) 59.6% versus 66.8% ($p = 0.35$) were achieved. The proposed methodology for generating a cohort of VPs is capable of mimicking the BG metrics of a real cohort of T1D patients from the Hospital Clínic de Barcelona. It can adopt the inter-day variations in the BG profile, similar to the observed clinical data, and thus provide a benchmark for preclinical testing of control techniques and therapy strategies for T1D patients.

Keywords: type 1 diabetes; virtual patients; type 1 diabetes simulator; artificial pancreas

Citation: Ahmad, S.; Ramkissoon, C.M.; Beneyto, A.; Conget, I.; Giménez, M.; Vehi, J. Generation of Virtual Patient Populations That Represent Real Type 1 Diabetes Cohorts. *Mathematics* **2021**, *9*, 1200. https://doi.org/10.3390/math9111200

Academic Editor: Maria Laura Manca

Received: 31 March 2021
Accepted: 23 May 2021
Published: 25 May 2021

1. Introduction

Type 1 diabetes (T1D) is a disorder characterized by the destruction of insulin producing beta cells in the pancreas due to an autoimmune reaction [1]. A great amount of research effort has been made in the past few decades to automate the insulin delivery for the treatment of people with T1D, leading to a rapid increase in artificial pancreas (AP) technology. A basic AP system integrates a closed-loop control algorithm, continuous glucose monitoring (CGM), and subcutaneous continuous insulin infusion for optimum blood glucose (BG) control [2]. For pre-clinical testing and validation of therapeutic strategies used in AP technology, various simulators have been developed. The most well-known simulators used in AP research are the University of Virginia/Padova (UVa/Padova) simulator [3], Oregon Health and Science University (OHSU) simulator [4], and Cambridge Simulator [5]. The use of simulators is vital in the development of healthcare technologies, which allows significant research to be performed at an accelerated rate while circumventing unnecessary risks to the patient and costs related to animal or clinical testing [6]. Simulators have played a prominent role in the development of many important areas of biomedical research, such as anesthesia administration [7], HIV therapy [8], minimal

invasive surgery techniques [9], acute ischemic stroke treatment [10], drugs' cardiotoxicity assessment [11], drugs' metabolism prediction [12], vaccine target identification (COVID-19) [13], etc. Due to its flexibility and relative low patient risk, in silico experimentation via simulators is becoming increasingly used in healthcare research centers [14]. In addition, in silico testing of therapeutic approaches for the management of T1D is considered a prerequisite before proceeding to clinical trials [15]. However, it is important to note that computer simulations provide a safe ground for testing the new therapeutic strategies but cannot substitute clinical trials.

The main challenges for BG regulation in T1D are the disturbances in terms of meals, exercise, stress, and variability (inter-patient and intra-patient). The UVa/Padova simulator allows the incorporation of different meal scenarios for the virtual patient (VP) population, allowing researchers to analyze the effectiveness of a control algorithm [16–22], validate optimization and adaptation strategies for insulin delivery [23–26], develop disturbance detection algorithms for meals [27–29] and exercise [30], develop methods for mitigating the risks of hypoglycemia [31,32], and integrate machine learning algorithms into conventional diabetes therapy and bolus calculator for the treatment of T1D patients [33–35]. In the literature, the meal scenarios used for testing BG regulation effectiveness are based on typical values considering three meals per day [36–47]. However, in real life, the amount of carbohydrate intake and number of meals per day may vary patient to patient. Data from a cohort of 14 T1D patients obtained from the Hospital Clinic de Barcelona shows that daily carbohydrate (median) intake ranges from 36.79 to 186.43 g (SD = 46.85) with 3 to 7 (SD = 1.27) meals (median) per day. This difference in conventional and real scenarios can lead to under- or over-performance of the tool or methodology developed for the treatment of T1D patients. Therefore, real-life scenarios that include meals, exercise, inter-day BG variability, and other variations for use in simulation are still lacking. These types of scenarios for developing T1D management technology are vital in providing key information about safety and limitations of proposed treatment strategies, while avoiding unnecessary expenses.

Additionally, the virtual patients (VPs) available in the current T1D simulators exclude certain sub-cohorts of patients, such as high variability, hypoglycemic-prone, hypoglycemia unawareness, pregnant, menstruating, and additional comorbidities. Moreover, current simulators only offer a limited number of VPs. An academic version of the UVa/Padova simulator available for researchers consists of three groups of 10 VPs corresponding to children, adolescent, and adult populations. The original Oregon Health and Science University (OHSU) (2004) [48] simulator is composed of 6 VPs, and Chassin et al. introduced a cohort comprised of 18 virtual subjects [49].

To address current downfalls of VP populations in simulators, several methodologies to generate larger cohorts of T1D VPs have been developed. Haider et al. [50] proposed a probabilistic method for the generation of virtual subjects. Clinical data from 12 young T1D patients was used to test the methodology. Resalat et al. [4] proposed a statistical method to generate a population of T1D VPs mainly based on variants of the Hovorka model. The selection criterion of VP was based on clinical data from 20 patients undergoing artificial pancreas (AP) trials. The parameters used for comparison were percent of time (PoT) in normoglycemia (BG in 70 to 180 mg/dL), hyperglycemia (BG > 180 mg/dL), and hypoglycemia (BG < 70 mg/dL). Orozco-Lopez et al. [51] proposed a methodology to generate a large cohort of VPs. An already available cohort in the OHSU simulator was utilized to generate more VPs by establishing a relationship between the subject's parameters in terms of covariance illustrated in the Hovorka model.

The first UVa/Padova simulator was approved by the FDA in 2008 [3] for a single-meal scenario only. The VPs were represented by a set of model parameters which were extracted randomly from joint distributions of parameters. A new version was published in 2014 [52], in which improved glucose kinetics in hypoglycemia and glucagon dynamics were implemented. The virtual population was also improved in terms of clinical parameters such as carbohydrates ratio and correction factor. This version was also approved

by the FDA for single-meal scenarios only. It was mentioned that including day-by-day variations in parameters was under investigation. The latest version was published in 2018 [53]. The diurnal variations in glucose due to insulin sensitivity were added to extend the scope of the simulator from single-meal to multiple meal and multiple day scenarios.

T1D simulators should offer cohorts of VPs capable of mimicking BG dynamics of real patients. In real life, the BG curves vary from day to day due to significant nonlinearities and time varying effects. In available simulators, the BG curves usually follow certain patterns depending on the meals delivered per day. In order to analyze and validate the treatment and therapeutic strategies developed for patients with T1D, VPs offering more realist BG behavior are required. This study is focused on capturing the day-to-day variations in BG and generating VPs to reflect the BG outcomes of specific real T1D cohorts.

In this work, a methodology to modify the cohort of UVa/Padova's adult population is proposed to represent a cohort of T1D patients from the Hospital Clínic de Barcelona. This sub-cohort is not currently represented in the existing adult population. The novelty of the proposed scheme lies in the day-to-day adjustment of key physiological parameters that represent true phenomena in real T1D patients. Optimization of these parameters to capture the glycemic outcomes of the real cohort is a primary contribution of the presented study. This provides a benchmark for testing and validation of control algorithms and treatment strategies developed for AP under realistic scenarios, targeting a specific real-life cohort of T1D patients.

The paper is structured as follows: Section 2 explains the methodology used to extract real-life scenarios from clinical data and describes the modification of the adult population. Section 3 is devoted to the illustration of results, Section 4 presents a discussion of the results, and Section 5 concludes the paper.

2. Methodology

The methodology used for generating the VPs is shown in the schematic presented in Figure 1. It is composed of three steps, as explained below.

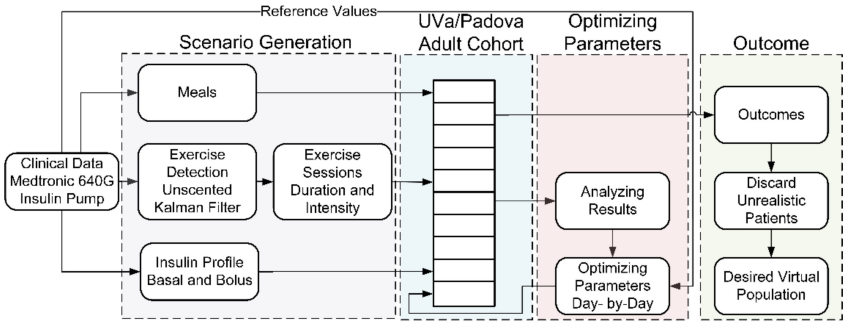

Figure 1. Block diagram showing the steps involved in methodology.

2.1. Generating the Scenario

To generate the scenario, glucose, meal, and insulin data were extracted reviewing the electronic medical records and databases of individuals with T1D followed at the Diabetes Unit, Endocrinology and Nutrition Department at the Hospital Clínic of Barcelona. In the current analysis, patients with T1D using SAP therapy with the 640G Medtronic-Minimed system (Medtronic-Minimed, Northridge, CA, USA) linked to a glucometer (Contour Next 2.4®®, Ascensia Diebetes Care, Parsippany, NJ, USA) and a glucose sensor (Enlite®®, Medtronic-Minimed, Northridge, CA, USA) for at least 6 months were included. Demographic and clinical data were recorded from computerized clinical records. Data were collected from uploads from each patient including CGM data using CareLink Pro®® software. The study has been reviewed by the local ethics committee (HCB/2015/0683) and has therefore been performed in accordance with the ethical standards laid out in an

appropriate version of the 1964 Declaration of Helsinki. All subjects gave informed consent. The demographic data of the patients is provided in Table 1.

Table 1. Demographic data of the cohort.

Parameter	Patients (n = 14)		
Gender (Male/Female)	**6/8**		
	Mean (Standard Deviation)	**Maximum**	**Minimum**
Age (years)	41.57 (11.67)	74	30
Weight (kg)	70.07 (17.37)	116	50
Height (centimeters)	168.71 (9.38)	187	157
Time with Diabetes (years)	13.43 (7.31)	29	3
HbA1c	7.04 (0.82)	8.9	5.9

Data for 14 days is selected corresponding to each patient in a cohort with at least 70% of CGM data available during this time period, which has been defined as the required minimum amount of CGM data for the attainment of meaningful results [54]. The amount of carbohydrates and time of ingestion from the clinical data is included as meals into the simulator. Within a timeframe of 14 days, the days with less than 50% of CGM data are also excluded from the scenario and simulations. Using a previously developed algorithm [30], which requires BG and insulin profiles as inputs, disturbances that are not described by other parameters in the UVa/Padova model are detected and included in the simulator in the form of aerobic exercise. For the correct inclusion of the detected exercise sessions, a reference table was generated using the UVa/Padova simulator for a range of intensity values. The value of intensity that matched the BG profile from the clinical data was selected for each detected exercise session. The exercise model [55] considered in this work was previously fit by our group using clinical data [31]. The basal and bolus insulin values from the pump are implemented in the simulator. A single immediate dose of bolus insulin was used 99% of the time for insulin administration. Details of the scenarios extracted from the clinical data of 14 T1D patients are presented in Table 2.

Table 2. Details of the realistic scenarios extracted from the clinical data.

Scenario	Duration (Days)	Basal Insulin per Day (U)	Bolus Insulin per Day (U)	Total Insulin per Day (U)	Number of Meals per Day	Amount of Carbohydrates per Day (g)	Estimated Exercise Sessions per Day	CGM Active (%)
1	14	35.116	19.516	54.633	2.928	111.071	2	97.40
2	12	22.762	29.250	52.012	3.785	163.214	2	79.44
3	11	20.328	19.091	39.418	5.142	99.642	1	73.93
4	13	17.461	18.308	35.769	3.000	86.786	1	92.09
5	11	14.323	26.718	41.041	4.857	186.428	2	75.72
6	10	16.502	26.970	43.472	5.214	148.214	0	71.16
7	12	13.323	30.092	43.414	4.357	136.428	1	77.75
8	12	18.108	11.250	29.358	4.286	36.786	1	74.75
9	14	9.295	10.239	19.534	5.642	103.214	1	95.83
10	09	34.085	16.400	50.485	6.428	68.571	0	70.36
11	12	20.137	12.446	32.582	4.428	74.286	0	73.74
12	14	33.261	33.657	66.918	3.428	182.143	2	92.31
13	12	8.065	15.125	23.190	3.142	174.214	1	79.32
14	14	17.303	19.639	36.942	7.070	90.375	0	87.18

All values given are in median except percentage CGM was active.

2.2. Parameter Optimization

A total of 14 scenarios (corresponding to 14 patients) were extracted from the clinical data. Each scenario was simulated using the UVa/Padova's 10 adult patients. The adult population was modified to replicate the BG outcomes indicated by the clinical data.

The parameters that describe insulin sensitivity and baseline endogenous glucose production were optimized. In the case of the UVa/Padova VPs, the parameters V_{mx} and k_{p1} are considered for the purpose of modification. The model equations containing the parameters are provided below:

$$U_{id}(t) = \frac{[V_{mo} + V_{mx}.X(t)].G_t(t)}{K_{mo} + G_t(t)}, \quad (1)$$

$$EPG(t) = k_{p1} - k_{p2}.\, G_p(t) - k_{p3}.X^L(t), \quad (2)$$

where $U_{id}(t)$ represents the insulin-dependent utilization of glucose in the remote compartment, $X(t)$ is the insulin action, and the parameter V_{mx} (mg/kg/min per pmol/L) used for adjusting the BG profile is the sensitivity of insulin on glucose utilization. $EPG(t)$ describes endogenous glucose production and k_{p1}(mg/kg/min) is proportional to the basal endogenous glucose production. To replicate the BG values found in the clinical data of real patients in which insulin action varies and appears different on different days, the parameters were adjusted accordingly.

Firstly, the V_{mx} and k_{p1} parameters were optimized in a day-by-day fashion to mimic the median BG value from clinical data subjected to the constraints of maximum and minimum BG limits. Next, the adjusted day-by-day parameters were smoothed using a transition period of 4 h applied at 10 pm. The p-values were calculated using the Wilcoxon signed rank test.

The flow chart of the algorithm is presented in Figure 2, which is adopted for the optimization of parameters corresponding to each day. The description of parameters used in the flow chart can be seen in Table 3. The parameters are incremented or decremented in order to minimize the error (see Table 3). The parameters are subjected to maximum and minimum limits constraints. A threshold of 0.5 mg/dL is considered for the acceptable error in median BG. Once this threshold is met, the values of the parameters are selected for that particular day.

The flow chart in Figure 2 depicts the iterations involved in the numerical simulations required for the optimization of the parameters. The chart reflects the process of optimizing the parameters for a single particular day. The primary goal of the optimization is to find a solution that will restrict the error below the threshold. These parameter changes result in BG outcomes similar to the clinical data.

Table 3. Description of parameters used in flow chart.

Parameter	Description
Error	Reference Median CGM − Current Median CGM
Th	Threshold (0.5 mg/dL)
Upper Limit	480 mg/dL
Lower Limit	50 mg/dL
k_{p1}_Max	15 mg/kg/min
k_{p1}_Min	0.01 mg/kg/min
V_{mx}_Max	3 mg/kg/min per pmol/L
V_{mx}_Min	0.001 mg/kg/min per pmol/L

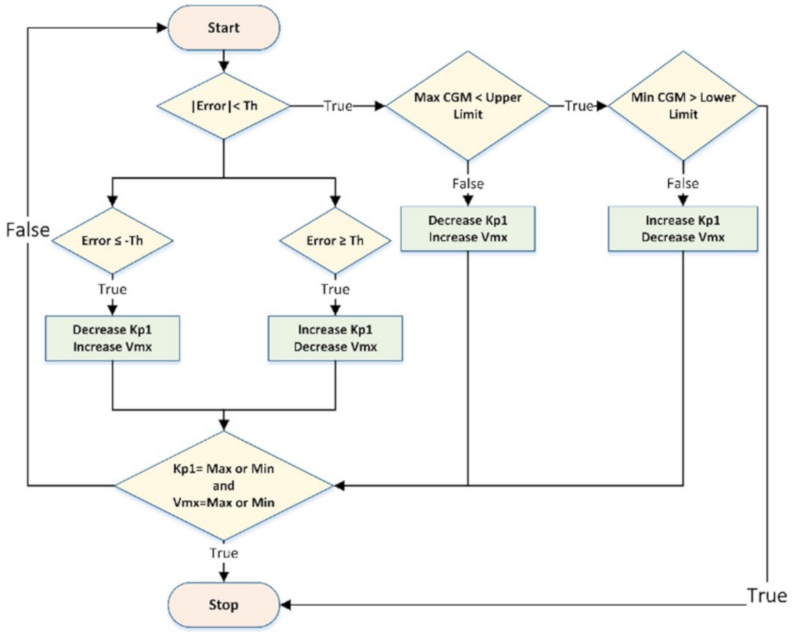

Figure 2. Optimization algorithm flow chart.

2.3. Discarding Unrealistic Patients

After optimizing the parameters, patients were discarded based on two criteria. Firstly, VP with maximum BG value above 500 mg/dL and/or the minimum BG value below 30 mg/dL were discarded. Secondly, patients with two times the standard deviation of saturation points as compared to the clinical data were discarded from the final cohort of VPs. Saturation points represent the end range of a CGM and are defined as BG values equal or greater than 400 mg/dL or equal or less than 40 mg/dL.

3. Results

Overall results are presented in Table 4. A cohort of 75 VPs (54%) was generated as compared to the maximum possible number of 140 VPs. A total of 65 VPs were discarded based on maximum and minimum BG limits (31%) and saturation points (15%), resulting in a cumulative total of 46%.

3.1. Blood Glucose Outcomes

The results are provided in median and interquartile range (IQ). The performance indicators of BG profile presented to draw a comparison between clinical data and simulation results can be divided into four categories. Firstly, the absolute BG values, which include the mean, median, maximum, and minimum values corresponding to the entire duration (Table 2) of the scenario. Secondly, the CV (indicator of the glycemic variability) and glucose management index (GMI), which is an indicator for average glycemic exposure. Thirdly, the percentage of time BG values lie in various ranges. Finally, the percentage of saturation points are reported.

Mathematics **2021**, *9*, 1200

Table 4. A comparison of clinical data and simulation results.

Parameter	Clinical Data	Simulation Results with Optimization	*p*-Value
Mean CGM (mg/dL)	162.2 (145.6–169.3)	166.3 (155.3–175.3)	0.194
Median CGM (mg/dL)	156.5 (135–165)	162.2 (146.8–171.9)	0.104
Maximum CGM (mg/dL)	345 (282–400)	322 (303.9–361.9)	0.715
Minimum CGM (mg/dL)	48.5 (41–52)	45.4 (41.6–49.4)	0.463
CV (Percentage)	33 (28.8–38.1)	32 (26.8–35.5)	0.542
GMI (Percentage)	7.2 (6.8–7.4)	7.3 (7–7.5)	0.194
% of time CGM			
Below 54 mg/dL	0.11 (0.031–0.636)	0.68 (0.221–1.116)	0.502
54 to 69 mg/dL	1.69 (0.779–3.39)	1.51 (0.521–3.212)	0.670
70 to 140 mg/dL	36.43 (30.682–48.742)	30.69 (23.512–39.323)	0.011
70 to 180 mg/dL	66.85 (57.402–71.563)	59.64 (56.313–70.362)	0.358
181 to 250 mg/dL	24.86 (20.649–30.788)	27.49 (22.960–31.250)	0.153
Above 250 mg/dL	4.27 (2.333–9.845)	5.44 (2.691–10.985)	0.426
Saturation Points 40 mg/dL (%)	0 (0–0)	0 (0–0.043)	0.688
Saturation Points 400 mg/dL (%)	0 (0–0.032)	0 (0–0)	0.438

3.2. Inter-Subject Variability

The original inter-patient variability provided by the UVa/Padova cohort is retained for the newly generated VPs. To demonstrate this, a scenario of three meals per day for a duration of 14 days, which is the maximum length of simulation considered in this study, was used to analyze the inter-subject variability. The breakfast (45 g), lunch (70 g), and dinner (60 g) were delivered at 7:00, 13:00, and 20:00, respectively. Open loop insulin therapy was used for simulations with adjusted basal rates for the generated VPs. The results for real scenario 9 (see Table 2) are presented in Figure 3. The simulation results of all real scenarios are provided as Supplementary Materials (Figures S1–S11). The results for real scenarios 5 and 7 are not included because they were composed of only 1 acceptable VP.

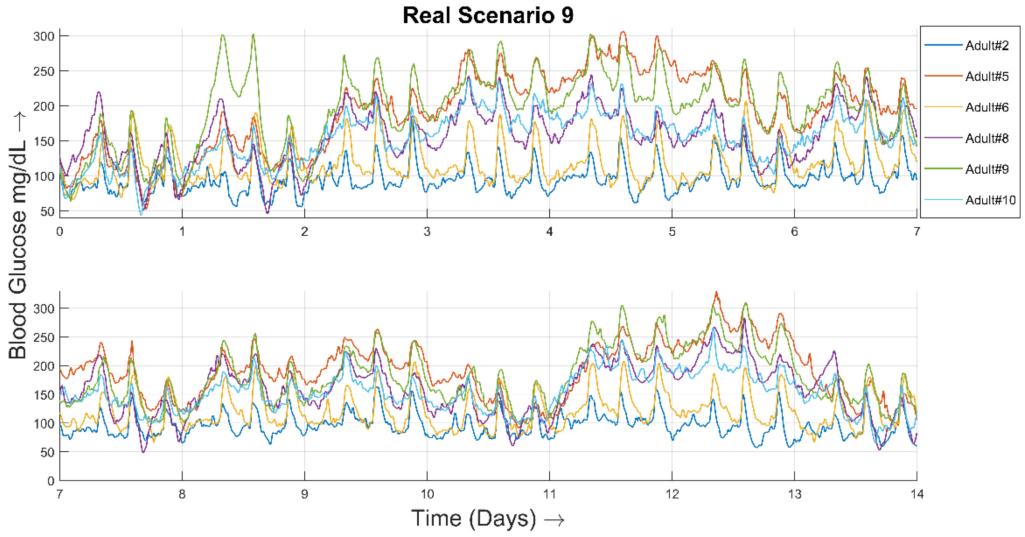

Figure 3. Representative inter-patient variability simulation from 6 of the newly generated VPs. These VPs were based on the real scenario 9 and their parameters were tuned based on the clinical data.

The overall BG curve corresponding to all 75 generated VPs is shown in Figure 4. The BG curve is calculated as the mean ± standard deviation BG value of all VPs. The duration of the simulation study was 14 days, with three meals a day.

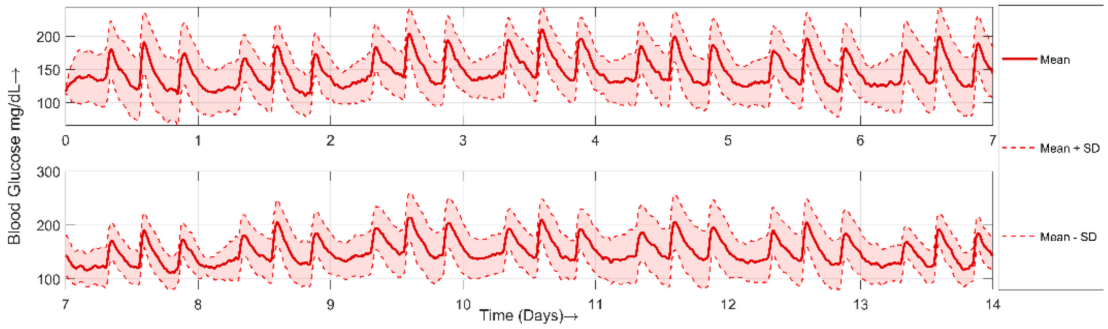

Figure 4. A plot showing mean (red solid line), mean + standard deviation (upper red dashed line) and mean − standard deviation (lower red dashed line) BG curve of all 75 generated VPs.

3.3. Optimized Parameters

The distinct realistic scenarios considered are reflected in a set of optimized parameters which are significantly different. To illustrate this, box plots of the parameters (V_{mx} and k_{p1}) are presented in Figures 5 and 6, respectively. The values of all parameters for 75 VPs are provided in the Supplementary Materials.

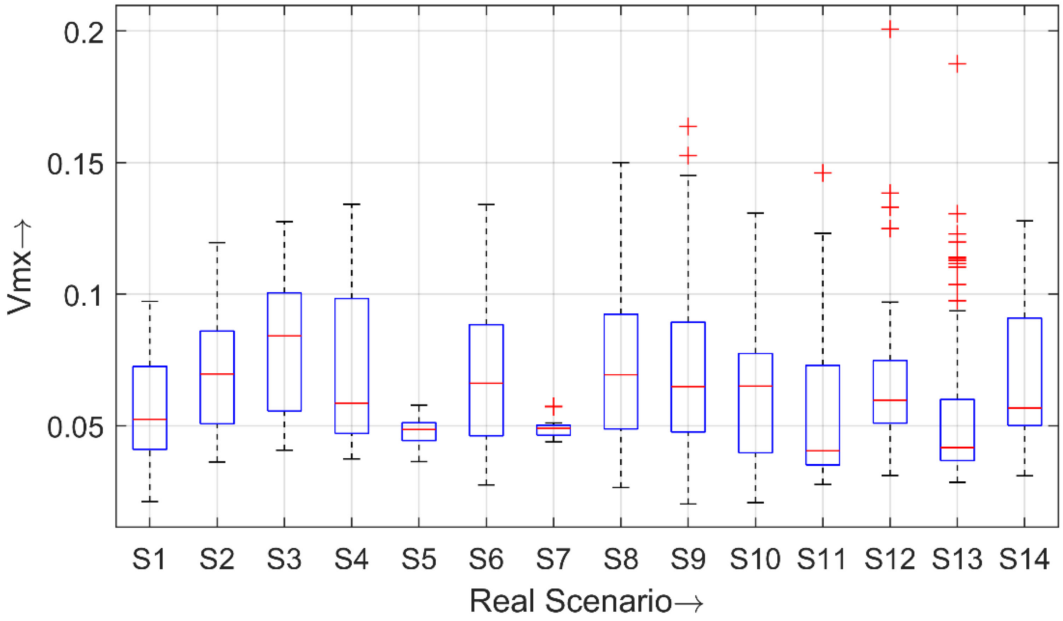

Figure 5. Boxplot of the parameter V_{mx} for all 14 real scenarios.

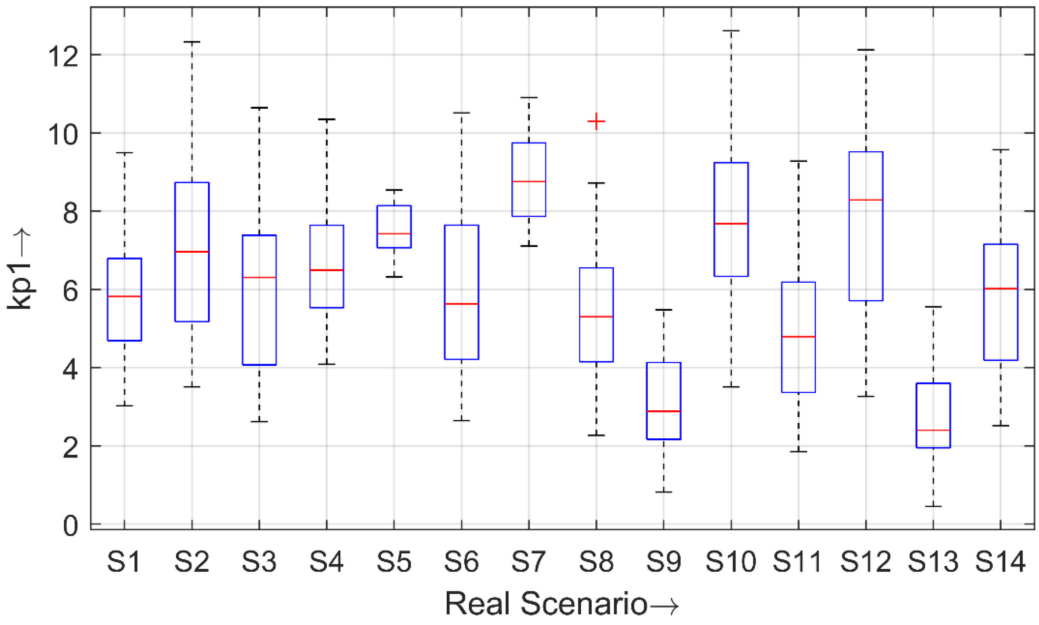

Figure 6. Boxplot of the parameter k_{p1} for all 14 real scenarios.

3.4. Mapping Scenarios against VPs

The details of valid generated VPs and discarded VPs are presented in Table 5. The real scenarios are corresponding to clinical data (see Table 2). The adult patients are the modified VPs from the UVa/Padova's adult cohort.

Table 5. Mapping of real scenarios into valid VPs.

Real Scenario	Adult 1	Adult 2	Adult 3	Adult 4	Adult 5	Adult 6	Adult 7	Adult 8	Adult 9	Adult 10
1	✓	✗	✓	✗	✓	✓	✗	✓	✗	✓
2	✗	✓	✓	✗	✗	✓	✗	✗	✗	✓
3	✓	✓	✗	✓	✗	✓	✗	✓	✗	✓
4	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
5	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
6	✗	✓	✗	✓	✓	✓	✗	✓	✗	✓
7	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
8	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
9	✗	✓	✗	✗	✓	✓	✗	✓	✓	✓
10	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓
11	✗	✗	✗	✓	✓	✓	✗	✓	✗	✓
12	✓	✓	✗	✓	✗	✓	✗	✓	✗	✓
13	✓	✗	✗	✗	✓	✓	✗	✓	✓	✓
14	✗	✓	✗	✓	✓	✓	✗	✓	✓	✓

The symbols tick and cross represent a valid and discarded VP, respectively.

4. Discussion

Several methods have already been proposed in the literature for the generation of VPs, mainly focused on generating large cohorts of patients with T1D. In this work, the proposed methodology is focused on replicating a specific cohort of T1D patients. The treatment strategies developed for existing cohorts of VPs are prone to over- or under-

perform for such sub-cohorts of patients with T1D. Therefore, the proposed methodology will provide a benchmark for in silico experimentation to develop T1D treatment strategies for sub-cohorts, allowing tight BG control to be achieved in these patient populations. The application of the proposed methodology can be extended to generate VPs replicating various sub-cohorts of patients with T1D by optimizing the parameters according to the clinical data of the targeted sub-cohort. Another contribution of this work is to provide real-life scenarios (meals, exercise, and glycemic variability) for testing and validating the treatment strategies developed for patients with T1D.

Utilizing the scenario attributes and insulin profiles from the clinical data, the adult population of the UVa/Padova simulator shows significant deviation in results as compared to the clinical data. This indicates that the adult population of the UVa/Padova simulator does not represent the cohort considered in this study. The effectiveness of the proposed methodology is demonstrated by achieving outcomes similar to the considered cohort of patients with T1D. Therefore, in this work, a cohort comprised of 75 VPs was generated to reflect the BG metrics of a cohort from the Hospital Clinic de Barcelona.

To demonstrate the contribution of meals in CV and BG outcomes, a comparison of simulation results with real meal scenario (RMSc), typical meal scenario (TMSc), and clinical data is presented in Table 6. RMSc is composed of the meals pattern extracted from the clinical data, whereas TMSc is composed of four meal scenarios (three with 3 meals per day and one with additional snacks per day) [38,39,56,57]. The results for TMSc are presented as median of simulation results for all individual four-meal scenarios. Open loop control and UVa/Padova's adult cohort are used for simulations. The mean BG (129.2 mg/dL, 131.8 mg/dL) versus 166.3 mg/dL, CV (19.6%, 25.2%) versus 33%, and PoT in range 70 to 180 mg/dL (95.3%, 90.15%) versus 66.85% are reported for RMSc and TMSc, respectively. The CV in case of TMSc is 5.6% greater compared to the RMSc. The clinical data shows that meals consumed by real patients (considered in this study) are smaller than those used in traditional in silico simulations. Therefore, meals only account for a small portion of CV. The BG outcomes for RMSc and TMSc are somehow close to each other but significantly different from the clinical data. It implies that including only meals in the simulation scenario is not enough to achieve realistic BG outcomes. Therefore, in this study, insulin pump data was added to the simulator and a methodology was proposed to adjust the model parameters to replicate clinical BG outcomes. The day-to-day optimization of parameters and the smoothing of daily parameter transitions cumulatively allow for the achievement of the glycemic variability of real patients.

The results presented in Table 4 reflect the BG profile of a cohort with T1D. The primary goal was to generate a virtual cohort of patients with T1D to mimic real patients. The performance indicators considered to compare the BG profile of the cohorts were statistically similar ($p > 0.05$), except the PoT BG values lie in a range of 70 to 140 mg/dL. The mean and median BG values reported were very close, but a rise of about 10 mg/dL was observed in the IQ range in case of the simulation results. The maximum BG value reported was 23 mg/dL lower in simulation results. However, the VPs generated showed significantly close results as compared to the clinical data ($p = 0.71$). The minimum BG value reported was 3 mg/dL lower as compared to the clinical data and the IQ range reported was almost identical. CV reported was lower by 1% and the IQ range differed by about 2%. The GMI reported was almost the same as indicated by the clinical data.

Table 6. A comparison of clinical data and simulation results using UVa/Padova VPs in open loop control. Typical meal scenario results are the ones obtained using typical three meal or three meal plus snack scenarios as published in the literature. Real meal scenarios include the same meals as in clinical data.

Parameter	Clinical Data	Typical Meal Scenario	Real Meal Scenario
Mean CGM (mg/dL)	162.2 (145.6–169.3)	131.8 (126.4–140.8)	129.2 (126.6–134.9)
Median CGM (mg/dL)	156.5 (135–165)	127.2 (120.8–134.3)	125.1 (124.3–128.3)
Maximum CGM (mg/dL)	345 (282–400)	249.6 (221.5–266.9)	229.2 (205.6–273.9)
Minimum CGM (mg/dL)	48.5 (41–52)	57.0 (53.2–62.0)	63.5 (60.1–66.2)
CV (Percentage)	33 (28.8–38.1)	25.2 (22.6–27.0)	19.6 (17.3–25.8)
GMI (Percentage)	7.2 (6.8–7.4)	6.4 (6.3–6.7)	6.4 (6.3–6.5)
% of time CGM			
Below 54 mg/dL	0.11 (0.031–0.636)	0 (0.00–0.05)	0 (0.00 – 0.00)
54 to 69 mg/dL	1.69 (0.779–3.39)	1.07 (0.35–1.41)	0.3 (0.10–0.48)
70 to 140 mg/dL	36.43 (30.682–48.742)	64.28 (54.82–70.61)	71.7 (62.68–77.12)
70 to 180 mg/dL	66.85 (57.402–71.563)	90.15 (82.60–94.15)	95.3 (88.75–98.35)
181 to 250 mg/dL	24.86 (20.649–30.788)	7.97 (4.41–13.36)	4.5 (1.22–8.58)
Above 250 mg/dL	4.27 (2.333–9.845)	0.025 (0.00–0.55)	0.0 (0.00–0.75)

The performance indicators related to PoT of the BG values in a specific range reflected quite close behavior as compared to the clinical data, except in the range 70 to 140 mg/dL. The PoT in the very low range (<54 mg/dL) observed was 0.58% greater. However, the PoT in low range of 54 to 69 mg/dL appeared to replicate the clinical data. The PoT reported in a range of 70 to 140 mg/dL showed a decline of about 6%. The PoT in normoglycemia (70 to 180 mg/dL), which is the target range to achieve, showed a decrease of 6% in simulation results with an identical IQ range. The PoT reported in ranges 181 to 250 mg/dL and >250 mg/dL almost replicated the clinical data.

Moreover, the saturation points were also considered for the selection of VPs to prevent the use of the CGM limits (40–400 mg/dL) for parameter fitting. The results are presented in terms of percent of saturation points in the entire duration of the scenario. The durations for all considered scenarios are presented in Table 2. The percent of saturation points (40 and 400 mg/dL) reported are similar as compared to the clinical data. The saturation points correspond to PoT BG in very low range (<54 mg/dL) and very high range (>250 mg/dL). Therefore, this criterion results in achieving the PoT BG in ranges mentioned similar to the clinical data.

The detailed mapping of valid and discarded VPs is provided in Table 5.

As we expected, no original VP can be adjusted for all scenarios nor is any scenario likely to be adjusted for all patients. In fact, UVa/Padova adult 7 cannot be properly adjusted for any scenario and is discarded in all cases. Modification of the parameters for this patient resulted in out-of-range glycaemia values (>500 mg/dL) in all 14 cases. For scenario 5, only adult 10 resulted in a valid VP. Three of the ten VPs were discarded because they violated the saturation point criteria. The remaining VPs were discarded because blood glucose was out of range (>500 or <30 mg/dL). For scenario 7, only adult 2 is a valid VP. The other 9 VPs were discarded because BG values were out of range.

There may exist mismatches between real and detected exercise sessions, since proper detection may require accepting a certain false positive rate to obtain a high true positive rate. Despite this, the goal of this work is to demonstrate that the proposed algorithm can be used to cope with all possible scenario elements existing in real life, which can be integrated into the simulator. Clinical data that includes exercise details will allow that information to be used directly without need for the detection of exercise sessions, resulting in even more accurate scenario development.

The parameter V_{mx} has been extensively used by the Padova group to explain intraday variability of insulin sensitivity, and intra- and inter-patient variability [53,58], and is therefore a parameter that can be varied both between patients and within the same patient, as applied in this work. To account for inter-day variability in basal and postprandial endogenous glucose production, the parameter k_{p1} in UVa/Padova VPs is considered for modification in the methodology presented in this study.

The main limitation of this study is in forcing the 10 adult UVa/Padova VPs to replicate the BG metrics of the same real subject. This led to the rejection of about 46% generated VPs. The VPs in the UVa/Padova simulator that cannot be forced to match the outcomes of a real patient were discarded through the process explained in the methodology section. However, inter-patient variability is still retained, as it is reflected by the distinct set of parameters defining the VP in the UVa/Padova simulator. In case of the same UVa/Padova VP corresponding to different real subjects, this variability is retained in the parameters considered for modification in this study, which change day-by-day. The other limitation is that it is not possible to exactly replicate the BG variability of real patients due to the causes of variability (stress, illness, lifestyle, etc.), which are difficult to model because of their unpredictable nature. However, the goal is to minimize the gap between simulation environment and reality in terms of BG outcomes.

5. Conclusions

In this work, a novel algorithm to generate a virtual cohort of T1D patients was presented. The novelty of the proposed scheme lies in the optimal daily adjustment of the parameters to achieve the glycemic outcomes reflected by clinical data. The daily variation of parameters represents realistic daily changes in real patients with T1D that influence BG curves. The clinical data was exclusively taken into account to modify the parameters, resulting in more realistic BG outcomes in terms of generated VPs.

The algorithm is based on optimizing the parameters of virtual adult patients from the UVa/Padova simulator to replicate the BG profile of a targeted cohort of real patients. The targeted BG profile is replicated in a day-by-day manner by optimizing the parameters accordingly. A virtual cohort of 75 patients has been generated for a cohort of 14 patients with T1D from the Hospital Clinic de Barcelona. The cohort of VPs generated potentially represents the cohort from the Hospital Clinic de Barcelona in terms of BG performance indicators. The statistical similarity index in terms of p-values (Wilcoxon signed rank test) was presented to validate the effectiveness of the proposed algorithm.

This algorithm can be used to test the controllers and therapeutic strategies developed for the treatment of T1D patients. It provides testing under the realistic scenarios based on the clinical data and a challenging variable behavior of patients as the parameters are changing day-by-day. Moreover, it is based on the FDA-approved UVa/Padova simulator and can be utilized for the preclinical validation.

The presented study can be extended in two possible directions. Firstly, a greater number of parameters for the purpose of modification and optimization can be explored. This may better capture the inter-subject variability and is expected to increase the number of VPs generated. Secondly, the parameters can be adjusted hourly instead of daily to achieve the intraday variability in the BG curve, as shown by the real patients.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/math9111200/s1, Figure S1: Simulation results for real scenario 1, Figure S2: Simulation results for real scenario 2, Figure S3: Simulation results for real scenario 3, Figure S4: Simulation results for real scenario 4, Figure S5: Simulation results for real scenario 6, Figure S6: Simulation results for real scenario 8, Figure S7: Simulation results for real scenario 10, Figure S8: Simulation results for real scenario 11, Figure S9: Simulation results for real scenario 12, Figure S10: Simulation results for real scenario 13, Figure S11: Simulation results for real scenario 14. Table S1: Values of optimized parameters corresponding to real patient 1, Table S2: Values of optimized parameters corresponding to real patient 2, Table S3: Values of optimized parameters corresponding to real patient 3, Table S4: Values of optimized parameters corresponding to real patient 4, Table S5: Values of optimized

parameters corresponding to real patient 5, Table S6: Values of optimized parameters corresponding to real patient 6, Table S7: Values of optimized parameters corresponding to real patient 7, Table S8A: Values of optimized parameters corresponding to real patient 8, Table S8B: Values of optimized parameters corresponding to real patient 8, Table S9: Values of optimized parameters corresponding to real patient 9, Table S10A: Values of optimized parameters corresponding to real patient 10, Table S10B: Values of optimized parameters corresponding to real patient 10, Table S11: Values of optimized parameters corresponding to real patient 11, Table S12: Values of optimized parameters corresponding to real patient 12, Table S13: Values of optimized parameters corresponding to real patient 13, Table S14: Values of optimized parameters corresponding to real patient 14.

Author Contributions: S.A. conceived, designed, implemented, and validated the methodology, performed the statistical analysis and wrote the manuscript; C.M.R. and A.B. contributed to the development of the methodology and critically revised the manuscript; I.C. and M.G. provided the clinical data and critically revised the manuscript; J.V. supervised all the simulation study, analyzed, interpreted data and critically revised the manuscript. All authors contributed to the review of the report and approved the final version for submission. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Spanish Ministry of Science and Innovation through grant PID2019-107722RB-C22/AEI/10.13039/501100011033 and the Government of Catalonia under 2017SGR1551 and 2019 FI_B 01200.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Hospital Clínic de Barcelona (protocol code HCB/2015/0683 and date of approval 22 September 2015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from the Spanish Consortium on Artificial Pancreas and Diabetes Technology and are available on request from the corresponding author with the permission of the Spanish Consortium on Artificial Pancreas and Diabetes Technology.

Conflicts of Interest: All authors have no conflict of interest related to this manuscript.

References

1. Katsarou, A.; Gudbjörnsdottir, S.; Rawshani, A.; Dabelea, D.; Bonifacio, E.; Anderson, B.J.; Jacobsen, L.M.; Schatz, D.A.; Lernmark, Å. Type 1 Diabetes Mellitus. *Nat. Rev. Dis. Prim.* **2017**, *3*, 1–17. [CrossRef]
2. Hernando, M.E.; García-Sáez, G.; Gómez, E.J.; Pérez-Gandía, C.; Rodríguez-Herrero, A. Automated Insulin Delivery: The Artificial Pancreas Technical Challenges. *Am. J. Ther.* **2020**, *27*, e62–e70. [CrossRef]
3. Cobelli, C.; Dalla Man, C.; Sparacino, G.; Magni, L.; De Nicolao, G.; Kovatchev, B.P. Diabetes: Models, Signals, and Control. *IEEE Rev. Biomed. Eng.* **2009**, *2*, 54–96. [CrossRef] [PubMed]
4. Resalat, N.; El Youssef, J.; Tyler, N.; Castle, J.; Jacobs, P.G. A Statistical Virtual Patient Population for the Glucoregulatory System in Type 1 Diabetes with Integrated Exercise Model. *PLoS ONE* **2019**, *14*, e0217301. [CrossRef] [PubMed]
5. Wilinska, M.E.; Chassin, L.J.; Acerini, C.L.; Allen, J.M.; Dunger, D.B.; Hovorka, R. Simulation Environment to Evaluate Closed-Loop Insulin Delivery Systems in Type 1 Diabetes. *J. Diabetes Sci. Technol.* **2010**, *4*, 132–144. [CrossRef] [PubMed]
6. Ciorap, R.; Luca, C.; Andriţoi, D. Realistic patient simulators for education in medicine and bioengineering. In Proceedings of the The International Scientific Conference eLearning and Software for Education, Bucharest, Romania, 19–20 April 2018; Volume 3, pp. 389–394.
7. Soltész, K. *On Automation in Anesthesia*; Department of Automatic Control, Lund University: Lund, Sweden, 2013.
8. Zurakowski, R.; Teel, A.R. A Model Predictive Control Based Scheduling Method for HIV Therapy. *J. Theor. Biol.* **2006**, *238*, 368–382. [CrossRef] [PubMed]
9. Halvorsen, F.H.; Elle, O.J.; Fosse, E. Simulators in Surgery. *Minim. Invasive Ther. Allied Technol.* **2005**, *14*, 214–223. [CrossRef] [PubMed]
10. Konduri, P.R.; Marquering, H.A.; van Bavel, E.E.; Hoekstra, A.; Majoie, C.B.; INSIST Investigators. In-Silico Trials for Treatment of Acute Ischemic Stroke. *Front. Neurol.* **2020**, *11*, 1062. [CrossRef]
11. Rudy, Y. In Silico Pipeline for Drug Cardiotoxicity Assessment. *Circ. Res.* **2020**, *126*, 965–967. [CrossRef]
12. Kazmi, S.R.; Jun, R.; Yu, M.-S.; Jung, C.; Na, D. In Silico Approaches and Tools for the Prediction of Drug Metabolism and Fate: A Review. *Comput. Biol. Med.* **2019**, *106*, 54–64. [CrossRef] [PubMed]
13. Lee, C.H.; Koohy, H. In Silico Identification of Vaccine Targets for 2019-NCoV. *F1000Research* **2020**, *9*, 145. [CrossRef]
14. Lateef, F. Simulation-Based Learning: Just like the Real Thing. *J. Emergencies Trauma Shock* **2010**, *3*, 348. [CrossRef] [PubMed]

15. Kovatchev, B.P.; Breton, M.; Dalla Man, C.; Cobelli, C. In Silico Preclinical Trials: A Proof of Concept in Closed-Loop Control of Type 1 Diabetes. *J. Diabetes Sci. Technol.* **2009**. [CrossRef]
16. Soru, P.; De Nicolao, G.; Toffanin, C.; Dalla Man, C.; Cobelli, C.; Magni, L.; AP@ Home Consortium. MPC Based Artificial Pancreas: Strategies for Individualization and Meal Compensation. *Annu. Rev. Control* **2012**, *36*, 118–128. [CrossRef]
17. Percival, M.W.; Wang, Y.; Grosman, B.; Dassau, E.; Zisser, H.; Jovanovič, L.; Doyle, F.J., III. Development of a Multi-Parametric Model Predictive Control Algorithm for Insulin Delivery in Type 1 Diabetes Mellitus Using Clinical Parameters. *J. Process. Control* **2011**, *21*, 391–404. [CrossRef]
18. Colmegna, P.; Pena, R.S.S.; Gondhalekar, R.; Dassau, E.; Doyle, F.J., III. Reducing Risks in Type 1 Diabetes Using H_∞ Control. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2939–2947. [CrossRef]
19. Colmegna, P.; Sánchez-Peña, R.S.; Gondhalekar, R. Linear Parameter-Varying Model to Design Control Laws for an Artificial Pancreas. *Biomed. Signal Process. Control* **2018**, *40*, 204–213. [CrossRef]
20. Colmegna, P.H.; Sanchez-Pena, R.S.; Gondhalekar, R.; Dassau, E.; Doyle, F.J. Switched LPV Glucose Control in Type 1 Diabetes. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 1192–1200. [CrossRef]
21. MohammadRidha, T.; Aït-Ahmed, M.; Chaillous, L.; Krempf, M.; Guilhem, I.; Poirier, J.Y.; Moog, C.H. Model Free IPID Control for Glycemia Regulation of Type-1 Diabetes. *IEEE Trans. Biomed. Eng.* **2017**, *65*, 199–206. [CrossRef]
22. Herrero, P.; Bondia, J.; Adewuyi, O.; Pesl, P.; El-Sharkawy, M.; Reddy, M.; Toumazou, C.; Oliver, N.; Georgiou, P. Enhancing Automatic Closed-Loop Glucose Control in Type 1 Diabetes with an Adaptive Meal Bolus Calculator—In Silico Evaluation under Intra-Day Variability. *Comput. Methods Programs Biomed.* **2017**, *146*, 125–131. [CrossRef]
23. Vettoretti, M.; Facchinetti, A.; Sparacino, G.; Cobelli, C. Predicting Insulin Treatment Scenarios with the Net Effect Method: Domain of Validity. *Diabetes Technol. Ther.* **2016**, *18*, 694–704. [CrossRef] [PubMed]
24. Toffanin, C.; Messori, M.; Cobelli, C.; Magni, L. Automatic Adaptation of Basal Therapy for Type 1 Diabetic Patients: A Run-to-Run Approach. *Biomed. Signal Process. Control* **2017**, *31*, 539–549. [CrossRef]
25. Torrent-Fontbona, F. Adaptive Basal Insulin Recommender System Based on Kalman Filter for Type 1 Diabetes. *Expert Syst. Appl.* **2018**, *101*, 1–7. [CrossRef]
26. Campos-Náñez, E.; Layne, J.E.; Zisser, H.C. In Silico Modeling of Minimal Effective Insulin Doses Using the UVa/Padova Type 1 Diabetes Simulator. *J. Diabetes Sci. Technol.* **2018**, *12*, 376–380. [CrossRef] [PubMed]
27. Turksoy, K.; Samadi, S.; Feng, J.; Littlejohn, E.; Quinn, L.; Cinar, A. Meal Detection in Patients with Type 1 Diabetes: A New Module for the Multivariable Adaptive Artificial Pancreas Control System. *IEEE J. Biomed. Health Inform.* **2015**, *20*, 47–54. [CrossRef] [PubMed]
28. Xie, J.; Wang, Q. Meal Detection and Meal Size Estimation for Type 1 Diabetes Treatment: A Variable State Dimension Approach. *Dyn. Syst. Control Conf.* **2015**, *57243*, V001T15A003.
29. Samadi, S.; Turksoy, K.; Hajizadeh, I.; Feng, J.; Sevil, M.; Cinar, A. Meal Detection and Carbohydrate Estimation Using Continuous Glucose Sensor Data. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 619–627. [CrossRef]
30. Ramkissoon, C.M.; Bertachi, A.; Beneyto, A.; Bondia, J.; Vehi, J. Detection and Control of Unannounced Exercise in the Artificial Pancreas without Additional Physiological Signals. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 259–267. [CrossRef] [PubMed]
31. Bertachi, A.; Beneyto, A.; Ramkissoon, C.M.; Vehi, J. Assessment of Mitigation Methods to Reduce the Risk of Hypoglycemia for Announced Exercise in a Uni-Hormonal Artificial Pancreas. *Diabetes Technol. Ther.* **2018**, *20*, 285–295. [CrossRef] [PubMed]
32. Garcia-Tirado, J.; Colmegna, P.; Corbett, J.; Ozaslan, B.; Breton, M.D. Ensemble model predictive control strategies can reduce exercise hypoglycemia in type 1 diabetes: In silico studies. In Proceedings of the 2019 American Control Conference (ACC), Philadelphia, PA, USA, 10–12 July 2019; pp. 4752–4758.
33. Aiello, E.M.; Wu, Z.; Christofides, P.D.; Toffanin, C.; Cobelli, C.; Magni, L. Improving Diabetes Conventional Therapy via Machine Learning Modeling. In Proceedings of the 2019 American Control Conference (ACC), Philadelphia, PA, USA, 10–12 July 2019; pp. 4136–4143.
34. Zhu, T.; Li, K.; Georgiou, P. A Dual-Hormone Closed-Loop Delivery System for Type 1 Diabetes Using Deep Reinforcement Learning. *arXiv* **2019**, arXiv:1910.04059.
35. Cappon, G.; Vettoretti, M.; Facchinetti, A.; Sparacino, G. Smart Bolus Calculator for Personalized Insulin Dosing Using Continuous Glucose Monitoring Data and Patient Characteristics. *Diabetes Technol. Ther.* **2019**, *21*, A16–A17.
36. Campos-Delgado, D.U.; Hernández-Ordoñez, M.; Femat, R.; Gordillo-Moscoso, A. Fuzzy-Based Controller for Glucose Regulation in Type-1 Diabetic Patients by Subcutaneous Route. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2201–2210. [CrossRef]
37. Magni, L.; Raimondo, D.M.; Dalla Man, C.; De Nicolao, G.; Kovatchev, B.; Cobelli, C. Model Predictive Control of Glucose Concentration in Type I Diabetic Patients: An in Silico Trial. *Biomed. Signal Process. Control* **2009**, *4*, 338–346. [CrossRef]
38. Incremona, G.P.; Messori, M.; Toffanin, C.; Cobelli, C.; Magni, L. Model Predictive Control with Integral Action for Artificial Pancreas. *Control Eng. Pract.* **2018**, *77*, 86–94. [CrossRef]
39. Bertachi, A.; Ramkissoon, C.M.; Beneyto, A.; Vehí, J. Exercise-Induced Hypoglycemia in Type 1 Diabetes: In-Silico Comparison between Announced and Unannounced Strategies in Closed-Loop Control. *IFAC PapersOnLine* **2019**, *52*, 1000–1005. [CrossRef]
40. Ahmad, S.; Ahmed, N.; Ilyas, M.; Khan, W. Super Twisting Sliding Mode Control Algorithm for Developing Artificial Pancreas in Type 1 Diabetes Patients. *Biomed. Signal Process. Control* **2017**, *38*, 200–211. [CrossRef]
41. Nath, A.; Deb, D.; Dey, R.; Das, S. Blood Glucose Regulation in Type 1 Diabetic Patients: An Adaptive Parametric Compensation Control-Based Approach. *IET Syst. Biol.* **2018**, *12*, 219–225. [CrossRef] [PubMed]

42. Djouima, M.; Azar, A.T.; Drid, S.; Mehdi, D. Higher Order Sliding Mode Control for Blood Glucose Regulation of Type 1 Diabetic Patients. *Int. J. Syst. Dyn. Appl.* **2018**, *7*, 65–84. [CrossRef]
43. Beneyto, A.; Bertachi, A.; Bondia, J.; Vehi, J. A New Blood Glucose Control Scheme for Unannounced Exercise in Type 1 Diabetic Subjects. *IEEE Trans. Control Syst. Technol.* **2018**, *7*, 65–84. [CrossRef]
44. Meneghetti, L.; Terzi, M.; Susto, G.A.; Del Favero, S.; Cobelli, C. Fault detection in artificial pancreas: A model-free approach. In Proceedings of the 2018 IEEE Conference on Decision and Control (CDC), Miami, FL, USA, 17–19 December 2018; pp. 303–308.
45. Paoletti, N.; Liu, K.S.; Chen, H.; Smolka, S.; Lin, S. Data-Driven Robust Control for a Closed-Loop Artificial Pancreas. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *17*, 1981–1993. [CrossRef]
46. Güemes, A.; Herrero, P.; Georgiou, P. A novel glucose controller using insulin sensitivity modulation for management of type 1 diabetes. In Proceedings of the 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 26–29 May 2019; pp. 1–5.
47. Nath, A.; Deb, D.; Dey, R. An Augmented Subcutaneous Type 1 Diabetic Patient Modelling and Design of Adaptive Glucose Control. *J. Process. Control* **2020**, *86*, 94–105. [CrossRef]
48. Hovorka, R.; Canonico, V.; Chassin, L.J.; Haueter, U.; Massi-Benedetti, M.; Federici, M.O.; Pieber, T.R.; Schaller, H.C.; Schaupp, L.; Vering, T.; et al. Nonlinear Model Predictive Control of Glucose Concentration in Subjects with Type 1 Diabetes. *Physiol. Meas.* **2004**, *25*, 905. [CrossRef] [PubMed]
49. Chassin, L.J. In Silico Testing of Glucose Controllers: Methodology and Sample Application. Ph.D. Thesis, City University London, London, UK, 2005.
50. Haidar, A.; Wilinska, M.E.; Graveston, J.A.; Hovorka, R. Stochastic Virtual Population of Subjects with Type 1 Diabetes for the Assessment of Closed-Loop Glucose Controllers. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3524–3533. [CrossRef] [PubMed]
51. Orozco-López, O.; Rodríguez-Herrero, A.; Castañeda, C.E.; García-Sáez, G.; Hernando, M.E. Method to Generate a Large Cohort In-Silico for Type 1 Diabetes. *Comput. Methods Programs Biomed.* **2020**, *193*, 105523. [CrossRef]
52. Man, C.D.; 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. [CrossRef] [PubMed]
53. Visentin, R.; Campos-Náñez, E.; Schiavon, M.; Lv, D.; Vettoretti, M.; Breton, M.; Kovatchev, B.P.; Dalla Man, C.; Cobelli, C. The UVa/Padova Type 1 Diabetes Simulator Goes from Single Meal to Single Day. *J. Diabetes Sci. Technol.* **2018**, *12*, 273–281. [CrossRef] [PubMed]
54. Huffman, J.L.; McNeil, G.; Bismilla, Z.; Lai, A. Essentials of scenario building for simulation-based education. In *Comprehensive Healthcare Simulation: Pediatrics*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 19–29.
55. Dalla Man, C.; Breton, M.D.; Cobelli, C. Physical Activity into the Meal Glucose—Insulin Model of Type 1 Diabetes: In Silico Studies. *J. Diabetes Sci. Technol.* **2009**. [CrossRef] [PubMed]
56. Schiavon, M.; Dalla Man, C.; Cobelli, C. Physiology-based run-to-run adaptation of insulin to carbohydrate ratio improves type 1 diabetes therapy: Results from an in silico study. In Proceedings of the 2019 American Control Conference (ACC), Philadelphia, PA, USA, 10–12 July 2019; pp. 4124–4129.
57. Moscardö, V.; Herrero, P.; Díez, J.L.; Giménez, M.; Rossetti, P.; Bondia, J. In silico evaluation of a parallel control-based coordinated dual-hormone artificial pancreas with insulin on board limitation. In Proceedings of the 2019 American Control Conference (ACC), Philadelphia, PA, USA, 10–12 July 2019; pp. 4759–4764.
58. Visentin, R.; Dalla Man, C.; Kudva, Y.C.; Basu, A.; Cobelli, C. Circadian Variability of Insulin Sensitivity: Physiological Input for in Silico Artificial Pancreas. *Diabetes Technol. Ther.* **2015**, *17*, 1–7. [CrossRef]

2.2 Bolus Insulin Calculation without Meal Information. A Reinforcement Learning Approach

In this publication, we present RL algorithm for the calculation of insulin bolus prior to meal intake in people with T1D. The primary contribution to the field is that the proposed insulin bolus calculator does not require meal information, i.e., CHO content nor patient specific parameters are required. On the other hand, SBCs utilize all this information resulting in CHO misestimation errors as well as putting an extra management burden on patients. The results are presented in terms of standard glycemic metrics and a comparison of the proposed algorithm to the SBC is provided. The candidate's contributions to this publication included developing, designing, and implementing the RL algorithm, as well as validation of the approach in in-silico trials. The candidate also contributed to writing the paper, participating in discussion, and editing it throughout the review process. Dr. Aleix Beneyto and Dr. Ivan Contreras participated in the discussions and reviewed the article. Dr. Josep Vehí supervised the candidate during the development of the work.

Title: Bolus Insulin Calculation without Meal Information. A Reinforcement Learning Approach

Authors: **Sayyar Ahmad**, Aleix Beneyto, Ivan Contreras and Josep Vehí

Journal: Artificial Intelligence In Medicine

Volume: 134, Page: 102436, Published: November 2022

DOI: https://doi.org/10.1016/j.artmed.2022.102436

Quality index: SCIE 2022 Biomedical Engineering, Impact factor: 7.5, Q1 (17/96)

CHAPTER 2. REINFORCEMENT LEARNING STRATEGIES FOR INSULIN BOLUS CALCULATION IN A REALISTIC COHORT OF PATIENTS WITH TYPE 1 DIABETES

Artificial Intelligence In Medicine 134 (2022) 102436

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Research paper

Bolus Insulin calculation without meal information. A reinforcement learning approach

Sayyar Ahmad [a], Aleix Beneyto [a], Ivan Contreras [a], Josep Vehi [a,b,*]

[a] Department of Electrical, Electronic and Automatic Engineering, University of Girona, 17004 Girona, Spain
[b] Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 28001 Madrid, Spain

ARTICLE INFO

Keywords:
Reinforcement learning
Type 1 diabetes
Insulin bolus calculator
Artificial pancreas

ABSTRACT

In continuous subcutaneous insulin infusion and multiple daily injections, insulin boluses are usually calculated based on patient-specific parameters, such as carbohydrates-to-insulin ratio (CR), insulin sensitivity-based correction factor (CF), and the estimation of the carbohydrates (CHO) to be ingested. This study aimed to calculate insulin boluses without CR, CF, and CHO content, thereby eliminating the errors caused by misestimating CHO and alleviating the management burden on the patient. A Q-learning-based reinforcement learning algorithm (RL) was developed to optimise bolus insulin doses for in-silico type 1 diabetic patients. A realistic virtual cohort of 68 patients with type 1 diabetes that was previously developed by our research group, was considered for the in-silico trials. The results were compared to those of the standard bolus calculator (SBC) with and without CHO misestimation using open-loop basal insulin therapy. The percentage of the overall duration spent in the target range of 70–180 mg/dL was 73.4% and 72.37%, <70 mg/dL was 1.96 and 0.70%, and >180 mg/dL was 23.40 and 24.63%, respectively, for RL and SBC without CHO misestimation. The results revealed that RL outperformed SBC in the presence of CHO misestimation, and despite not knowing the CHO content of meals, the performance of RL was similar to that of SBC in perfect conditions. This algorithm can be incorporated into artificial pancreas and automatic insulin delivery systems in the future.

1. Introduction

Beta cells are responsible for the secretion of insulin for achieving homoeostasis of the glucose concentration in blood plasma [1,2]. Type 1 diabetes (T1D) is a chronic metabolic disorder caused by an autoimmune reaction that destroys pancreatic beta cells. Additionally, insulin deficiency results in elevated blood glucose (BG) concentrations and leads to several complications, such as cardiovascular complications, retinopathy, nephropathy, and neuropathy. [3]. Therefore, individuals with T1D rely on the administration of exogenous insulin to maintain their BG levels in the healthy range of 70–180 mg/dL to avoid diabetic complications.

Two types of insulin are generally used to reproduce the behaviour of insulin-producing beta cells based on the duration of insulin action. Long-acting and short-acting insulin are usually used to maintain BG levels during fasting and to achieve euglycaemia in the postprandial period, respectively. Short-acting insulin is used in continuous insulin infusion systems and is divided into two basic types: basal insulin that is used throughout the day during fasting periods and bolus insulin that is used to compensate for meals [4,5]. Basal-bolus insulin treatment regimens are used worldwide in two modes: multiple daily injections (MDI) and continuous subcutaneous insulin infusion (CSII). Among the two modes, MDI therapies are the most common and cost-effective; however, they are invasive and less efficient in terms of glycaemic control performance compared with CSII [6,7].

Recently, the development of automated insulin delivery (AID) systems has been progressing rapidly owing to the advancements in continuous glucose monitoring (CGM) technology and the achievement of reduced HbA1c levels. Additionally, the risk of severe hypoglycaemia has reduced and the quality of life has improved due to the application of CSII therapy [8]. An AID system comprises three components: a CGM sensor that provides the estimated BG concentration in real-time (usually every 5 min), a control algorithm that calculates the amount of insulin to be delivered, and an insulin pump to inject insulin subcutaneously. The rate of insulin infusion in AID systems is usually based on the action from the controller, and the bolus insulin is calculated based on patient-specific parameters, such as insulin-to-carbohydrate ratio (CR), correction factor (CF), target glucose (TG), and insulin on board (IOB) [9,10].

In clinical trials, AID systems have been largely successful in glycaemic control [11]. However, postprandial glucose control remains

* Corresponding author at: Department of Electrical, Electronic and Automatic Engineering, University of Girona, 17004 Girona, Spain.
E-mail address: josep.vehi@udg.edu (J. Vehi).

https://doi.org/10.1016/j.artmed.2022.102436
Received 13 May 2022; Received in revised form 24 October 2022; Accepted 28 October 2022
Available online 7 November 2022

a major challenge for AID systems because of several process characteristics, such as high intra-patient variability, delays in insulin action owing to the subcutaneous route of administration, slow dynamics of the existing insulin, and variations in the insulin sensitivity of the patient. Moreover, modelling the exact dynamics of glucose absorption in patients is not feasible because of uncertainty and intra and inter-day variability. To overcome these challenges, hybrid AID systems have been introduced that require the estimated carbohydrate (CHO) content of meals as an input to recommend insulin boluses [12].

Research groups worldwide have proposed various bolus insulin advisors to achieve optimal postprandial glucose control. A bolus insulin-dosing algorithm based on glucose measurements was proposed to match the bolus insulin recommended by expert clinicians [13]. A run-to-run control strategy and case-based reasoning (CBR) were applied to construct a decision support system that recommends an insulin bolus, and this system was tested on the UVa/Padova simulator [14,15]. The results of a six-week pilot study were subsequently published [16] and integrated into the Imperial College artificial pancreas (AP) controller for in-silico evaluation [17]. A K-nearest neighbours classification algorithm was used to predict postprandial glucose, and this algorithm was used to suggest adjustments to the bolus insulin for administration to an adult cohort using the UVa/Padova simulator [18]. Additionally, temporal CBR was developed for the personalised calculation of bolus insulin for a decision-support system [19]. A neural-network-based bolus calculator was implemented using CGM data and patient-specific parameters as features in a cohort of 100 patients using the UVa/Padova simulator [20]. The proposed bolus insulin advisors in the literature mentioned above rely on the meal information, estimation of CHO, CR and CF where as the proposed bolus insulin calculator is independent of these parameters.

Reinforcement learning (RL) is an emerging field of machine learning and artificial intelligence. The implementation of RL involves a hit and trial-based algorithm that trains an agent by exposing it to an unknown environment to learn the optimal actions required in a particular state based on the rewards received by the agent from the environment [21]. Furthermore, RL has been successful in several domains, such as beating the human champion in the Alpha Go game [22], optimal performance in playing Atari games [23], training the soccer-playing robot [24], optimal dosing of medications with sensitive treatment windows [25], and self-driving vehicles [26]. Additionally, a survey highlighting the applications of RL in intelligent healthcare systems has been recently published [27]. In recent years, the potential of RL has been exploited for BG regulation. An early review on the feasibility of RL for AID concluded that RL can provide adaptive personalised insulin administration to individuals with T1D [28]. A recent detailed systematic review suggested the application of advanced RL techniques and their combination with other machine learning algorithms to solve the BG regulation problem [29]. In the most recent developments, several researchers proposed RL-based bolus insulin advisors. A deep RL-based bolus insulin recommender system was designed by adapting the standard bolus calculator (SBC) by introducing gains for CR, CF, and IOB to be learned by the RL agent [30]. A Q-learning (QL)-based RL algorithm was implemented for bolus calculation by considering CR as a learning parameter and including programmable basal rates using the hybrid AP platform [31]. An RL-based decision support system was implemented and tested in-silico to recommend daily basal rates and insulin boluses to patients [32]. By using meal information to define the states, RL-based BG regulation was implemented for minimising risk parameters [33].

The RL methods developed in the literature discussed above require an estimation of the CHO content of meals as the input, which is prone to misestimation. Additionally, the patient is burdened with calculating the CHO content and providing the estimation to the system. Counting errors and misestimation of CHO deteriorate glycaemic performance and decrease duration over which the CGM values remain in the target range of 70–180 mg/dL [34–36]. In this study, a QL–RL strategy was developed to provide optimal bolus insulin to patients without requiring patient-specific parameters, such as CR and CF, and the estimations of the CHO content of meals. The developed bolus insulin calculator was integrated with an open-loop CSII therapy. Moreover, RL works on the principle of hit and trial to learn from experience; however, in the BG regulation problem, a failure may be dangerous and result in severe hypoglycaemia that can be lethal. Therefore, allowing the RL agent to randomly deliver insulin boluses from the action space and learn from its experience is impractical. To solve this issue, a multiple action space strategy was adopted to limit the amount of bolus insulin delivery, and the bolus on board (BIOB) was considered during the learning process.

2. Methodology

As an artificial intelligence method, RL progressively trains an agent by learning from direct interactions with its environment. The framework of interactions between an agent and the environment is defined in terms of states, actions, and rewards. Additionally, this is usually assumed to be a Markov decision process (MDP) represented by a tuple $M(S,\ A,\ P,\ r)$ with state space S, action space A, transition probability matrix $P(s_{t+1}|s_t,\ a_t)$, and immediate reward r. At each time step t, the agent is rewarded with $r_{t+1}\ (s_t,\ a_t,\ s_{t+1}) \in \mathbb{R}$ for action a_t taken in state s_t and transitioning into the new state s_{t+1}. The objective of the agent is to maximise the long-term return $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$, and $\gamma \in [0,\ 1)$ is defined as a discount factor that determines the present value of the future rewards [21].

The states are mapped into actions using a stochastic rule known as a policy and is denoted by $\pi : S \mapsto A$. Each policy π exhibits an action-value function $q_\pi(s,\ a) = \mathbb{E}_\pi[R_t|s_t = s,\ a_t = a]$, where $\mathbb{E}_\pi$ is the expected value of the long-term return R_t. Additionally, $q_\pi(s,\ a)$ quantifies the effect of action a in state s. An optimal action-value function $q^*(s,\ a) = \underbrace{max}_{\pi}\ q_\pi(s,\ a)$ corresponds to a unique optimal policy. Further, this optimal policy can be estimated using three major approaches namely, model-based dynamic programming, the Monte Carlo method, and the temporal difference (TD) method [21].

Furthermore, QL is an off-policy form of the TD method, in which the learning process is initialised using an arbitrary action-value function based on the transitions $(s_t,\ a_t,\ r_{t+1},\ s_{t+1})$ at each time step t. The action-value function is updated according to the following TD relation:

$$q_{t+1} = q_t(s_t, a_t) + \alpha[r_t + \gamma \underbrace{max}_{a}(q_t(s_{t+1}, a)) - q_t(s_t, a_t)], \tag{1}$$

where t denotes the current time step, s denotes the state, a denotes the action, r denotes the immediate reward, $\alpha \in [0,\ 1)$ denotes the learning rate, and γ denotes the discount factor.

The ultimate goal of a QL agent (QLA) is to develop an optimal policy that maximises the total expected reward. The agent is expected to deliver an optimal amount of bolus insulin according to a given state to achieve a desirable postprandial (PoP) glucose response after meals. The CHO content of the meal is unknown to the QLA in the approach presented in this study. A block diagram of the methodology is shown in Fig. 1.

The PoP BG regulation problem is defined in the MDP framework and the key elements are described in the following subsection.

2.0.1. States

The states were defined in terms of the glucose area under the curve (AUC), as well as maximum (MaxG) and minimum (MinG) CGM values, in the PoP period for the bolus calculation pertaining to PoP glucose control. The states were calculated as the current states, in which the action was taken, and the next states, in which the action was evaluated. For breakfast, the current state was calculated based on pre-prandial CGM data for a window of 4 h between 4 AM to 8 AM, whereas the next state was based on PoP CGM data for a window

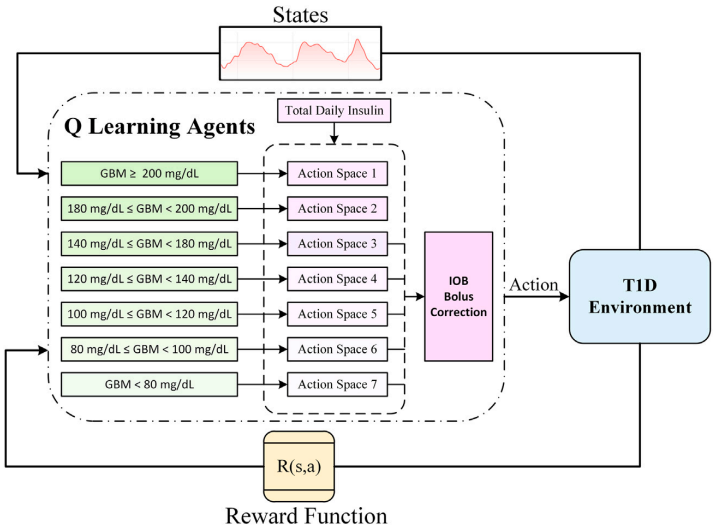

Fig. 1. Block diagram of the developed RL algorithm.

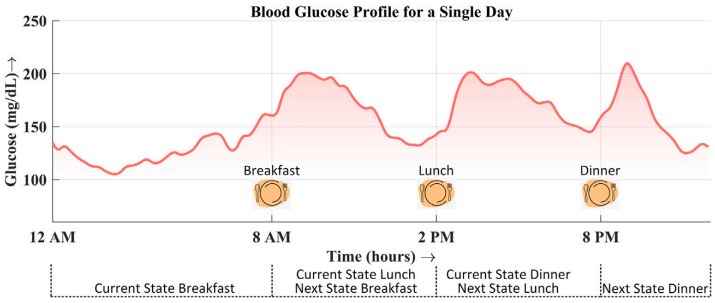

Fig. 2. State definition for the RL algorithm.

of 4 h between 8 AM to 12 PM. A similar symmetry was followed to obtain the current and next states associated with lunch and dinner. The state-space is expressed as follows:

$$S = \{AUC, MaxG, MinG\} \tag{2}$$

The mapping of states is illustrated in Fig. 2.

2.0.2. Actions

The action to be taken by the QLA is the unit of bolus insulin delivered to the patient after a meal. Seven action subspaces were defined based on the range of CGM values before meal intake. In the BG regulation problem, the QLA cannot explore freely because some actions may lead to severe conditions resulting in extreme CGM values. The action space was classified into seven subspaces to ensure safe exploration by the QLA in the environment. The range of bolus insulin values in a specific action space is defined based on the total daily insulin requirement of the patient, which is readily available from their medical history. The action space is given as follows:

$$A_m = \{a_1, a_2, \ldots .., a_L\}, \tag{3}$$

where $m = 1, 2, \ldots, 7$ and L are the numbers of actions in an action subspace, which is considered to be 20 in this study.

Once an action is selected by the QLA, the bolus insulin is further adjusted based on the estimation of BIOB according to the following relation:

$$u_b = \begin{cases} u_{RL} - \frac{u_{BIOB}}{k_{BIOB}} & : u_{RL} > \frac{u_{BIOB}}{k_{BIOB}} \\ u_b - 0.05 \times u_b & : u_{RL} < \frac{u_{BIOB}}{k_{BIOB}} \ \& \ 140 \leq G_{BM} < 180 \\ u_b - 0.1 \times u_b & : u_{RL} < \frac{u_{BIOB}}{k_{BIOB}} \ \& \ 120 \leq G_{BM} < 140 \\ u_b - 0.2 \times u_b & : u_{RL} < \frac{u_{BIOB}}{k_{BIOB}} \ \& \ 80 \leq G_{BM} < 120 \\ 0 & : u_{RL} < \frac{u_{BIOB}}{k_{BIOB}} \ \& \ G_{BM} < 80, \end{cases} \tag{4}$$

where u_b represents the adjusted bolus insulin, u_{RL} is the bolus insulin selected by QLA, u_{BIOB} is the estimated BIOB, G_{BM} is the CGM value before meals in mg/dL, and k_{BIOB} is a hyperparameter defining the amount of BIOB to be considered. Parameter k_{BIOB} was tuned for all action spaces and is different for different action spaces and meals such as breakfast, lunch, and dinner.

Eq. (4) shows that there is a reduction in the bolus insulin selected by the QLA. This reduction was applied to avoid any severe hypoglycaemia since the QLA does not rely on the CHO content of the meal and other relevant parameters. Reductions of 5, 10, and 20% were applied when the BIOB factor was greater than the bolus insulin selected by the QLA for several ranges of G_{BM}, as shown in Eq. (4). This was a rare occurrence, and usually, the insulin delivery was saturated at 0 during such conditions. However, as the proposed methodology used QLAs that were classified based on the ranges of G_{BM}, the same ranges were used here, and a reduction proportional to the risk of hypoglycaemia in the bolus insulin was applied. Specific values in percentages for the

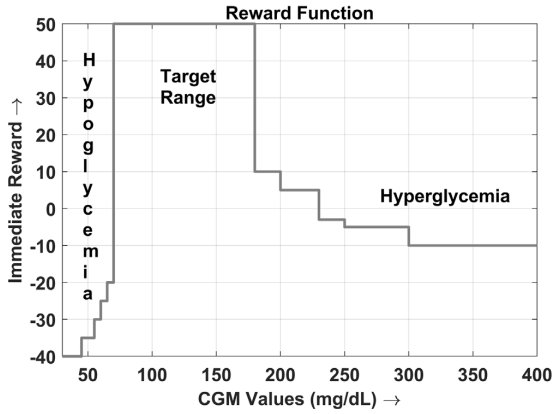

Fig. 3. Rewards assigned for the actions performed by the agent.

reduction in bolus insulin were considered based on experimentation. In extreme cases where $G_{BM} < 80$ mg/dL, the bolus insulin was saturated at 0.

Furthermore, u_{BIOB} was estimated according to a two-compartment dynamic model [37] expressed as follows:

$$\begin{cases} \dot{x_1}(t) = & u_b - k_{dia}x_1(t) \\ \dot{x_2}(t) = & k_{dia}(x_1(t) - x_2(t)) \\ u_{BIOB} = & x_1(t) + x_2(t), \end{cases} \tag{5}$$

where $x_1(t)$ and $x_2(t)$ are the two compartments and u_b is the bolus insulin. A populational value of 0.013 min^{-1} was considered for the constant k_{dia}, as discussed in [38].

2.0.3. Reward function

Immediate rewards for specific actions are defined in terms of numerical values. The actions leading to the desired states are assigned high values, whereas those resulting in undesirable states, such as hyperglycaemia and hypoglycaemia, are penalised by assigning negative reward values. The mapping of rewards and states is as follows:

$$R(s_t, a_t) = \begin{cases} 50 & : \text{MinG} \geq 70 \ \& \ \text{MaxG} \leq 180 \\ -10 \leq R(s_t, a_t) \leq 10 & : \text{MaxG} > 180 \\ -40 \leq R(s_t, a_t) \leq -15 & : \text{MinG} < 70 \end{cases} \tag{6}$$

The reward function is crucial in the training of the QLA as the feedback of the actions performed by the QLA is communicated in terms of numerical rewards. The CGM values being within the normal range can be considered the most desirable PoP state. Therefore, the QLA is incentivised by a high reward of 50 if the action results in the most desirable state. Additionally, PoP hypoglycaemia is avoided by penalising the QLA for actions that result in hypoglycaemia. The penalty is proportional to the severity of hypoglycaemia, and a penalty of −40 is imposed if the minimum CGM value in the PoP period is below 40 mg/dL. Furthermore, PoP hyperglycaemia is an undesirable state and is similarly avoided by assigning negative rewards; however, these reward values are lower than those for hypoglycaemia because hypoglycaemia is a common threat to PoP glycaemic control.

The complete reward function in relation to (6) is illustrated in Fig. 3.

2.0.4. Training

Seven QLAs were trained based on the classification of action spaces for each meal namely, breakfast, lunch, and dinner. Each QLA corresponded to the action space based on a specific CGM value before meal intake. Additionally, the QLAs were trained simultaneously while switching between the various defined action spaces on a certain day during the training process. The QLAs were expected to converge to an optimal policy after the training phase was completed.

The training process was initiated with an arbitrary action-value function $q\ (s,\ a)$ and continued until the terminal number of iterations was reached. Each iteration represented a simulation for a single day. This indicated that the data for the current and next states, actions performed, and rewards for all three meals were calculated each day. In every iteration, the corresponding entity was updated in $q\ (s,\ a)$ based on the TD formula described in Eq. (1). The maximum number of iterations considered in the simulations was 1500. Therefore, the total duration of the training session was 1500 days for all the QLAs. The number of training iterations for each QLA differed depending on the number of occurrences of the CGM value before meal intake corresponding to each QLA. This implied that the sum of training iterations for a certain meal for all QLAs is ≤ maximum number of iterations (1500).

To identify the actions with the highest contributions to maximising the long-term reward, the QLA must visit all the action–state pairs. To achieve this, the action is chosen according to an ϵ-greedy policy. This indicates that a random action is chosen with a probability of ϵ, and the best possible action corresponding to a state is chosen with a probability of $1 - \epsilon$, as described by the following relation:

$$a_c = \begin{cases} arg \underbrace{max}_{a_c} q(s_t,\ a_c) & : with\ probability\ 1 - \epsilon \\ random\ action & : with\ probability\ \epsilon \end{cases} \tag{7}$$

Initially, the value of ϵ was set to 0.9 to allow the RL agent to explore for most of the duration. Gradually, the value of ϵ was reduced exponentially according to $\epsilon = e^{-\frac{i}{800}}$ (i refers to the current iteration) with a minimum value of 0.1 to continue the exploration with a probability of 10% and prevent the QLA from falling into the local optimum policy. The learning rate α was maintained as dynamic starting with a large value of 0.9 to allow the new values to contribute more toward updating the action-value function and exponentially decay with a minimum value of 0.3 as $\alpha = e^{-\frac{i}{1000}}$.

2.0.5. Training scenario

The scenario considered a three-meal protocol with breakfast at 08:00 containing 30–50 g of CHO, lunch at 14:00 containing 50–70 g of CHO, and dinner at 20:00 containing 60–80 g of CHO. Intra-patient variability was introduced as sinusoidal variations in meal absorption, insulin pharmacodynamics, and insulin sensitivity [39], and the variability in meal composition was additionally included during the training session of the QLA.

2.0.6. Algorithm

The main steps involved in the developed algorithm are described in Table 1.

3. In-silico validation scenario and setup

The in-silico evaluation was performed in the FDA-approved UVa/Padova simulator. A cohort of 68 patients, previously developed by our research group [40], was considered in this study because this resembled a real cohort of challenging patients with T1D from the Hospital Clínic de Barcelona. Here, SBC was implemented on 68 patients, and this was considered to be a baseline for comparison. The patient-specific parameters, such as CR and CF, were first determined for all patients according to clinical guidelines [41]. The SBC [42] can be expressed as follows:

$$u_{SBC} = \frac{CHO}{CR} - \frac{(BG_C - BG_T)}{CF} - \widehat{IOB}, \tag{8}$$

where u_{SBC} is the insulin bolus, BG_C is the current value of BG, BG_T is the target value of BG, CHO is the carbohydrate content of the meal,

Table 1
Q-learning algorithm for bolus insulin calculation.

1. **Initialize** $\gamma = 0.8$, $\alpha = 0.9$ *and* $\epsilon = 0.9$
2. **Initialize** $q(s_t, a_t)$, $\forall\ s \in S$, $\forall\ a \in A$ to 0
3. **Evaluate** the current state s_t
4. **Repeat**
   - **a.** Select the action space based on the CGM value before the meal
   - **b.** Choose the action according to ϵ greedy policy
   - **c.** Apply the IOB bolus adjustment, if required
   - **d.** Apply the action and observe the reward and next state
   - **e.** Update the action-value function: $q_{t+1} \leftarrow q_t(s_t, a_t) + \alpha[r_t + \gamma \underbrace{max}_{a}(q_t(s_{t+1}, a)) - q_t(s_t, a_t)]$
   - **f.** Exponential decay of α and ϵ: $\alpha \leftarrow e^{-\frac{t}{1000}}$, $\epsilon \leftarrow e^{-\frac{t}{800}}$
   - **g.** $s_t \leftarrow s_{t+1}$
5. **Obtain** the optimal policy after maximum number of iterations are completed

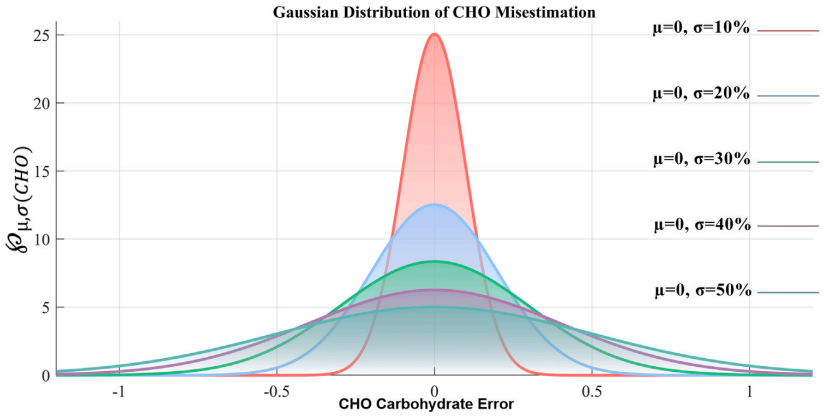

Fig. 4. Gaussian distribution curves for CHO misestimation, in which CHO $Carb = 10g$, μ = mean and σ = SD.

CR is the carbohydrate-to-insulin ratio, CF is the correction factor, and $\widehat{IOB}$ is the estimated insulin on board. In this study, the duration between meals was considered to be 6 h and $\widehat{IOB}$ was assumed to be negligible in the implementation of the SBC.

Additionally, a scenario of 14 days was considered for in-silico testing. A description of the meals is provided in the training scenario section. Two schemes were considered for including CHO misestimations by patients to show its implication while using SBC. The CHO misestimation is implemented as a Gaussian distribution, as explained in a recent study [43], and is given by:

$$u_{SBC} = \frac{\widetilde{CHO}}{CR} - \frac{(BG_C - BG_T)}{CF} - IOB, \tag{9}$$

where

$$\widetilde{CHO} = CHO + \wp CHO \tag{10}$$

Here, $\wp$ represents the relative error owing to CHO misestimation and is represented by a Gaussian distribution $\wp(Mean,\ Standard\ Deviation)$. The mean corresponds to the systematic error, and the standard deviation (SD) reflects the random error. The two schemes considered in this study correspond to $\wp(0, 20\%)$ and $\wp(0, 40\%)$ to show the effect of CHO misestimation on the BG profile. A mean of zero was selected because the CHO misestimation focused on CHO counting errors by the patients and not by the system. The distribution curves for the CHO misestimations are shown in Fig. 4.

4. Results

To investigate the performance of the proposed RL algorithm, the results are presented in terms of the standardised CGM metrics for clinical care described in a consensus report by the American Diabetes Association and the European Association for the Study of Diabetes (EASD) [44]. The results were based on a scenario of 14 days, as recommended by the consensus report, and were presented as median (interquartile range, 25%–75%). The SBC results were presented as a baseline for comparison. Two cases of CHO misestimations were additionally reported for SBC therapy.

The standardised CGM metrics and insulin information are presented in Table 2. The mean and median CGM values slightly improved compared to those obtained as a result of SBC therapy (with and without CHO misestimation). The coefficient of variation (CV) of the BG values was found to be better in the case of SBC with no CHO misestimation. However, the CV was better for the RL algorithm when CHO misestimation was included in the SBC. The glucose monitoring index (GMI) was slightly improved by using the RL algorithm. The percentage of the overall duration, during which the CGM values (PCGM) were present in the tight target ranges of 70–140 mg/dL and 70–180 mg/dL, was greater in the case of the RL algorithm than that in the case of SBC, SBC20, and SBC40. Additionally, the PCGM values below 54 mg/dL and in the range of 54–69 mg/dL were lower in case of SBC with no CHO misestimation. However, the performance of the RL algorithm was better than that of SBC with CHO misestimation corresponding to an SD of 40% and is almost similar when the SD of CHO misestimation is set to 20%.

Subsequently, the p-values of the RL algorithm for SBC and the two versions of SBC with CHO misestimation are calculated and indicated in Table 2. The RL algorithm outperformed SBC40, and also exhibited significant improvement in mean, median, and maximum CGM values, as well as in PCGM values in the target range, as compared to SBC20. However, the RL algorithm exhibited results similar to those of SBC without CHO misestimation.

Table 2
Standardised glycaemic and insulin metrics for RL, SBC, SBC20, and SBC40.

Performance indicator	SBC	SBC20	SBC40	RL
Mean CGM	153.6 (145.5–166.9)*	151.2 (142.3–165.4)*	156 (146.7–169.3)*	149.9 (141.7–161.5)
Median CGM	152.7 (143–164.9)*	150.2 (139.8–162.3)*	155.3 (144.5–168)*	150 (139.7–157.8)
Max CGM	281 (263.3–310.8)	288.2 (267.9–309)*	299.9 (280.3–325.6)*	282.8 (260.5–308.9)
Min CGM	55.6 (43.6–68.3)	40.4 (28.8–51.8)	36.4 (26.5–53.9)	43 (32.3–51.5)
CV	25.9 (23.6–29.6)	28.6 (25.6–32.5)	30.2 (25.8–33.5)*	27.5 (25.6–30.9)
GMI	6.98 (6.79–7.30)*	6.93 (6.72–7.27)*	7.04 (6.815–7.36)*	6.9 (6.7–7.175)
Below 54	0 (0–0.53)	0.71 (0.17–2.37)	0.91 (0.075–2.378)	0.73 (0.1–1.48)
54 to 69	0.45 (0.04–1.15)	1.31 (0.64–1.99)	1.03 (0.465–1.64)	1.05 (0.58–2)
70 to 140	36.5 (28.3–46.3)*	35.8 (29.4–45.8)*	33.8 (26–41.4)*	38.8 (29.8–47.2)
70 to 180	72.4 (60–81.9)	71.5 (60.2–78)*	67.3 (55.6–74.6)*	73.4 (65.8–81.8)
181 to 250	22.1 (15.6–30.6)*	22.4 (16.2–30)*	25.6 (19.4–32.8)*	22.1 (15.7–27.2)
Above 250	1.4 (0.3–4.7)	1.4 (0.4–4.4)	2.4 (1.2–6.3)*	1 (0.3–3.6)
Basal Per Day	6.5 (5.5–8)	6.5 (5.5–8)	6.5 (5.5–8)	6.5 (5.5–8)
Bolus Per Day	15.5 (11.9–20.3)	16.4 (12.4–21.4)	15.4 (11.7–20.2)	17.2 (13.5–21.4)
TDI	22.5 (18.7–26)	23.1 (19.2–27.1)	22.3 (18.6–26)	23.6 (20.4–27.3)

p values (wilcoxon signed rank test) are considered in comparison with RL.
SBC20: SBC with CHO misestimation, with an SD of 20%.
SBC40: SBC with CHO misestimation, with an SD of 40%.
*$p\ value\ <\ 0.05$.

5. Discussion

Different approaches have been proposed in the existing literature to calculate the optimal bolus insulin dose for people with T1D. Most of these approaches are based on the concept of SBC, which is used in insulin infusion pumps. Additionally, several researchers have demonstrated the benefits of RL by adapting the tuning of the basic formulation of SBC. More specifically, CR, CF, and IOB have been optimised using RL, which requires the CHO content to be digested by patients with T1D as input.

Existing methods for bolus infusion in insulin pumps require the mandatory entry of CHO estimates each time a meal is consumed. Two major problems are associated with this approach. First, these methods are prone to errors because precisely estimating the CHO content in a meal is not straightforward, thereby resulting in the common misestimation by T1D patients [45]. Second, calculating CHO places an extra burden on patients each time they consume a meal.

This study focused on designing a bolus insulin calculator that is independent of CHO estimation. To achieve this objective, a bolus insulin calculator that uses RL was designed. A TD-based QL topology of RL was utilised in this study. Additionally, RL overcomes the problem of personalisation because learning was based on individual users. The main challenge for the RL agents was being unaware of the CHO content while delivering the insulin bolus to the patient. To overcome this problem, the action space was classified into subspaces based on the CGM value immediately before meal intake and the actions were defined according to the total daily insulin requirement of the patients. Second, the states were defined using an alternative method (the current method corresponds to preprandial CGM values next to postprandial CGM values) on the same day as the meal is consumed, which increases the effectiveness of the learning process. Finally, BIOB was included as a means to decrease the occurrence of severe hypoglycaemia.

The primary BG outcomes are presented in Table 2. To demonstrate the effect of CHO misestimation in SBC on BG outcomes and compare this with the developed bolus insulin calculator, two scenarios of CHO misestimation were considered in this study. In the first scenario, SBC40 refers to the CHO misestimation error with an SD of 40%. The results clearly revealed that RL outperformed SBC40 because the patient did not need to enter the CHO content, thereby making RL independent of CHO. The p values for PCGM below 54 mg/dL and in the range of 54–69 mg/dL indicated that the performances of SBC40 and RL were statistically similar. This was because RL did not calculate the bolus insulin based on CHO content, and a single meal that contained low CHO over the 14 days resulted in an increased PCGM below 70 mg/dL. The second scenario, SBC20, refers to a CHO

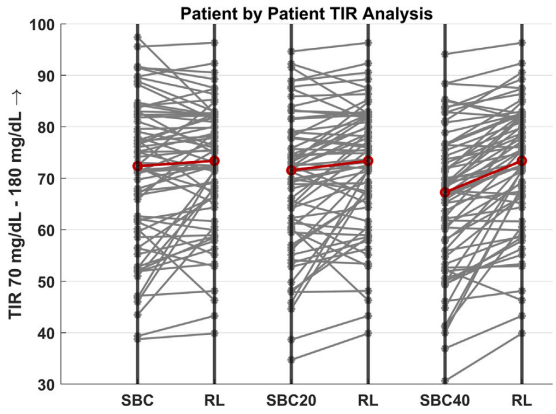

Fig. 5. Head-to-head comparison of the patients for the percentage of time spent in euglycemia. The grey lines represent individual patients, and the red lines correspond to the median values of the cohort. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

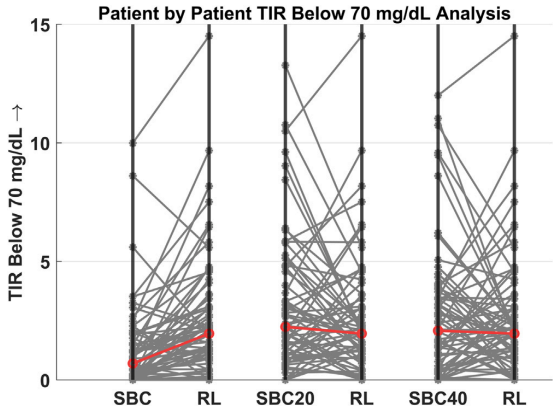

Fig. 6. Head-to-head comparison of patients for the percentage of the overall duration spent in hypoglycaemia. The grey lines represent individual patients, and the red lines correspond to the median values of the cohort. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

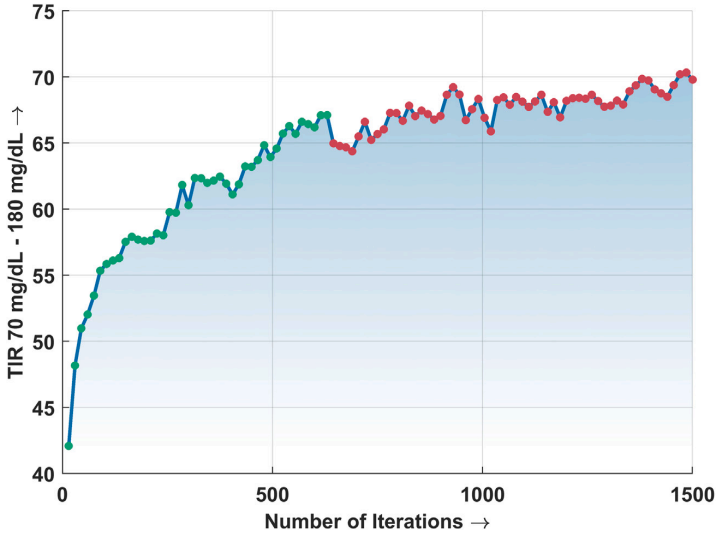

Fig. 7. Performance analysis of the QLAs with respect to the number of iterations.

misestimation error with an SD of 20%. Similarly, RL exhibited a better performance in terms of mean, median, and maximum CGM values, as well as PCGM, in the target range of 70–180 mg/dL and above the target range of 180–250 mg/dL as compared to SBC20. However, the performance of the bolus insulin calculator was similar in the case of the minimum CGM value and PCGM below 54 mg/dL, in the range of 54–69 mg/dL, and above 250 mg/dL. Comparing SBC with RL demonstrated that both delivered the same performance statistically; however, the performance of RL in the case of PCGM in the tight target range of 70–140 mg/dL was superior. However, SBC outperformed RL in the case of PCGM below 54 mg/dL and the range of 54–69 mg/dL. The overall performance of RL was satisfactory, as it did not use the CHO content in meals to calculate the bolus insulin. This is a primary step towards effectively compensating for the unannounced meals.

The patient-by-patient comparison of PCGM values in the target range (TR) of 70–180 mg/dL is illustrated in Fig. 5, which reveals that RL improves the performance of patients with a TR below 60%. For patients with approximately 80% TR, RL either maintained the glycaemic performance or slightly improved or degraded it. In a few patients, RL was substantially degrading the glycaemic performance for several reasons. First, the sub-action spaces influenced bolus insulin delivery by providing a range of bolus insulin values, such that the RL agent can identify the optimal value in a particular state, and the action space was based on the TDI requirement of the patient. The TDI information of these patients, which is based on clinical data, was possibly not optimal, thereby leading to the degradation in glycaemic performance. Second, the hyperparameters of the RL algorithm were tuned for the entire cohort, as it is impractical to tune these for individuals in large cohorts of T1D patients. The glycaemic performance of these patients can possibly be improved if the algorithm is individually tuned. However, the overall performance of RL was slightly superior to those of SBC and SBC20 and substantially higher than that of SBC40.

The patient by patient comparison of PCGM values below the target range of < 70 mg/dL (TBR) is illustrated in Fig. 6. The RL algorithm maintained the TBR in the same range for patients with TBR less than 4%, which is also a clinical target for TBR. Notably, in a few patients, the TBR was increased, thereby resulting in the degradation of glycaemic performance, and these were the same patients as discussed in the above paragraph. For one patient whose TBR was approximately 10%, the TBR increased substantially when using RL. This patient was from the group, in which the glycaemic performance was degraded. Fig. 6 clearly reveals that in many patients, the TBR was considerably decreased owing to RL, especially compared with SBC20 and SBC40.

The improvement in the policy owing to the increase in the number of iterations is demonstrated in Fig. 7. The median value of the cohort for PCGM within the target range of 70–180 mg/dL is shown with respect to the number of iterations. The points on the graph represent the policy results after every 15 iterations. The overall nature of policy growth was exponential, and the green dots represented a faster improvement in the early stages of training and the red dots represented a steady improvement later in the policy of the QLAs.

The RL algorithms proposed in the literature exclusively use CHO information and the concept of SBC. A deep RL algorithm was demonstrated [30] and tested on ten adults and ten adolescents using the UVa/Padova platform. The PCGM in the target range improved from 74.1% to 80.9% and from 54.9% to 61.6% for adults and adolescents, respectively. A QL algorithm was presented in a hybrid artificial pancreas setting to learn optimal CRs and basal rates using model predictive control as the primary controller for 50 subjects in a simulation environment based on Hovorka's model [31]. The PCGM in the target range improved from 67% to 86.7% over 5 weeks. In the present study, the PCGM achieved in the target range was 73.4% because the adult cohort used in this study comprised 68 patients, which was different from those in the previously reported literature. Second, the calculation of the bolus was not based on patient-specific parameters, which is a key advantage of this study. Finally, open-loop basal insulin therapy was implemented with constant basal rates while implementing the proposed bolus insulin calculator.

The main limitation of this study was the validation of RL on a virtual cohort, as implementation in clinical settings is challenging because of uncertain conditions in real-life scenarios. However, a modified virtual cohort was considered in this study. Variability was also incorporated into the simulator, including the random rate of absorption for meals, random CHO content in meals, and circadian variability in insulin sensitivity, to capture real-life conditions. Translation of the proposed algorithm to a clinical setting can be achieved following four major steps. (1) Generation of the virtual cohort that represents the real patients with type 1 diabetes in terms of CGM metrics as considered in this study. (2) Training of the Reinforcement Learning (RL) agents in the simulation environment before testing in the clinical trials as a

starting point. 3) Preclinical testing and validation of the RL algorithm under real-life scenarios with real patient data. The CGM data of the real patient will be used by the RL agent to recommend bolus insulin for meals in this phase. 4) Implementation of the personalised trained agents for real patients in clinical trials with safety precautions. Second, the patients were able to cope with severe hypoglycaemia. Learning about unexpected small meals without using the CHO content in meals is complicated for the agent. During the 14 days, a few unexpected small meals may result in a significant increase in the PCGM below 70%. This is because, for most of the patients, this value is close to 0%, as presented in Fig. 6. In a few patients, PCGM below 70% was increased considerably, thereby causing an overall increase in PCGM below 70%, as shown in Table 2. The standardised CGM metrics presented in Table 2 reveal that the designed bolus insulin calculator can be a better choice than SBC because the exact CHO content in meals to be digested is not required.

6. Conclusions

In this study, a novel bolus insulin calculator was designed based on the QL concept of RL with the novelty of not requiring CHO content estimations, CR, and CF. An RL-based multi-agent strategy was used based on the CGM data immediately before meal intake.

The proposed algorithm outperformed SBC with CHO misestimation and exhibited a performance similar to that of SBC. The main advantage of the proposed algorithm is the control of PoP glucose without knowing the CHO content in a meal, thereby eliminating the errors caused by CHO misestimation and alleviating the burden of entering CHO information into CSII systems.

The present study can be extended in two major directions. First, the Q-learning algorithm can be replaced with more sophisticated algorithms, such as deep Q-learning or deep actor-critic, to achieve an improved performance to outperform SBC. Second, the developed bolus insulin controller can be incorporated into AP and AID systems to improve glucose control performance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Katsarou A, Gudbjörnsdottir S, Rawshani A, Dabelea D, Bonifacio E, Anderson BJ, et al. Type 1 diabetes mellitus. Nat Rev Dis Primers 2017;3(1):1–17.
[2] Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. Lancet 2014;383(9911):69–82.
[3] Marcovecchio ML, Lucantoni M, Chiarelli F. Role of chronic and acute hyperglycemia in the development of diabetes complications. Diabetes Technol Ther 2011;13(3):389–94.
[4] Attri B, Goyal A, Gupta Y, Tandon N. Basal-bolus insulin regimen for hospitalised patients with COVID-19 and diabetes mellitus: a practical approach. Diabetes Therapy 2020;11(9):2177–94.
[5] Russell-Jones D, Bode BW, De Block C, Franek E, Heller SR, Mathieu C, et al. Fast-acting insulin aspart improves glycemic control in basal-bolus treatment for type 1 diabetes: results of a 26-week multicenter, active-controlled, treat-to-target, randomized, parallel-group trial (onset 1). Diabetes Care 2017;40(7):943–50.
[6] Chen BZ, Zhang LQ, Xia YY, Zhang XP, Guo XD. A basal-bolus insulin regimen integrated microneedle patch for intraday postprandial glucose control. Sci Adv 2020;6(28):eaba7260.
[7] Tsarkova P, Chakarova N, Dimova R, Grozeva G, Todorova A, Serdarova M, et al. CSII is related to more stable glycemia in adults with type 1 diabetes. Endocrine 2021;1–5.
[8] Priesterroth L, Grammes J, Clauter M, Kubiak T. Diabetes technologies in people with type 1 diabetes mellitus and disordered eating: A systematic review on continuous subcutaneous insulin infusion, continuous glucose monitoring and automated insulin delivery. Diabetic Med 2021;e14581.
[9] Gross TM, Kayne D, King A, Rother C, Juth S. A bolus calculator is an effective means of controlling postprandial glycemia in patients on insulin pump therapy. Diabetes Technol Ther 2003;5(3):365–9.
[10] Lewis D. How it started, how it is going: the future of artificial pancreas systems (automated insulin delivery systems). J Diabetes Sci Technol 2021;15(6):1258–61.
[11] Christiansen M, Bartee A, Lalonde A, Jones RE, Katz M, Wolpert H, et al. Performance of an automated insulin delivery system: results of early phase feasibility studies. Diabetes Technol Ther 2021;23(3):187–94.
[12] Tsoukas MA, Majdpour D, Yale J-F, El Fathi A, Garfield N, Rutkowski J, et al. A fully artificial pancreas versus a hybrid artificial pancreas for type 1 diabetes: a single-centre, open-label, randomised controlled, crossover, non-inferiority trial. Lancet Digit Health 2021;3(11):e723–32.
[13] Palerm CC, Zisser H, Bevier WC, Jovanovič L, Doyle FJ. Prandial insulin dosing using run-to-run control: application of clinical data and medical expertise to define a suitable performance metric. Diabetes Care 2007;30(5):1131–6.
[14] Herrero P, Pesl P, Reddy M, Oliver N, Georgiou P, Toumazou C. Advanced insulin bolus advisor based on run-to-run control and case-based reasoning. IEEE J Biomed Health Inf 2014;19(3):1087–96.
[15] Pesl P, Herrero P, Reddy M, Xenou M, Oliver N, Johnston D, et al. An advanced bolus calculator for type 1 diabetes: system architecture and usability results. IEEE J Biomed Health Inf 2015;20(1):11–7.
[16] Reddy M, Pesl P, Xenou M, Toumazou C, Johnston D, Georgiou P, et al. Clinical safety and feasibility of the advanced bolus calculator for type 1 diabetes based on case-based reasoning: a 6-week nonrandomized single-arm pilot study. Diabetes Technol Ther 2016;18(8):487–93.
[17] Herrero P, Bondia J, Adewuyi O, Pesl P, El-Sharkawy M, Reddy M, et al. Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator–in silico evaluation under intra-day variability. Comput Methods Programs Biomed 2017;146:125–31.
[18] Aiello EM, Toffanin C, Messori M, Cobelli C, Magni L. Postprandial glucose regulation via KNN meal classification in type 1 diabetes. IEEE Control Syst Lett 2018;3(2):230–5.
[19] Brown D, Aldea A, Harrison R, Martin C, Bayley I. Temporal case-based reasoning for type 1 diabetes mellitus bolus insulin decision support. Artif Intell Med 2018;85:28–42.
[20] Cappon G, Vettoretti M, Marturano F, Facchinetti A, Sparacino G. A neural-network-based approach to personalize insulin bolus calculation using continuous glucose monitoring. J Diabetes Sci Technol 2018;12(2):265–72.
[21] Sutton RS, Barto AG. Reinforcement learning: an introduction. MIT Press; 2018.
[22] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science 2018;362(6419):1140–4.
[23] Hosu I-A, Rebedea T. Playing atari games with deep reinforcement learning and human checkpoint replay. 2016, arXiv preprint arXiv:1607.05077.
[24] Riedmiller M, Gabel T, Hafner R, Lange S. Reinforcement learning for robot soccer. Auton Robots 2009;27(1):55–73.
[25] Nemati S, Ghassemi MM, Clifford GD. Optimal medication dosing from suboptimal clinical examples: A deep reinforcement learning approach. In: 2016 38th annual international conference of the IEEE engineering in medicine and biology society. EMBC, IEEE; 2016, p. 2978–81.
[26] Sallab AE, Abdou M, Perot E, Yogamani S. Deep reinforcement learning framework for autonomous driving. Electron Imaging 2017;2017(19):70–6.
[27] Coronato A, Naeem M, De Pietro G, Paragliola G. Reinforcement learning for intelligent healthcare applications: A survey. Artif Intell Med 2020;109:101964.
[28] Bothe MK, Dickens L, Reichel K, Tellmann A, Ellger B, Westphal M, et al. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. Expert Rev Med Devices 2013;10(5):661–73.
[29] Tejedor M, Woldaregay AZ, Godtliebsen F. Reinforcement learning application in diabetes blood glucose control: A systematic review. Artif Intell Med 2020;104:101836.
[30] Zhu T, Li K, Kuang L, Herrero P, Georgiou P. An insulin bolus advisor for type 1 diabetes using deep reinforcement learning. Sensors 2020;20(18):5058.
[31] Jafar A, El Fathi A, Haidar A. Long-term use of the hybrid artificial pancreas by adjusting carbohydrate ratios and programmed basal rate: A reinforcement learning approach. Comput Methods Programs Biomed 2021;200:105936.
[32] Sun Q, Jankovic MV, Budzinski J, Moore B, Diem P, Stettler C, et al. A dual mode adaptive basal-bolus advisor based on reinforcement learning. IEEE J Biomed Health Inf 2018;23(6):2633–41.
[33] Zhu J, Zhang Y, Rao W, Zhao Q, Li J, Wang C. Reinforcement learning for diabetes blood glucose control with meal information. In: International symposium on bioinformatics research and applications. Springer; 2021, p. 80–91.
[34] Brazeau A, Mircescu H, Desjardins K, Leroux C, Strychar I, Ekoé J, et al. Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. Diabetes Res Clin Pract 2013;99(1):19–23.
[35] Kawamura T, Takamura C, Hirose M, Hashimoto T, Higashide T, Kashihara Y, et al. The factors affecting on estimation of carbohydrate content of meals in carbohydrate counting. Clin Pediatr Endocrinol 2015;24(4):153–65.

[36] Roversi C, Vettoretti M, Del Favero S, Facchinetti A, Sparacino G, Consortium H-R. Modeling carbohydrate counting error in type 1 diabetes management. Diabetes Technol Ther 2020;22(10):749–59.
[37] Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R. Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Trans Biomed Eng 2004;52(1):3–12.
[38] Beneyto A, Bertachi A, Bondia J, Vehi J. A new blood glucose control scheme for unannounced exercise in type 1 diabetic subjects. IEEE Trans Control Syst Technol 2018;28(2):593–600.
[39] León-Vargas F, Garelli F, De Battista H, Vehí J. Postprandial blood glucose control using a hybrid adaptive PD controller with insulin-on-board limitation. Biomed Signal Process Control 2013;8(6):724–32.
[40] Ahmad S, Ramkissoon CM, Beneyto A, Conget I, Giménez M, Vehi J. Generation of virtual patient populations that represent real type 1 diabetes cohorts. Mathematics 2021;9(11):1200.
[41] Walsh J, Roberts R. Pumping insulin: everything you need for success on a smart insulin pump. vol. 4, Torrey Pines Press San Diego, CA; 2006.
[42] Zisser H, Robinson L, Bevier W, Dassau E, Ellingsen C, Doyle III FJ, et al. Bolus calculator: a review of four "smart" insulin pumps. Diabetes Technol Ther 2008;10(6):441–4.
[43] Roversi C, Vettoretti M, Del Favero S, Facchinetti A, Choudhary P, Sparacino G. Impact of carbohydrate counting error on glycemic control in open-loop management of type 1 diabetes: Quantitative assessment through an in silico trial. J Diabetes Sci Technol 2021;19322968211012392.
[44] Holt RI, DeVries JH, Hess-Fischl A, Hirsch IB, Kirkman MS, Klupa T, et al. The management of type 1 diabetes in adults. a consensus report by the American diabetes association (ADA) and the European association for the study of diabetes (EASD). Diabetologia 2021;1–44.
[45] Buck S, Krauss C, Waldenmaier D, Liebing C, Jendrike N, Högel J, et al. Evaluation of meal carbohydrate counting errors in patients with type 1 diabetes. Exp Clin Endocrinol Diabetes 2021.

2.3 An Automatic Deep Reinforcement Learning Bolus Calculator for Automated Insulin Delivery Systems

This publication proposes a FAID system for people with T1D. The proposed system consist of a PD CL controller for continuous insulin delivery, a DRL insulin bolus calculator for insulin bolus calculation and an UKF meal detector for the detection of meals to fully automate the process. The CLC algorithm and meal detector were previously developed by our group. The insulin bolus calculator is an advanced version of the algorithm presented in section 2.2 and is explained in this article. The proposed FAID system is compared to the HAID system that relies on SBC and the results showed a comparable performance. The candidate's contributions to this publication included developing, designing, and implementing the FIAD, as well as its validation in in-silico trials. The candidate also contributed to writing the paper, participating in discussion, and editing it throughout the review process. Dr. Taiyu Zhu participated in the development of the DRL algorithm and was supervised by Dr. Pantelis Georgiou during this phase. Dr. Aleix Beneyto and Dr. Ivan Contreras participated in the discussions and reviewed the article. Dr. Josep Vehí supervised the candidate during the development of the entire work.

Title: An Automatic Deep Reinforcement Learning Bolus Calculator for Automated Insulin Delivery Systems
Authors: **Sayyar Ahmad**, Aleix Beneyto, Taiyu Zhu, Ivan Contreras, Pantelis Georgiou and Josep Vehí
Journal: Scientific Reports
Status: Submitted, November 2023
Quality index: SCIE 2023 Multidisciplinary Sciences, Impact factor: 4.6, Q2 (22/73)

An automatic deep reinforcement learning bolus calculator for automated insulin delivery systems

Sayyar Ahmad[1], Aleix Beneyto[1], Taiyu Zhu[2], Ivan Contreras[1], Pantelis Georgiou[2], Josep Vehi[1,3*]

[1*]Department of Electrical, Electronic and Automatic Engineering, University of Girona, Girona, 17004, Spain.
[2]Centre for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, London, U.K.
[3]Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, 28001, Spain.

*Corresponding author(s). E-mail(s): josep.vehi@udg.edu;
Contributing authors: sayyar.ahmad@udg.edu; aleix.beneyto@udg.edu; taiyu.zhu17@imperial.ac.uk; ivan.contreras@udg.edu; pantelis@imperial.ac.uk;

Abstract

In hybrid automatic insulin delivery (HAID) systems, meal disturbance is compensated by feedforward control, which requires the announcement of the meal by the patient with type 1 diabetes (DM1) to achieve the desired glycemic control performance. The calculation of insulin bolus in the HAID system is based on the amount of carbohydrates (CHO) in the meal and patient-specific parameters, i.e. carbohydrate-to-insulin ratio (CR) and insulin sensitivity-related correction factor (CF). The estimation of CHO in a meal is prone to errors and is burdensome for patients. This study proposes a fully automatic insulin delivery (FAID) system that eliminates patient intervention by compensating for unannounced meals. This study exploits the deep reinforcement learning (DRL) algorithm to calculate insulin bolus for unannounced meals without utilizing the information on CHO content. The DRL bolus calculator is integrated with a closed-loop controller and a meal detector (both previously developed by our group) to implement the FAID system. An adult cohort of 67 virtual patients based on the modified UVa/Padova simulator was used for in-silico trials. The percentage of the overall duration spent in the target range of **70–180** mg/dL was **71.2%** and **76.2%**, $<$ **70** mg/dL was **0.9%** and **0.1%**, and $>$ **180** mg/dL was **26.7%** and **21.1%**, respectively, for

the FAID system and HAID system utilizing SBC including CHO misestimation. The proposed algorithm can be exploited to realize FAID systems in the future.

Keywords: automatic insulin delivery, artificial pancreas, unannounced meals, deep reinforcement learning

1 Introduction

Type 1 diabetes (DM1) is a metabolic disorder caused by an autoimmune reaction that leads to the destruction of insulin-secreting beta cells in the pancreas. It leads to insulin deficiency and elevated levels of blood glucose (BG) referred to as hyperglycemia. Long-term complications as a consequence of chronic DM1 may be microvascular and macrovascular. Retinopathy, nephropathy, and neuropathy are microvascular complications, whereas cardiovascular disease, artery inflammation and injury in the peripheral system, and cerebrovascular disease are among the macro-vascular complications [1].

The BG of normal subjects is maintained in a narrow range of 70 – 180 mg/dL, which is called normoglycemia. In people with DM1, normoglycemia is achieved by the lifelong administration of exogenous insulin generally under the supervision of physicians [2]. Recent technological advancements have had a considerable effect on the management of DM1. Automatic insulin delivery (AID), also referred to as artificial pancreas (AP), systems are developed for the treatment of DM1 to overcome hypo and hyperglycemia and reduce long-term complications associated with DM1. The three core components of an AID system are a continuous glucose monitoring device (CGM) that generally provides BG measurements every 5 minutes, an insulin pump to continuously deliver insulin, and an algorithm to calculate the optimal insulin rate to be administered to the subject with DM1 [3].

Advancements in CGM technology make it possible to analyze glycemic trends, patterns, and key information with improved accuracy, increased duration, and mean absolute relative difference (MARD) $\leq 10\%$. CGM systems can be used to calculate insulin dosing rates [4]. AID systems have been reported to be a safe and effective approach to the treatment of DM1 [5]. However, optimal control of postprandial BG remains a concern for AID systems for various reasons, including significant delays in insulin action as a result of the subcutaneous route, slow response of the available insulin analogues, variability in the insulin sensitivity of DM1 subjects, and high intrapatient variability. Moreover, accurate modeling of glucose absorption is not possible because of uncertainty and intraday and interday variations. To improve glycemic performance, researchers have proposed hybrid AID (HAID) systems based on feedforward control schemes, usually proportional to the carbohydrates content (CHO) in meals [6].

HAID systems provide automated insulin delivery via closed-loop control algorithms and patient-initiated bolus insulin delivery to compensate for announced meals based on various insulin bolus calculators [7]. HAID systems have shown improved

glycemic control performance with a reduction in the risk of hypoglycemia and are among the most advanced insulin delivery systems available for DM1 subjects [8].

The CHO content in meals is one of the main parameters and nutritional determinants of postprandial BG levels in DM1. It is recommended to accurately measure CHO for improved BG control performance [9]. However, the task of CHO counting is burdensome and prone to estimation errors, with average misestimations of around 20% in adults [10]. The quality of life in people with DM1 is negatively influenced by CHO counting and makes them less confident while interacting with peers, especially around food. To maintain the precision of the CHO count, standardized foods are more likely to be chosen by people with DM1, which can negatively affect their dietary choices [11]. Furthermore, the level of literacy required to count CHO can be an obstacle for many patients with DM1, leading to the selection of packaged processed foods over whole foods (grains, fruits, etc.) due to the relative ease provided by the nutritional information label [12]. HAID systems possess the benefit of meal announcement but they must be robust to missed meals and other factors discussed above. Therefore, a fully closed-loop AID (FAID) system is highly desirable to avoid the need for CHO counting and announcing meals in patients with DM1 [13].

Several algorithms have been proposed to automate the process of detecting meals in patients with DM1. A few of the proposals include fuzzy logic [14], various Kalman filters [15, 16], model-based detection utilizing an autoregressive model and real-time CGM data [17], detection of an increase in the glucose rate [18], and artificial intelligence (AI)-based meal detection [19]. Attempts have also been made to compensate for unannounced meals. The algorithms proposed include the Kalman filter to avoid CHO counting for automatic glucose regulation [20], disturbance observer, and feedforward compensation of unannounced meals [21], an automatic bolus priming system [22], and a meal absorption model for AP [23].

Reinforcement learning (RL) is a rapidly developing field of AI that has found success in many domains. A detailed systematic review reported that advanced RL algorithms can play a vital role in developing AID systems [24]. Recently, several researchers have proposed insulin bolus calculators that exploit different models of the RL algorithm [25–27]. The reported methodologies rely on information about the CHO content in meals and the meal announcement, resulting in HAID systems. In comparison, this study attempts to develop a FAID system with a insulin bolus calculator that uses the DRL algorithm and achieves glycemic control performance near that of HAID.

This work aims to develop a FAID system based on a deep reinforcement learning (DRL) insulin bolus calculator to compensate for unannounced meals and to eliminate interventions from patients with DM1. A closed-loop proportional–derivative (PD) control algorithm is used for the computation of the continuous insulin delivery rate. For the detection of meals, unscented Kalman filter (UKF) predictions are utilized based on the CGM and insulin data. The FAID system is compared to two versions of the HAID system, one utilizing the standard bolus calculator (SBC) for the compensation of meal disturbances along with CHO misestimation and the other utilizing the proposed DRL insulin bolus calculator.

2 Methodology

In this work, a DRL-based insulin bolus calculator is designed and integrated with a closed-loop controller and a UKF-based meal detector to compensate for unannounced meals in patients with DM1. The proposed DRL-based insulin bolus calculator is an advanced version of an algorithm published by our group [28] driven by meal detection and does not require information on the CHO content in meals, thereby fully closing the AID control loop. Continuous insulin delivery is achieved by a closed-loop PD controller with a safety auxiliary feedback element (SAFE) introduced in [29]. The detection of meals is based on an in-house algorithm utilizing an augmented minimal model and a UKF along with the insulin and CGM data [30]. A schematic of the overall strategy is given in figure 1.

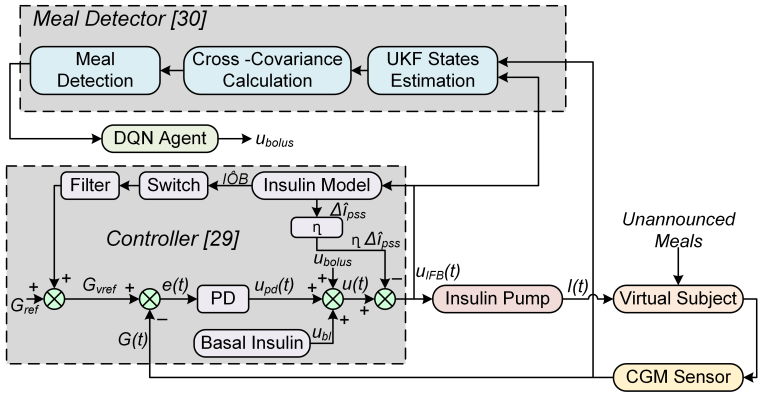

Fig. 1 Block diagram of the proposed FAID system.

2.1 PD Controller

The control strategy involves two loops: an inner loop comprising the insulin feedback system (IFB) that relies on the PD algorithm and an outer loop that provides a safety layer to exploit the concept of insulin on board (IOB).

The three insulin components constitute the inner control action: u_{bl} the basal insulin profile of the patient, u_{bolus} the insulin bolus, and the PD control action resulting in insulin action given by:

$$u(t) = k_p \left[e(t) + \tau_d \frac{dG(t)}{dt} \right] + u_{bl}(t) + u_{bolus} \tag{1}$$

where $k_p = \frac{60 \times TDI}{\tau_d \times 1500}$ (U/hr) is the proportional gain, TDI is the total daily insulin, $e(t)$ is the error in glucose concentration and $\tau_d = 90$ (min) is the derivative time constant.

The safety layer is based on sliding mode reference conditioning (SMRC) and comprises three parts: 1) a model to estimate IOB; 2) a sliding mode referencing block (SMR); and 3) a 1^{st}-order low-pass filter to smooth the reference adaptation. The outer safety layer modifies the reference glucose concentration (G_{ref}) under defined conditions to ensure that the IOB is bounded ($IOB \in [0, \overline{IOB}]$). Essentially, this is accomplished by a suspension of insulin infusion caused by the controller's reference modification. G_{ref} is modified to a virtual reference G_{vref} in case the estimated ($\widehat{IOB}$) approaches dangerously or exceeds the maximum allowed IOB ($\overline{IOB}$). This phenomenon provides robustness against delays in the subcutaneous route.

The insulin absorption model [31] is utilized to account for the estimated IOB and is given below.

$$\begin{aligned} \frac{dc_1(t)}{dt} &= u(t) - k_{dia}c_1(t) \\ \frac{dc_2(t)}{dt} &= k_{dia}(c_1(t) - c_2(t)) \\ \widehat{IOB}(t) &= c_1(t) + c_2(t) \end{aligned} \tag{2}$$

where $u(t) = u_{pd}(t) + u_{bl} + u_{bolus}$, $c_1(t)$ and $c_2(t)$ are two compartments representing the basal and bolus IOB conditions and k_{dia} is a constant time that accounts for the duration of insulin action.

The SMR block is based on the concept of invariance control [32] with $IOB(t)$ being the variable to be bounded and belonging to the set:

$$\sum = \{x(t) | s(t) = \widehat{IOB}(t) - \overline{IOB}(t) \leq 0\} \tag{3}$$

where $x(t)$ is the state of the system and $s(t)$ is the sliding surface, defined as:

$$s(t) = \widehat{IOB}(t) - \overline{IOB}(t) + \tau \left(\dot{\widehat{IOB}}(t) - \dot{\overline{IOB}}(t) \right) \tag{4}$$

The invariance of the region $\sum$ is achieved using the following discontinuous function.

$$\nu(t) = \begin{cases} \nu^+ & \text{if} \quad s(t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Finally, the smoothness of the reference change is achieved by applying a first-order low-pass filter:

$$\frac{d\nu_f(t)}{dt} = -\lambda(\nu_f(t) - \nu(t)) \tag{6}$$

A widely used mechanism of IFB in AP systems is also implemented. The plasma insulin concentration is estimated online; then, insulin control action is inhibited proportionally. This gives rise to a new insulin control action given by:

$$u_{IFB} = u(t) - \eta(\widehat{i}_p(t) - \widehat{i}_{pss}(t)) = u(t) - \eta\Delta\widehat{i}_{pss}(t) \tag{7}$$

where $\widehat{i}_p(t)$ is the estimated value and $\widehat{i}_{pss}(t)$ is the steady-state estimated value of the plasma insulin concentration. $\Delta\widehat{i}_{pss}(t)$ is the deviation of the plasma insulin concentration from the basal infusion. Further details are presented in [29].

2.2 Meal Detector

The meal detector algorithm [30] takes the rate of insulin infusion and CGM value as inputs and estimates a disturbance term via an extended minimal model utilizing the UKF. The glucose subsystem comprises Bergman equations [33] as follows:

$$\frac{G_{pl}(t)}{dt} = -(p_1 + X(t))G_{pl}(t) + p_1 G_{bl} + \frac{D(t)}{V_g} \tag{8}$$

where $G_{pl}(t)$ is the blood plasma glucose concentration, $X(t)$ reflects insulin in the remote compartment, G_{bl} is basal glucose, p_1 is the insulin-independent rate of plasma glucose, $D(t)$ is the disturbance term included as an extended model state, and V_g is the volume distribution.

Subcutaneous glucose is represented by a first-order system [34] as given below:

$$\frac{G_s(t)}{dt} = -\frac{1}{\tau}G_s(t) + \frac{g}{\tau}G_{pl}(t) \tag{9}$$

$$\frac{X(t)}{dt} = -p_2 X(t) + p_3 I(t) \tag{10}$$

where $G_s(t)$ is the subcutaneous glucose concentration, τ is the time constant of the system, and the static gain is represented by g. $X(t)$ reflects insulin in the remote compartment, p_2 is the disappearance rate of remote insulin and p_3 captures insulin sensitivity. The insulin subsystem model is the same as that represented by equation 2, and the concentration of plasma insulin [34] is given by:

$$\frac{I(t)}{dt} = -k_f I(t) + \frac{1}{V_i}.\frac{S_2(t)}{t_{max,I}} \tag{11}$$

where V_i is the distribution volume, k_f is the fractional rate of disappearance, and $t_{max,I}$ is the time to maximum absorption of insulin.

After estimation of the model states given by equations 2 and 8 to 11 through UKF, the cross-covariance is calculated between the two sequences $G_s(k)$ (from the CGM data) and $D_{dif}(k)$ (forward difference of disturbance term) over a window of specified length. G_{s_n} and D_{dif_n} are jointly stationary random processes, and their cross-covariance sequence is defined as the cross-correlation of mean-removed sequences [35], as given below:

$$\Psi_{G_s,D(m)} = E\{(G_s(n+m) - \mu_{G_s})(D_{diff}(n) - \mu_{D_{diff}})^*\} \tag{12}$$

where the mean values of the random processes are represented by μ_{G_s} and $\mu_{D_{diff}}$, E stands for the expectation operation, and $*$ represents the complex conjugate.

Meal consumption is assumed if a predefined threshold is exceeded by the cross-covariance between G_{sn} and D_{difn} with respect to the last three consecutive samples (15 minutes). As a safety measure, meals are not detected during the night period (23h - 6h).

The meal detector can be tuned regarding three settings with respect to the threshold and window size for cross-covariance [30]. The three settings refer to 1) highest sensitivity (high true positives), 2) trade-off (high true positives and low false positives), and 3) lowest false positives. In this study, trade-off tuning is used because the highest sensitivity is prone to false positives and will result in the delivery of insulin bolus at times other than meals, leading to extreme hypoglycemia. The third setting was not used because it decreases the true positives substantially.

A meal detection flag is triggered if:

$$Meal = \begin{cases} True \text{ if } cG_s, D_{dif}(m) \geq T \wedge D_{dif}(k) > 0 \wedge G_s(k) - G_s(k-3) > 0 \\ False \text{ otherwise} \end{cases} \quad (13)$$

where T is the predefined threshold and $cG_s, D_{dif}(m)$ represents the raw cross-covariance, as given in [30].

2.3 The DRL Algorithm

The problem is first formulated as a Markov decision process (MDP) to implement the training of the RL agent. An MDP is defined in terms of state space S, action space A, the transition probability $P(s_{t+1} \mid s_t, a_t)$ of the next state (s_{t+1}) given action (a_t) is taken in the current state (s_t), and an immediate reward r_t, mathematically represented as a tuple $M(S, A, P, r)$. In DRL, the agent is based on a combination of RL and a category of artificial neural networks (ANNs), specifically deep neural networks (DNNs), and is termed a deep Q-network (DQN). The DQN aims to learn actions that result in the maximum total expected reward. The total expected reward can be represented as $E_R = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ...]$, where $\gamma \in [0, 1)$ is the discount factor defining the contribution of future rewards and r_t is the immediate reward at time step t.

In DRL, the mapping of states into actions to be taken by the DQN is termed the policy and is represented by $\pi : S \rightarrow A$. The quality of the policy is represented by the action-value function $Q_\pi(s, a)$. The policy that leads to the maximum E_R is a unique optimal policy π^* and results in a unique optimal action-value function $Q^*(s, a)$. In this work, a fully connected DNN is used to learn π^* to approximate $Q^*(s, a, \theta) \approx Q^*(s, a)$, where θ refers to the parameters of the DNN. The final goal of training the DQN is to learn π^*, which implies that the agent will take the best possible action in a given state. In RL, the optimal action-value function is obtained on the basis of the notion of the Bellman equation [36] given below:

$$Q^*(s,a) = E_{s_{t+1}}[r + \gamma \underbrace{max}_{a} Q^*(s_{t+1}, a) \mid s, a] \tag{14}$$

The optimal policy is obtained by dynamic programming to iteratively evaluate:

$$Q_{t+1}(s,a) = Q_t(s_t, a_t) + \alpha[r_t + \gamma \underbrace{max}_{a_{t+1}} Q_t(s_{t+1}, a_{t+1}) \mid s, a] \tag{15}$$

According to Bellman's identity, Q_t converges to Q^* as $t \to \infty$, where $\alpha \in [0, 1)$ is the learning rate. This approach to RL (Q-Learning) requires the states to be discrete and lack generalization. Therefore, in DRL, $Q^*(s,a)$ is approximated by a nonlinear function approximator such as DNN. To estimate $Q^*(s,a)$, the DQN uses fixed Q-targets by maintaining the $Q(s,a,\theta)$ and the target $\hat{Q}(s,a,\hat{\theta})$, both having the same architecture. The two approximators improve the stability of optimization by updating the parameters of $\hat{Q}(s,a,\hat{\theta})$ periodically to the latest parameters of $Q(s,a,\theta)$ [37]. The parameters are updated every 15 iterations during the training phase in the proposed algorithm.

In this work, multi-DQNs are implemented and trained. Typically, there are three meals per day, i.e., breakfast, lunch, and dinner. The protocol for meals is described later in the scenario subsection under Results. For each meal, the action space is divided into 8 subaction spaces based on the 8 ranges defined for the CGM value before meal intake. The action space is explained later in subsection 2.3.2. A DQN is trained for each subaction space, resulting in the implementation of 8 DQNs for each meal and leading to a total of 24 DQNs corresponding to three meals a day.

The motivation behind introducing a multi-DQN strategy is to obtain a personalized DRL agent for each subaction space with respect to meals. This approach will limit the learning experience of each DQN to that specific subaction space and meal, thereby providing greater chances of better performance. In summary, it is the personalization of a DQN based on the meal and the CGM value before meal intake.

A fully connected ANN composed of three hidden layers is considered to represent a DQN for the approximation of $Q^*(s,a,\theta)$. Each hidden layer is composed of 28 nodes. The whole network consists of 5 layers, including the input and output layers. The input layer represents 15 parameters (defining the state), and the output layer shows the Q-value of each action taken in that particular state. The Q-value used in RL measures the effectiveness of the action taken in a certain state. The DQN architecture is presented in figure 2.

The main components of the MDP model considered in this study are explained below:

2.3.1 State Space

The states are represented as the current state and the next state. DQN takes the action in the current state, which is then evaluated in the next state during the training process. In DRL, the states are continuous in nature, and discretization of states is not required. The current state is based on the pre-prandial CGM data of 4 hours. The parameters considered are the maximum CGM value, minimum CGM value, area under the curve (AUC) of the CGM data, and the 12 CGM values (1-hour data)

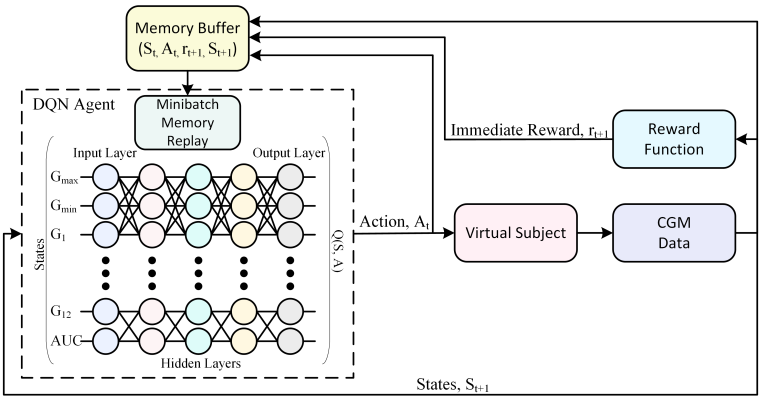

Fig. 2 Representation of the DRL algorithm based on DQN. The states feed the DQN to approximate the optimal policy $Q^*(s, a)$. A randomly extracted mini-batch of experiences is also utilized by the DQN. The action A_t corresponds to the maximum Q-value, which is the insulin bolus to be delivered to the patient. As a result, a transition occurs for the state S_{t+1}, and the memory buffer is updated with the new experience.

before meal intake, summing to 15 parameters. AUC is calculated for the CGM data representing hyper or hypoglycemia only. In the next state, the same parameters are calculated based on the 4-hour postprandial CGM data, and the 12 CGM values are considered for the last hour of the postprandial window. The states are based on the CGM data, so the ANN can learn hidden patterns in the BG profile. The state space can be represented as:

$$S = \{G_{max}, G_{min}, G_{t_m-1}, G_{t_m-2}, G_{t_m-3}, ...G_{t_m-k}, AUC\} \tag{16}$$

where G_{max} is the maximum CGM value, G_{min} is the minimum CGM value, t_m is the meal detection time, k is the sample, G_{t_m-k} is the CGM value at $t_m - k$ and AUC is the area under the curve over 4 hours of CGM data corresponding to hyper and hypoglycemia only.

2.3.2 Action Space

The action space for a certain meal is classified into 8 subaction spaces (SASs) corresponding to 8 different BG ranges. The number of SASs in a previous study [28] was 7, but the number has now been increased to 8 to enhance safety based on BG before a meal and to provide greater flexibility to the agent in the choice of insulin bolus. According to the CGM value (sample) before meal intake (G_{BM}), belonging to one of the 8 defined ranges, the corresponding SAS is selected for action by the DQN agent. The actions considered in this study are discrete and are the bolus insulin units to be delivered to the patient, as described in [28]. The action space can be represented as:

$$A = \begin{cases} A_1 & G_{BM} \geq 200 \\ A_2 & 180 \leq G_{BM} < 200 \\ A_3 & 160 \leq G_{BM} < 180 \\ A_4 & 140 \leq G_{BM} < 160 \\ A_5 & 120 \leq G_{BM} < 140 \\ A_6 & 100 \leq G_{BM} < 120 \\ A_7 & 80 \leq G_{BM} < 100 \\ A_8 & G_{BM} < 80 \end{cases} \tag{17}$$

where A is the action space and $A_i \mid i = 1, 2...8$ represents the SASs. $A_i = \{a_1, a_2 \dots a_j\}$, where $a_1 \dots a_j$ are the bolus insulin units calculated based on the total daily insulin requirement of the patient and the value of G_{BM}. In this study, $j = 15$, i.e., an agent can choose among 15 actions from a chosen SAS. The selection of SAS for a single iteration is demonstrated in figure 3.

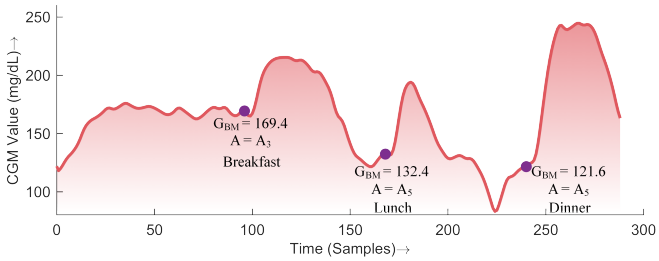

Fig. 3 Demonstration of the selection of a subaction space based on the CGM value before a meal.

The insulin bolus selected as an action is further adjusted according to the bolus insulin on board (BOB) to ensure safety and avoid extreme hypoglycemic events. The adjustment can be represented as a piece-wise function:

$$u_{ad} = \begin{cases} a_j - \widehat{BOB}/k_{BOB} & a_j > \widehat{BOB}/k_{BOB} \ \& \ G_{BM} \geq 180 \\ decrease\ a_j\ by\ 5\% & a_j < \widehat{BOB}/k_{BOB} \ \& \ 140 \leq G_{BM} < 180 \\ decrease\ a_j\ by\ 10\% & a_j < \widehat{BOB}/k_{BOB} \ \& \ 120 \leq G_{BM} < 140 \\ decrease\ a_j\ by\ 20\% & a_j < \widehat{BOB}/k_{BOB} \ \& \ 80 \leq G_{BM} < 120 \\ a_j & otherwise \end{cases} \tag{18}$$

where u_{ad} is the adjusted insulin bolus to be delivered, a_j is the action chosen by the agent, $\widehat{BOB}$ is the estimated BOB and k_{BOB} is a hyperparameter that is tuned separately for all SASs and three meals. A two-compartment model is used to estimate BOB [38].

2.3.3 Reward Function

An immediate reward is assigned to the actions of the DQN based on the next state. If the postprandial blood glucose is in the normal range (70-180 mg/dL), a high reward is given to the DQN. If the action taken by the DQN results in hyper or hypoglycemia, the agent is penalized. The numerical values assigned to the immediate rewards are illustrated in figure 4 and can be expressed as a piece-wise defined function:

$$r_t = \begin{cases} 50 & 70 \leq G_{maxp} \leq 180 \\ 20 & 180 \leq G_{maxp} < 200 \\ 10 & 200 \leq G_{maxp} < 230 \\ -5 & 230 \leq G_{maxp} < 250 \\ -15 & 250 \leq G_{maxp} < 300 \\ -20 & G_{maxp} \geq 300 \\ -30 & 65 \leq G_{minp} < 70 \\ -40 & 60 \leq G_{minp} < 65 \\ -50 & 55 \leq G_{minp} < 60 \\ -60 & 50 \leq G_{minp} < 55 \\ -70 & 45 \leq G_{minp} < 50 \\ -80 & G_{minp} < 45 \end{cases} \tag{19}$$

where G_{maxp} and G_{minp} represent the maximum and minimum glucose values in the postprandial period, respectively. In the case of the simultaneous occurrence of G_{maxp} and G_{minp}, the value associated with G_{minp} is considered. The reward function is designed to reward the DQN agent for optimal performance, i.e., maintaining postprandial glucose in the normal range. The reward values are considered positive for mild hyperglycemia to avoid hypoglycemic episodes. There exists a trade-off between avoiding hyper and hypoglycemia, as no information on the meal content is available. On the other hand, the occurrence of hypoglycemia is penalized proportionally to the intensity of the event to avoid severe postprandial hypoglycemia.

2.3.4 Implementation

The concept of memory replay is typically used in DRL for stability and convergence of the DNN [37]. This concept is also implemented in the proposed methodology. Memory is defined for each DQN. The memory buffer (MB) consists of the past experiences of the agent and can be represented as:

$$MB = \{\xi_1, \xi_2, \xi_3, ..., \xi_n\} \tag{20}$$

where n is the size of the MB and ξ is a single iteration experience given by:

$$\xi = \{s_t, a_t, r_t, s_{t+1}\} \tag{21}$$

To generate the memory, a simulation is performed for 1500 days, where the actions are taken randomly and the experiences are stored in MB. The size of the MB varies for each DQN and depends on the number of occurrences of a specific A_i during the whole simulation.

A cohort of 67 virtual patients previously developed by our group is considered in this study [39]. A protocol of three meals (breakfast at 08:00 of 30 g-50 g, lunch

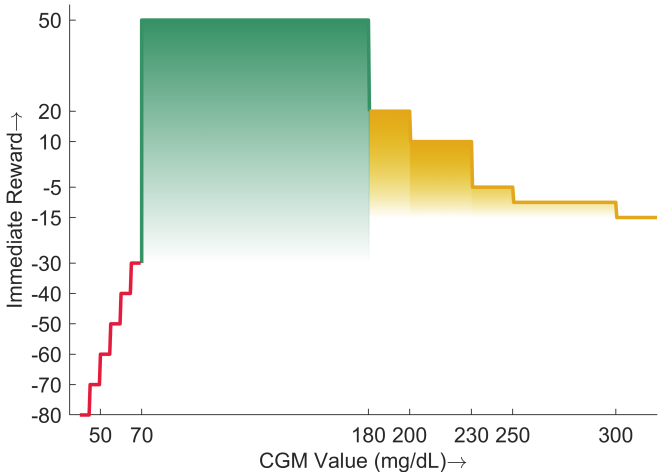

Fig. 4 Reward function for the proposed DRL algorithm. The green region represents the immediate reward when G_{pp} is in a healthy range, yellow for hyperglycemia and red for hypoglycemia.

at 14:00 of 50 g-70 g, and dinner at 20:00 of 60 g-80 g) was considered during the training session. The CHO content in meals was chosen randomly from the amounts indicated. All the meals were unannounced, and the agent only took action whenever it received a positive indicator from the meal-detector. The sources of intrapatient variability included sinusoidal variations in insulin pharmacodynamics and insulin sensitivity (circadian variability) and randomness in the rate of absorption of meals [40]. An epsilon greedy policy is used to choose the action, and an immediate reward is assigned to the DQN agent according to the reward function presented in equation 19. In a single iteration, the corresponding MB is updated with the new experience, and the weights of the DQN are updated based on past experiences from MB. The loss function used to optimize the DQN's weights is based on the Bellman equation and is given for a k_{th} iteration as follows:

$$L_k(\theta_k) = E_{(s_t,a_t,r_t,s_{t+1})} \sim U(MB) \left[\left(r_t + \gamma \underbrace{max}_{a_{t+1}} \left(\hat{Q}(s_{t+1}, a_{t+1}; \hat{\theta_k}) - Q(s, a; \theta_k) \right) \right)^2 \right] \tag{22}$$

During learning, the Q-learning updates are applied to the mini-batches $(s_t, a_t, r_t, s_{t+1}) \sim U(MB)$ extracted randomly from MB through uniform distribution, where γ is the discount factor, $\hat{Q}(s_{t+1}, a_{t+1}; \hat{\theta_k})$ is the target DQN in iteration k, whose weights $\hat{\theta_k}$ are updated periodically with the DQN $Q(s, a; \theta_k)$ weights. The

DRL algorithm implemented in this study to calculate the insulin bolus is presented in Algorithm 1.

Algorithm 1 Deep Reinforcement Algorithm for Insulin Bolus Calculation for the FAID

Generate the replay MB to capacity n
Initialize the action-value function $Q(s, a; \theta)$ with weights θ from a uniform random distribution with bounds [0, 1]
Initialize the target action-value function $\hat{Q}(s, a; \hat{\theta})$ with weights $\hat{\theta} = \theta$
Observe the current state s_t
while $iterations < 1500$ **do**
    if meal detection flag is triggered **then**
        Choose the SAS A_i based on the value of G_{BM}
        Explore with probability ϵ ; a random action a_j
        Exploit with probability 1-ϵ; $a = \underbrace{max}_{a_j} Q(s, a_j; \theta)$
        Apply the BIOB adjustment according to equation 18 and take action
        Observe the next state and assign the immediate reward
        Modify the MB with a new experience $\{s_k, a_k, r_k, s_{(k+1)}\}$
        Sample a random mini-batch of N experiences from MB
        Set (double DQN algorithm) $A_{max} \leftarrow arg \underbrace{max}_{a_{k+1}} Q(s_{k+1}, a_{k+1}; \theta)$
        $y_k \leftarrow r_k + \gamma \hat{Q}(s_{k+1}, A_{max}; \hat{\theta})$
        Perform a gradient descent step on $(y_k - Q(s_k, a_k; \theta))^2$ with respect to weights θ
        if iterations count == 15 **then**
            Set $\hat{Q}(s, a; \hat{\theta}) = Q(s, a; \theta)$
            Reset iteration counter
        end if
    else[no action]
    end if
end while
Obtain the set of trained DQNs

3 Results

3.1 In-silico scenario and benchmark

A virtual cohort with 67 virtual patients based on a modified version of the FDA-approved UVa/Padova simulator is used [39]. The simulation time for in-silico trials is 14 days. The meals delivered include breakfast at 07:00, lunch at 13:00, a snack at 17:00, and dinner at 20:00, composed of a CHO content selected randomly from 30 g-50 g, 50 g-70 g, 30 g-50 g, and 60 g-80 g, respectively. During the simulations, the

meal time is varied ±30 minutes around the time mentioned above. Variability is also incorporated, including randomness in the rate of absorption for meals, random CHO content in meals, and circadian variability in insulin sensitivity, to emulate real-life conditions [40].

Three insulin delivery systems are compared in this study, and they all utilize a PD closed-loop controller for continuous insulin delivery. First, the HAID system is implemented utilizing SBC for the insulin bolus calculation, and the CHO misestimation error is included to be more realistic. This baseline system is represented as HAID SBC MCHO. The CHO misestimation error is incorporated as a Gaussian distribution according to the recently published methodology [41]. To implement the SBC, the parameters required are the carbohydrate-to-insulin ratio (CR) and correction factor (CF), calculated based on clinical guidelines [42]. Then, the formula for SBC used in this study is given below [43]:

$$u_{bolus} = \frac{CHO}{CR} + \frac{(BG_k - BG_T)}{CF} - \widehat{IOB} \tag{23}$$

where u_{bolus} is the bolus insulin, BG_k is the CGM value at the time of delivering the bolus, BG_T is the target glucose value and $\widehat{IOB}$ is the estimated insulin on board.

Second, the HAID system with the proposed DRL insulin bolus calculator is represented as HAID DRL. As the DRL bolus calculator is independent of the CHO content in meals, CHO misestimation is not an issue in this case. In both HAID systems, all the meals are announced, hence the name hybrid. In this case (HAID DRL), the DRL algorithm was tuned and trained in the setting of announced meals. This implies that the meal detector was not used and the insulin bolus was delivered at meal time during the training session of DQN agents. The simulation performed for generating the memory (required for the memory replay concept in the DRL algorithm) was also based on announced meals. HAID DRL is included to explicitly show the difference in the glycemic performance induced by unannounced meals.

Finally, the proposed FAID system is the main contribution of this study. The FAID system is based on the DRL algorithm for bolus insulin dosing, but all the meals are unannounced. The delivery of insulin bolus is triggered by a signal from the meal detector whenever a meal is detected.

3.2 Comparison

To draw a comparison and investigate the performance of the proposed FAID system, the outcomes of the in-silico simulations are presented in the standardized core CGM metrics, as reported in a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) [44].

The standardized CGM metrics and insulin information are presented in table 1. The mean and median CGM values reported for the FAID system were statistically similar to those of the HAID systems, as indicated by the p-values. The extreme CGM values, i.e., minimum and maximum in the FAID system, were more spread, leading to a slightly higher glycemic variability, as indicated by the higher CV compared to that of the HAID systems. The FAID system achieved a similar glucose monitoring index (GMI), as reflected by the p-value.

The percentage of the CGM values (PCGM) reported for the ranges provided in table 1 showed an overall increase of 5% in the PCGM below 70 mg/dL and above 250 mg/dL (hypoglycemia and hyperglycemia) for the FAID system. Specifically, the difference in hypoglycemia (below 70 mg/dL) was 0.9%, and that in hyperglycemia (above 250 mg/dL) was 4.1%, which is in accordance with the designed reward function. Hypoglycemia was penalized more than hyperglycemia since a hypoglycemic excursion is riskier than a hyperglycemic excursion of the same magnitude.

According to the p-values, the differences in PCGM ranges are significant, except for the tight target range (70 - 140 mg/dL). Importantly, all the values achieved were in the range recommended by the ADA consensus report [44].

The glycemic risk index (GRI), a measure of the quality of glycemia based on hypoglycemia and hyperglycemia components using CGM tracings [45], is also provided. However, the overall performance of the FAID system was comparable to that of HAID versions, despite unannounced meals.

Table 1 Comparison of standardized CGM metrics and insulin data for the FAID system.

Performance Indicator	*HAID SBC MCHO*[1]	*HAID DRL*[2]	*FAID*[3]
Mean CGM	153.1 (147.3 - 161.1)	155.8 (149.7 - 160.3)	156.1 (148 - 167.5)
Median CGM	146.7 (140.9 - 155.9)	149.2 (144.8 - 154.9)	147.9 (140.2 - 158.4)
Max CGM	306.7 (283.6 - 324.6)	293.6 (265 - 347.3)	317.7 (296.2 - 347.5)*
Min CGM	66.6 (43.9 - 74.7)	72.3 (49.7 - 81.7)	43.1 (32.3 - 61.1)*
CV	25.6 (23.3 - 28.5)	24.1 (21.9 - 28.2)	30.6 (27.7 - 32.3)*
GMI	7 (6.8 - 7.2)	7 (6.9 - 7.1)	7 (6.8 - 7.3)
Below 54	0 (0 - 0.3)	0 (0 - 0.2)	0.4 (0 - 0.9)*
54 to 69	0.1 (0 - 0.5)	0 (0 - 0.3)	0.5 (0.2 - 1)*
70 to 140	42.2 (33.7 - 48.3)	38.9 (33.1 - 43.4)	41.1 (33.9 - 48.1)
70 to 180	76.2 (69.6 - 82.1)	75.7 (71 - 81.3)	71.2 (60.2 - 77.2)*
181 to 250	19.2 (14.7 - 23.4)	19.8 (15.5 - 24.1)	22.6 (18.8 - 27.5)*
Above 250	1.9 (1.1 - 3.1)	1.6 (0.7 - 4)	4.1 (1.8 - 8.4)*
GRI	20.9 (16.6 - 26.5)	20.9 (16.2 - 26.5)	27.8 (21.6 - 41.4)*
Con Insulin Per Day	6.3 (5 - 8)	6.3 (5.1 - 7.8)	8.4 (6.9 - 10)
Bolus Per Day	22.1 (16.5 - 26.7)	21.5 (16.7 - 26.4)	10.1 (7.8 - 14)
TDI	28.5 (23.1 - 32.3)	27.5 (24.4 - 33.1)	19.1 (15.5 - 23)

[1]HAID SBC MCHO = Hybrid automatic insulin delivery (closed-loop) with standard bolus calculator and CHO misestimation.

[2]HAID DRL = Hybrid automatic insulin delivery (closed-loop) with proposed DRL bolus calculator.

[3]FAID = Fully automatic insulin delivery with proposed DRL bolus calculator.

$\star$ p value < 0.01. The p values (FAID vs HAID SBC MCHO) are based on the Wilcoxon signed-rank test.

The performance of the FAID system is coupled with the accuracy of the meal detector and the time duration of detection. The performance metrics of the meal detector are presented in table 2, which summarizes the populational detection performance of meals. The detection of lunch and dinner was better, as evidenced by sensitivity and true positives, whereas the snacks were barely detected. The detection of breakfast was approximately 60%. The time taken to detect a meal ranged between 30 and 40 minutes. As reported in table 2 false positives (FPs) amounted to fewer than 1 meal in the cases of breakfast, lunch, and snacks, and none resulted in a

hypoglycemic event. However, in the case of dinner, this number is approximately 2.4 meals, and a total of 8 hypoglycemic events were observed.

Table 2 Performance metrics of the meal detector.

Sensitivity (%)	Detection Time (min)	TP	FP	FN
Breakfast				
57.74 ± 14.43	37.92 ± 2.34	8.08 ± 2.02	0.67 ± 0.78	5.92 ± 2.02
57.14 (35.71 - 84.29)	38.75 (35 - 40)	8 (5 - 11.8)	0.5 (0 - 2)	6 (2.2 - 9)
Lunch				
95.24 ± 5.56	35 ± 0	13.33 ± 0.78	0.67 ± 0.98	0.67 ± 0.78
96.43 (85.71 - 100)	35 (35 - 35)	13.5 (12 - 14)	0 (0 - 2.9)	0.5 (0 - 2)
Snacks				
8.33 ± 5.96	29.17 ± 17.88	1.17 ± 0.83	0.33 ± 0.49	12.83 ± 0.83
7.14 (0 - 14.29)	37.5 (0 - 44.75)	1 (0 - 2)	0 (0 - 1)	13 (12 - 14)
Dinner				
95.83 ± 4.78	34.38 ± 1.55	13.42 ± 0.67	2.42 ± 1.31	0.58 ± 0.67
96.43 (86.43 - 100)	35 (30.25 - 35)	13.5 (12.1 - 14)	2 (0.1 - 4.9)	0.5 (0 - 1.9)

Values reported as mean ± standard deviation and median (25% - 75%).
Abbreviations: TP, true positive; FP, false positive; FN, false negative.

To exemplify the performance of the approach, the four-hour postprandial BG curves for each meal are illustrated in figures 5, 6, and 7. The BG followed a similar trajectory in all three cases. The postprandial peak BG values were higher in the case of the FAID system, reflecting the 30 to 40 minute delay in the delivery of the insulin bolus as a consequence of meal detection. The populational values of the meal detection time in minutes are represented by filled circles (pink) in the case of the FAID. Points on top of each other represent meals on different days with the same time of detection, whereas points along the x-axis represent meals with different times of detection. The time of detection is represented by the x-axis in minutes, with the meal appearing at $t = 0$.

4 Discussion

Several attempts have been made in the pursuit of a reliable FAID system. A learning-MPC algorithm was validated in an inpatient clinical study for a single unannounced meal in 29 patients with DM1[46]. No severe hypoglycemia was recorded, and it was suggested to extend the time of clinical trials and the number of unannounced meals in a future study. Analysis of the initial safety and efficacy of a FAID system based on a multiple-model probabilistic controller was presented for patients with DM1 [47]. Thirty hours of inpatient study in 10 patients and 54 hours of supervised hotel study in 15 patients were performed, challenging the controller with unannounced meals. It was concluded that there exists a greater risk of hypoglycemia compared to that of the HAID algorithms. A meal detection and estimation module was presented, relying on the fuzzy logic algorithm [48]. The algorithm was evaluated in a retrospective study for a total of 117 meals and 11 patients. The percentage of FPs reported was 20.8%. The detector was integrated with the AP system, but the calculation of insulin bolus was also dependent on the patient's CR. In a more recent study, an internal model

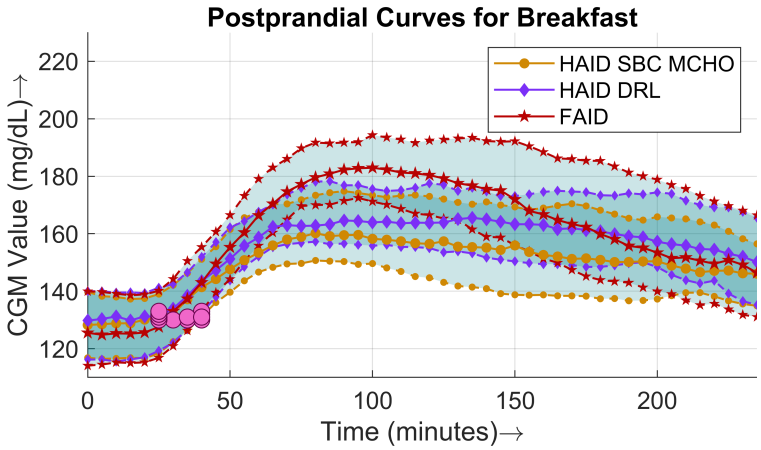

Fig. 5 Four-hour postprandial BG curves for breakfast. The solid lines represent median values, whereas the dotted lines correspond to the interquartile range. The filled circles are points where meals were detected, plotted against the time of detection in minutes in the case of the FAID system.

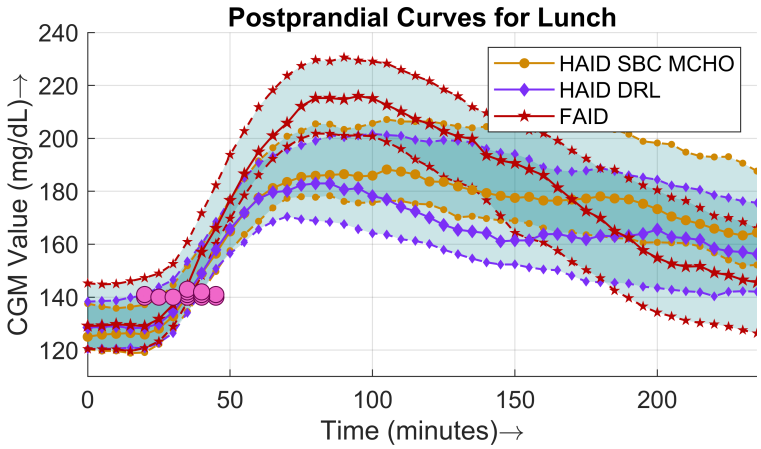

Fig. 6 Four-hour postprandial BG curves for lunch. The solid lines represent median values, whereas the dotted lines correspond to the interquartile range. The filled circles are points where meals were detected, plotted against the time of detection in minutes in the case of the FAID system.

control approach was used to derive a feedback controller for the FAID system and was tested in the UVa/Padova DM1 simulator. The outcome was presented in terms of the CGM curve and compared with open-loop therapy, and it was reported that the postprandial peak was reduced by approximately 8% [49].

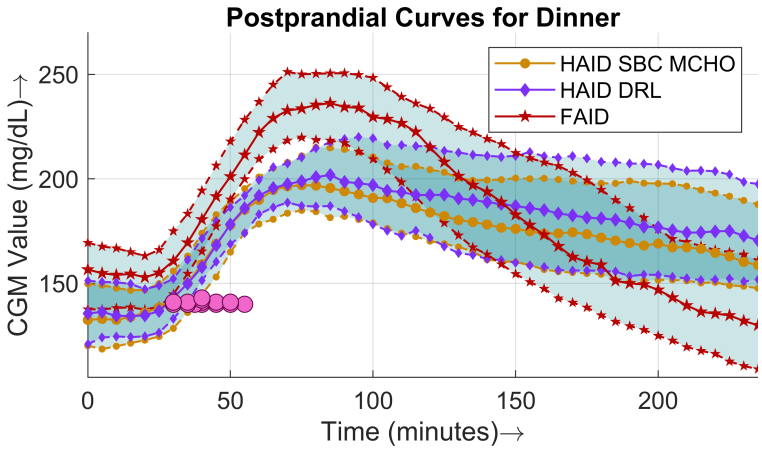

Fig. 7 Four-hour postprandial BG curves for dinner. The solid lines represent median values, whereas the dotted lines correspond to the interquartile range. The filled circles are points where meals were detected, plotted against the time of detection in minutes in the case of the FAID system.

In this work, a FAID system is proposed to compensate for meal disturbances by utilizing a DRL insulin bolus calculator. Three core components were integrated to implement the FAID system, i.e., a closed-loop PD controller for continuous insulin delivery, a detection algorithm for meal disturbances, and the DRL-based insulin bolus calculator. The proposed system can also accommodate announced meals without knowing the CHO content, unlike the methodologies presented in the literature. In such cases, the insulin bolus calculator is fed by meal announcement instead of the meal detector.

The primary CGM metrics are presented in table 1. CHO misestimation is included in the HAID with SBC to depict a real-life scenario. The absolute CGM values (mean, median, and maximum) are similar, whereas the minimum CGM is lower in the case of the FAID system because the insulin bolus calculation does not utilize CHO information and there is an inherent delay in bolus delivery due to the meal detection. The CV was slightly higher for the FAID system but was in the acceptable range of $< 35\%$. The GMI, an approximation of the A1C level based on the average BG from CGM [50], was similar in all cases.

The PCGM in the tight target range ($70 - 140$ mg/dL) was similar, and that in the target range ($70 - 180$ mg/dL) was lower by 5% in the FAID system. First, the PCGM in the range below 70% accounted for approximately 1% owing to the reasons mentioned above. Second, an increase was observed in the PCGM in the range above 180 mg/dL. This increase was induced by a delay in the bolus insulin delivery proportional to the meal detection duration. Moreover, a less aggressive dosing of bolus insulin, as reflected by greater penalties for hypoglycemia, also results in a lowering of PCGM in the target range ($70 - 180$ mg/dL). However, the PCGM in various target ranges presented for the FAID system is comparable to that of the HIAD systems.

The performance of the FAID system was coupled with the meal detector's accuracy and the delayed detection time. Greater accuracy and faster detection lead to better overall glycemic control performance of the FAID system. Thus, the performance metrics of the meal detector are presented in table 2. The detection of breakfast was better but had almost 40% false negatives (FNs). Lunch was very well detected and controlled as the amount of CHO in lunch was greater than that in breakfast or snacks. The snacks were rarely detected but were compensated for well by the closed-loop PD controller, suggesting that no feedforward compensation is needed for small meals. In the case of dinner, the detection was not desirable in terms of false positives, which may lead to nocturnal hypoglycemia, and a total of 8 hypoglycemic events were reported. This was one of the main reasons for lower CGM values in the case of the FAID system compared to the HAID systems. Regardless, the overall performance of the FAID system was satisfactory and comparable to that of the baseline HAID system given that the patient's intervention for meals was avoided successfully.

There is a trade-off in that adjusting the sensitivity of the meal detector to minimize FNs may also increase FPs, which are more dangerous. Based on two parameters of meals, the FNs of the meal detector were compensated well by the closed-loop PD controller. First, the dynamics and appearance of CHO in BG were considered, i.e., meals having slow dynamics and rate of appearance. Second, the amount of CHO in meals, i.e., meals with minimal CHO content such as snacks, was accounted for. In the abovementioned cases, when the meals were not detected, the disturbance was compensated for by the closed-loop controller, resulting in overall satisfactory performance, as evident from table 2. Therefore, it is not necessary to detect all meals in the presence of a closed-loop controller.

A comparison of the postprandial performance is explicitly presented in terms of populational postprandial BG curves for the three major meals in figures 5, 6, and 7. For all three meals, a similar pattern was observed, i.e., the peak was higher and the slope of the BG dip was steeper in the case of the FAID system as a consequence of the delay in insulin bolus delivery. Despite the steeper slope of the BG dip, there was no risk of severe hypoglycemic events owing to the higher peaks in the postprandial period. It is evident from the figures that the postprandial performance of the FAID system is comparable to that of the HAID systems and follows similar behavior.

The phenomenon of exploration in the DRL algorithm is not a safety concern in the in-silico trials. However, in clinical settings, it can be dangerous, for example, the management of DM1 without taking into account safety constraints [51]. A four-step approach suggested in [28] can be followed to move from in-silico to clinical trials. The improvement in policy during the training session is presented in terms of the total number of hypoglycemia events in figure 8. Each point in the plot represents a median of the number of hypoglycemia events per day for all patients for 25 days. A window of 25 days was selected to highlight the trend in the number of hypoglycemia events as training progressed. During training, an epsilon greedy policy that consists of both exploitation and exploration was considered; therefore, the trend was not downward throughout, but the overall impact was. As is clear from table 1, the time spent in hypoglycemia was approximately 1% when the trained DRL agents were deployed.

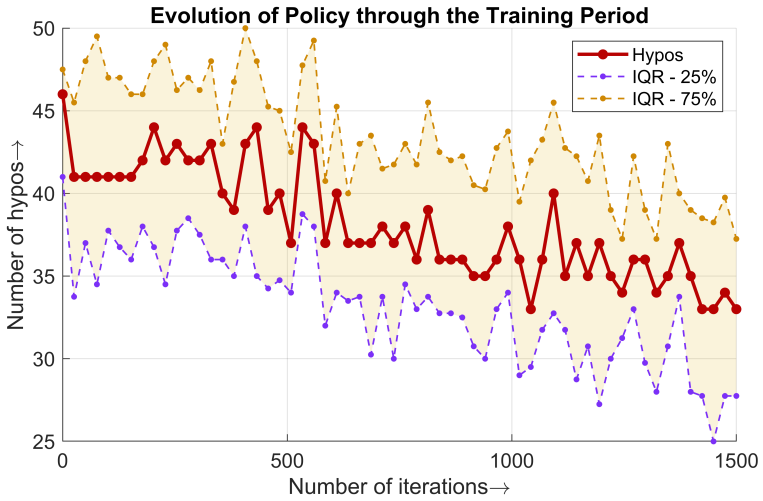

Fig. 8 Populational number of hypoglycemia events throughout the training period lasting for 1500 iterations. An epsilon greedy policy was followed for the purpose of training.

The main limitation of this study was the implementation of the FAID system in a virtual environment, as clinical settings would be more challenging owing to uncertain conditions in real-life scenarios. However, a customized virtual cohort was considered. Second, the dependency of the FAID system's performance on the meal detection algorithm limits this research. Despite having a suitable DRL insulin bolus calculator, the poor detection of unannounced meals may degrade the overall glycemic performance.

5 Conclusions

In this paper, a new machine learning-based FAID system was presented by integrating a closed-loop PD controller, a UKF-based meal detector, and a DRL-driven insulin bolus calculator. The proposed DRL algorithm was based on DQN and the feature of memory replay to calculate the insulin bolus without requiring information regarding CHO content, CR, and CF, thereby paving the way for the elimination of meal announcement.

The proposed FAID system showed similar performance to that of the HAID system, without a significant increase in hypoglycemia. The main objective of the FAID system is to eliminate patient intervention in the closed-loop system to avoid errors caused by CHO misestimation and to relieve the unnecessary burden on patients of calculating the CHO content.

Future research will include the use of a more accurate meal detector to minimize the effect of false positives and false negatives on the overall glycemic performance of

the FAID system. Furthermore, the use of more advanced DRL algorithms will boost the performance, enabling the FAID system to compete with HAID systems.

Declarations

Ethical Approval. Not applicable.

Availability of data and materials. Not applicable.

Competing Interests. The authors declare no competing interests.

Funding. This work was partially supported by the Spanish Ministry of Science and Innovation under Grant numbers PID2019-107722RB-C22 and PDC2021-121470-C22, by the Autonomous Government of Catalonia under Grant number 2021 SGR 01598 and by the program for researchers grant 2019 FI_B 01200.

References

[1] Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B.J., Jacobsen, L.M., Schatz, D.A., Lernmark, Å.: Type 1 diabetes mellitus. Nature reviews Disease primers **3**(1), 1–17 (2017)

[2] Iqbal, A., Novodvorsky, P., Heller, S.R.: Recent updates on type 1 diabetes mellitus management for clinicians. Diabetes & metabolism journal **42**(1), 3–18 (2018)

[3] Beck, R.W., Bergenstal, R.M., Laffel, L.M., Pickup, J.C.: Advances in technology for management of type 1 diabetes. The Lancet **394**(10205), 1265–1273 (2019)

[4] Rodbard, D.: Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. Diabetes technology & therapeutics **19**(S3), 25 (2017)

[5] Calhoun, P.M., Buckingham, B.A., Maahs, D.M., Hramiak, I., Wilson, D.M., Aye, T., Clinton, P., Chase, P., Messer, L., Kollman, C., *et al.*: Efficacy of an overnight predictive low-glucose suspend system in relation to hypoglycemia risk factors in youth and adults with type 1 diabetes. Journal of diabetes science and technology **10**(6), 1216–1221 (2016)

[6] Tsoukas, M.A., Majdpour, D., Yale, J.-F., El Fathi, A., Garfield, N., Rutkowski, J., Rene, J., Legault, L., Haidar, A.: A fully artificial pancreas versus a hybrid artificial pancreas for type 1 diabetes: a single-centre, open-label, randomised controlled, crossover, non-inferiority trial. The Lancet Digital Health **3**(11), 723–732 (2021)

[7] Leelarathna, L., Choudhary, P., Wilmot, E.G., Lumb, A., Street, T., Kar, P., Ng, S.M.: Hybrid closed-loop therapy: Where are we in 2021? Diabetes, Obesity and Metabolism **23**(3), 655–660 (2021)

[8] Bekiari, E., Kitsios, K., Thabit, H., Tauschmann, M., Athanasiadou, E., Karagiannis, T., Haidich, A.-B., Hovorka, R., Tsapas, A.: Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis. bmj **361** (2018)

[9] Reiterer, F., Freckmann, G., Re, L.: Impact of carbohydrate counting errors on glycemic control in type 1 diabetes. IFAC-PapersOnLine **51**(27), 186–191 (2018)

[10] Brazeau, A., Mircescu, H., Desjardins, K., Leroux, C., Strychar, I., Ekoé, J., Rabasa-Lhoret, R.: Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. Diabetes research and clinical practice **99**(1), 19–23 (2013)

[11] Lawton, J., Blackburn, M., Rankin, D., Allen, J., Campbell, F., Leelarathna, L., Tauschmann, M., Thabit, H., Wilinska, M., Hovorka, R., *et al.*: The impact of using a closed-loop system on food choices and eating practices among people with type 1 diabetes: a qualitative study involving adults, teenagers and parents. Diabetic Medicine **36**(6), 753–760 (2019)

[12] Mehta, S.N., Haynie, D.L., Higgins, L.A., Bucey, N.N., Rovner, A.J., Volkening, L.K., Nansel, T.R., Laffel, L.M.: Emphasis on carbohydrates may negatively influence dietary patterns in youth with type 1 diabetes. Diabetes Care **32**(12), 2174–2176 (2009)

[13] Ware, J., Hovorka, R.: Recent advances in closed-loop insulin delivery. Metabolism **127**, 154953 (2022)

[14] Samadi, S., Turksoy, K., Hajizadeh, I., Feng, J., Sevil, M., Cinar, A.: Meal detection and carbohydrate estimation using continuous glucose sensor data. IEEE journal of biomedical and health informatics **21**(3), 619–627 (2017)

[15] Turksoy, K., Samadi, S., Feng, J., Littlejohn, E., Quinn, L., Cinar, A.: Meal detection in patients with type 1 diabetes: a new module for the multivariable adaptive artificial pancreas control system. IEEE journal of biomedical and health informatics **20**(1), 47–54 (2015)

[16] Dassau, E., Bequette, B.W., Buckingham, B.A., Doyle III, F.J.: Detection of a meal using continuous glucose monitoring: implications for an artificial β-cell. Diabetes care **31**(2), 295–300 (2008)

[17] Meneghetti, L., Facchinetti, A., Del Favero, S.: Model-based detection and classification of insulin pump faults and missed meal announcements in artificial pancreas systems for type 1 diabetes therapy. IEEE Transactions on Biomedical Engineering **68**(1), 170–180 (2020)

[18] Harvey, R.A., Dassau, E., Zisser, H., Seborg, D.E., Doyle III, F.J.: Design of the glucose rate increase detector: a meal detection module for the health monitoring

system. Journal of diabetes science and technology **8**(2), 307–320 (2014)

[19] Mosquera-Lopez, C., Wilson, L.M., El Youssef, J., Hilts, W., Leitschuh, J., Branigan, D., Gabo, V., Eom, J.H., Castle, J.R., Jacobs, P.G.: Enabling fully automated insulin delivery through meal detection and size estimation using artificial intelligence. npj Digital Medicine **6**(1), 39 (2023)

[20] Fushimi, E., Colmegna, P., De Battista, H., Garelli, F., Sánchez-Peña, R.: Artificial pancreas: Evaluating the arg algorithm without meal announcement. Journal of Diabetes Science and Technology **13**(6), 1035–1043 (2019)

[21] Sanz, R., García, P., Díez, J.-L., Bondia, J.: Artificial pancreas system with unannounced meals based on a disturbance observer and feedforward compensation. IEEE Transactions on Control Systems Technology **29**(1), 454–460 (2020)

[22] Garcia-Tirado, J., Lv, D., Corbett, J.P., Colmegna, P., Breton, M.D.: Advanced hybrid artificial pancreas system improves on unannounced meal response-in silico comparison to currently available system. Computer Methods and Programs in Biomedicine **211**, 106401 (2021)

[23] Diamond, T., Cameron, F., Bequette, B.W.: A new meal absorption model for artificial pancreas systems. Journal of Diabetes Science and Technology **16**(1), 40–51 (2022)

[24] Tejedor, M., Woldaregay, A.Z., Godtliebsen, F.: Reinforcement learning application in diabetes blood glucose control: A systematic review. Artificial intelligence in medicine **104**, 101836 (2020)

[25] Sun, Q., Jankovic, M.V., Budzinski, J., Moore, B., Diem, P., Stettler, C., Mougiakakou, S.G.: A dual mode adaptive basal-bolus advisor based on reinforcement learning. IEEE journal of biomedical and health informatics **23**(6), 2633–2641 (2018)

[26] Zhu, T., Li, K., Kuang, L., Herrero, P., Georgiou, P.: An insulin bolus advisor for type 1 diabetes using deep reinforcement learning. Sensors **20**(18), 5058 (2020)

[27] Jafar, A., El Fathi, A., Haidar, A.: Long-term use of the hybrid artificial pancreas by adjusting carbohydrate ratios and programmed basal rate: A reinforcement learning approach. Computer Methods and Programs in Biomedicine **200**, 105936 (2021)

[28] Ahmad, S., Beneyto, A., Contreras, I., Vehi, J.: Bolus insulin calculation without meal information. a reinforcement learning approach. Artificial Intelligence in Medicine **134**, 102436 (2022)

[29] Beneyto, A., Bertachi, A., Bondia, J., Vehi, J.: A new blood glucose control scheme for unannounced exercise in type 1 diabetic subjects. IEEE Transactions

on Control Systems Technology **28**(2), 593–600 (2018)

[30] Ramkissoon, C.M., Herrero, P., Bondia, J., Vehi, J.: Unannounced meals in the artificial pancreas: detection using continuous glucose monitoring. Sensors **18**(3), 884 (2018)

[31] Hovorka, R., Canonico, V., Chassin, L.J., Haueter, U., Massi-Benedetti, M., Federici, M.O., Pieber, T.R., Schaller, H.C., Schaupp, L., Vering, T., *et al.*: Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiological measurement **25**(4), 905 (2004)

[32] Revert, A., Garelli, F., Picó, J., De Battista, H., Rossetti, P., Vehí, J., Bondia, J.: Safety auxiliary feedback element for the artificial pancreas in type 1 diabetes. IEEE transactions on Biomedical Engineering **60**(8), 2113–2122 (2013)

[33] Bergman, R.N., Phillips, L.S., Cobelli, C., *et al.*: Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. The Journal of clinical investigation **68**(6), 1456–1467 (1981)

[34] Facchinetti, A., Sparacino, G., Cobelli, C.: Enhanced accuracy of continuous glucose monitoring by online extended kalman filtering. Diabetes technology & therapeutics **12**(5), 353–363 (2010)

[35] Larsen, J.: Correlation functions and power spectra. Section for Cognitive Systems, Informatics and Mathematical Modelling (2009)

[36] Watkins, C.J., Dayan, P.: Q-learning. Machine learning **8**, 279–292 (1992)

[37] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., *et al.*: Human-level control through deep reinforcement learning. nature **518**(7540), 529–533 (2015)

[38] Wilinska, M.E., Chassin, L.J., Schaller, H.C., Schaupp, L., Pieber, T.R., Hovorka, R.: Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Transactions on Biomedical Engineering **52**(1), 3–12 (2004)

[39] Ahmad, S., Ramkissoon, C.M., Beneyto, A., Conget, I., Giménez, M., Vehi, J.: Generation of virtual patient populations that represent real type 1 diabetes cohorts. Mathematics **9**(11), 1200 (2021)

[40] Visentin, R., Dalla Man, C., Kudva, Y.C., Basu, A., Cobelli, C.: Circadian variability of insulin sensitivity: physiological input for in silico artificial pancreas. Diabetes technology & therapeutics **17**(1), 1–7 (2015)

[41] Roversi, C., Vettoretti, M., Del Favero, S., Facchinetti, A., Choudhary, P., Sparacino, G.: Impact of carbohydrate counting error on glycemic control in open-loop

management of type 1 diabetes: quantitative assessment through an in silico trial. Journal of Diabetes Science and Technology **16**(6), 1541–1549 (2022)

[42] Walsh, J., Roberts, R.: Pumping Insulin: Everything You Need for Success on a Smart Insulin Pump vol. 4. Torrey Pines Press San Diego, CA, ??? (2006)

[43] Zisser, H., Robinson, L., Bevier, W., Dassau, E., Ellingsen, C., Doyle III, F.J., Jovanovic, L.: Bolus calculator: a review of four "smart" insulin pumps. Diabetes technology & therapeutics **10**(6), 441–444 (2008)

[44] Holt, R.I., DeVries, J.H., Hess-Fischl, A., Hirsch, I.B., Kirkman, M.S., Klupa, T., Ludwig, B., Nørgaard, K., Pettus, J., Renard, E., *et al.*: The management of type 1 diabetes in adults. a consensus report by the american diabetes association (ada) and the european association for the study of diabetes (easd). Diabetes Care **44**(11), 2589–2625 (2021)

[45] Klonoff, D.C., Wang, J., Rodbard, D., Kohn, M.A., Li, C., Liepmann, D., Kerr, D., Ahn, D., Peters, A.L., Umpierrez, G.E., et al.: A glycemia risk index (gri) of hypoglycemia and hyperglycemia for continuous glucose monitoring validated by clinician ratings. Journal of diabetes science and technology, 19322968221085273 (2022)

[46] Song, L., Liu, C., Yang, W., Zhang, J., Kong, X., Zhang, B., Chen, X., Wang, N., Shen, D., Li, Z., *et al.*: Glucose outcomes of a learning-type artificial pancreas with an unannounced meal in type 1 diabetes. Computer methods and programs in biomedicine **191**, 105416 (2020)

[47] Cameron, F.M., Ly, T.T., Buckingham, B.A., Maahs, D.M., Forlenza, G.P., Levy, C.J., Lam, D., Clinton, P., Messer, L.H., Westfall, E., *et al.*: Closed-loop control without meal announcement in type 1 diabetes. Diabetes technology & therapeutics **19**(9), 527–532 (2017)

[48] Samadi, S., Rashid, M., Turksoy, K., Feng, J., Hajizadeh, I., Hobbs, N., Lazaro, C., Sevil, M., Littlejohn, E., Cinar, A.: Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system. Diabetes technology & therapeutics **20**(3), 235–246 (2018)

[49] Sanz, R., García, P., Romero-Vivó, S., Díez, J., Bondia, J.: Near-optimal feedback control for postprandial glucose regulation in type 1 diabetes. ISA Transactions (2022)

[50] Bergenstal, R.M., Beck, R.W., Close, K.L., Grunberger, G., Sacks, D.B., Kowalski, A., Brown, A.S., Heinemann, L., Aleppo, G., Ryan, D.B., *et al.*: Glucose management indicator (gmi): a new term for estimating a1c from continuous glucose monitoring. Diabetes care **41**(11), 2275–2280 (2018)

[51] Yau, K.-L.A., Chong, Y.-W., Fan, X., Wu, C., Saleem, Y., Lim, P.-C.: Reinforcement learning models and algorithms for diabetes management. IEEE Access **11**, 28391–28415 (2023)

CHAPTER 3

DISCUSSION

RL algorithms have been proposed in this thesis to calculate the insulin bolus in people with T1D with a focus on developing a FAID system and a realistic cohort of VPs has been generated. Firstly, a methodology was developed for the generation of a realistic cohort of VPs that represents a cohort of people with T1D from Hospital Clinic de Barcelona. The clinical data of insulin and CGM was explicitly utilized in generating a virtual cohort. The methodology was developed to replicate the standard glycemic metrics of the real cohort to provide a realistic validation environment for the developed RL algorithms. Secondly, an RL algorithm based on Q-Learning was developed for the calculation of insulin bolus without taking into account meal information. The algorithm exploited CGM data to calculate insulin bolus and a comparison was made with the SBC. Finally, a FAID was developed with an advanced version of the RL algorithm based on deep Q networks. The proposed FAID system consists of a PD controller, a meal detector and a insulin bolus calculator based on RL algorithm. The following sections of this chapter provide a summary of the presented work and discuss the advantages as well as potential limitations of the proposed approaches.

3.1 Generation of T1D Virtual Cohort

In (Ahmad et al., 2021) we presented a methodology to generate a realistic cohort of subjects that resembles the glycemic metrics of a cohort of patients from the Hospital Clinic de Barcelona. The gathered clinical data was utilized to estimate the model parameters in a day-by-day fashion to achieve realistic glycemic variability in the in-silico simulations. The generated cohort has been used for the in-silico validation of the proposed RL algorithms.

There exist several methodologies in the literature to generate a large cohort of VPs with T1D. (Haidar et al., 2013) proposed a stochastic e-cloning concept for the generation of VPs for the assessment of CL controllers. A Bayesian approach was adopted to estimate the model parameters and a Monte Carlo methodology was used to obtain the probability distributions of the parameters. Clinical data from 12 young T1D patients was used to test the methodology. (Resalat et al., 2019) proposed a statistical method to generate a population of VPs for both single and double hormone AP systems. The clinical data of 20 patients undergoing AP trials was used for evaluation. The glycemic metrics used for comparison were percent of time BG in target range (70 to 180 mg/dL), hyperglycemia (> 180 mg/dL), and hypoglycemia (< 70mg/dL). (Orozco-López et al., 2020) proposed a methodology to generate a large cohort of VPs by establishing a relationship between the subject's parameters in terms of covariance utilizing the Hovorka model. These methodologies are based on distributions, statistical and stochastic estimation of parameters providing generic virtual cohorts of people with T1D. Contrarily, our approach is focused on generating a virtual cohort that represents a specific group of people with T1D. In this study we have considered 14 T1D subjects from the Hospital Clinic de Barcelona characterized by a relatively high coefficient of variation (CV) in their BG profile that are usually excluded from clinical trials.

The proposed methodology is based on the FDA approved UVa/Padova simulator. The academic version of UVa/Padova simulator consists of 3 groups of 10 VPs (children, adolescents, and adults). Additionally, 1 subject of each gruop represents an average T1D subject (Torrent-

Fontbona and López, 2018). The gathered clinical data from the electronic medical records of patients at the Diabetes Unit, Endocrinology and Nutrition Department from the Hospital Clínic of Barcelona include CHO content in meal, meal time and basal and bolus insulin dosage. The patients were using 640G Medtronic Minimed insluin pump system. Exercise sessions were included as disturbances and were derived from BG profile of the real patients. CGM data of 14 days from a real patient with at least 70% of valid data as recommended by (Battelino et al., 2019) was considered for the generation of VPs. The two model parameters i.e., sensitivity of insulin on glucose utilization and basal endogenous glucose production were modified in a day-by-day fashion to capture daily BG profiles of the real patients. Against each real patient, 10 VPs were generated as per the adult cohort of UVa/Padova simulator.

After the generation of the VPs, the VPs showing unrealistic BG profiles (BG values above 500 mg/dL or below 30 mg/dL) were discarded. Additionally, VPs with two times the standard deviation of saturation points (400 mg/dL and 40 mg/dL) as compared to the clinical data were excluded from the final cohort. Finally, the initial cohort of 140 VPs was reduced to a total of 75 VPs that were able to show statistically similar CGM metrics as compared to their real patients counterparts. The proposed methodology was able to generate valid VPs for all of the real subjects.

The major limitation of the proposed methodology is using as a baseline the adult UVa/Padova cohort to replicate the BG profile of each single real subject. This resulted in a high VP discard ratio of 46% (65 out of the 140 patients were removed). However, as a VP in UVa/Padova is defined by a set of distinct parameters; the inter-patient variability was still retained. Moreover, the generated VPs do not exactly replicate the BG variability (stress, illness, lifestyle, etc.) of real patients because of their unpredictable nature.

The proposed methodology can be used for the evaluation of therapeutic strategies developed for the treatment of people with T1D. Specifically, individuals with different characteristics as demonstrated in this work a real cohort that is usually excluded from clinical trials due to

higher CV. Moreover, the generated VPs provide more realistic and challenging scenarios for in-silico validation unlike the methodologies mentioned in literature.

3.2 Reinforcement Learning for Insulin Bolus Dosing

We have proposed an RL insulin bolus dosing algorithm (Ahmad et al., 2022) that does not require the amount of CHO in meals nor patient specific parameters such as CR and CF. Firstly, The goal was to eliminate the estimation of CHO amount in meals by patients to avoid CHO counting errors. Secondly, to release the CHO management burden on patients.

Currently available RL powered insulin bolus infusion techniques for pumps require entry of CHO estimates each time a meal is taken (Sun et al., 2018; Zhu, Li, Herrero and Georgiou, 2020; Jafar et al., 2021). This strategy has two significant drawbacks. First, these approaches are prone to inaccuracies since it is difficult to estimate the CHO content of a meal correctly, leading to the frequent misestimation by people with T1D. Second, patients face management burden every time they eat a meal as a result of calculating CHO. Moreover, it negatively influences the quality of life in people with T1D and makes them less confident while interacting with peers, especially around food (Lawton et al., 2019). On the other hand, the proposed RL algorithm only relies on CGM data for calculation of insulin bolus and eliminates the risk of CHO counting errors in people with T1D.

We proposed an RL algorithm based on a Q-Learning approach with a multiple action spaces strategy to avoid the need of CHO information and solely utilize CGM data. The classification of the action space into multiple sub spaces was based on the CGM value before the meal. The insulin bolus doses in sub spaces were defined in relation to the CGM value before the meal. The results were compared to the SBC as a baseline and a statistical comparison was made by providing the p-values in the results. We have also demonstrated SBC with CHO misestimation to show the real-life situation and the proposed algorithm outperformed the SBC with CHO misestimation.

The main limitation of this study is that the proposed RL algorithm is not deployed and tested in clinical settings. In real life scenarios, the trained RL agent may face new situations and it can be dangerous. Therefore, the RL agent should observe all the possible states during the training phase. This limitation also results in training for longer periods during the in-silico simulations. Moreover, a road-map of implementation in clinical setting was also provided in our study comprised of four steps as follows:

1. Generation of a virtual cohort that represents the target cohort of people with T1D.
2. Training of RL algorithm for the generated cohort.
3. Pre-clinical validation of the algorithm with utilizing CGM data from real patients for insulin dosing.
4. Implementaion of the personalized trained RL algorithm in clinical settings with safety measures.

3.3 Fully Automatic Insulin Delivery System

Finally, we have proposed a FAID system by integrating a PD CL controller for continuous insulin delivery, an UKF based meal detector for unannounced meals and an advanced version of the insulin bolus calculator discussed in the section 3.2. Evaluation of the FAID was performed on the virtual cohort of people with T1D discussed in the section 3.1.

FAID is the most recent advancement in the management of T1D. It is aimed to achieve the glycemic targets without increasing the risk of hypoglycemia to improve the quality of life by removing meal announcements from the daily activities of people with T1D (Giménez et al., 2021). FAID systems are expected to be in clinical practice in future but at present only hybrid AID systems are commercially available. In hybrid AID (HAID) system, patients are required to count the CHO and manage the insulin bolus at meal times. The clinical adoption of FAIDs remain in early stages despite of recent technological advances (Nwokolo and Hovorka, 2023).

Researchers have made attempts to develop FAID systems in the literature. In early attempts, (Wang et al., 2009) proposed an MPC based system that can learn from the patient's lifestyle and a model called an autoregressive exogenous model was utilized for control design. Meals were considered as disturbances in model during the design of controller. A pilot clinical study of this learning based AP was reported for an adult cohort of 10 patients (Wang et al., 2017). The AP was also evaluated in a 4 hours inpatient open-label study for a cohort of 29 patients with T1D with unannounced meal (Song et al., 2020). The convergence time of the presented system was less than 10 days which is a significant amount of time in real life scenarios. In (Cameron et al., 2017) an assessment of safety and efficacy of an MPC-based AP system was performed for patients with T1D and unannounced meals. Data from the National Health and Nutrition Examination Survey and the American Time Survey was used to anticipate future meals and compensate for unannounced meals. It was concluded that there exists more risk of hypoglycemia as compared to the algorithms relying on meal announcements. A methodology was proposed by (Samadi et al., 2018) for the detection and estimation of the amount of CHO in unannounced meals utilizing the CGM data and insulin delivery data with a focus on AP system. The percentages of detection reported for meals and snacks were 93.5% and 68% respectively. The integration of meal detector and CHO content estimator into an AP system was reported in future study recommendations (Samadi et al., 2018).

In this work, we proposed an advanced version of the insulin bolus calculator published (Ahmad et al., 2022) and discussed in section 3.2. The classical Q-Learning RL algorithm was replaced with a DRL algorithm called Deep Q Network (DQN). The major advantage is that DQN use continuous states allowing artificial neural network to learn hidden patterns in the CGM data to make decision for insulin bolus whereas Q-Learning is a tabular methodology based on discrete states. The continuous insulin delivery rate was modulated and infused by a closed-loop PD controller with a safety auxiliary feedback element (SAFE) introduced in (Beneyto et al., 2018). The detection of meals was based on an in-house algorithm utilizing

an augmented minimal model and UKF along with the insulin and CGM data (Ramkissoon et al., 2018). Integration of the DRL bolus calculator, PD closed-loop controller and UKF meal detector was performed for implementation of the FAID.

The performance of the FAID system was compared to the HAID system. The outcomes of the in-silico trials were reported in terms of CGM metrics as recommended for clinical practice by the American Diabetes Association (ADA), the European Association for the Study of Diabetes (EASD) (Holt et al., 2021) and in the international consensus conference (Aleppo, 2021). The overall performance of the FAID reported was comparable to HAID.

The main limitation of this study was the evaluation of the FAID in in-silico trials. Real-life scenarios could be more challenging due to uncertain conditions and unmodeled dynamics. However, a customized virtual cohort was considered in this study as discussed in the section 3.1. Secondly, the performance of the FAID was dependent on the accuracy and response of the meal detection algorithm. This implies that despite having a suitable DRL insulin bolus calculator the poor detection of unannounced meals may degrade the overall glycemic performance.

CHAPTER 4

CONCLUSIONS

This thesis proposed a DRL powered insulin bolus dosing system that does not require the information of CHO content in meals. This algorithm has been integrated with a PD CL controller and a meal detector to build a FAID system. Additionally, the developed insulin bolus calculator does not require patient specific parameters i.e., CR and CF. This work has also drawn a comparison between a RL based insulin bolus calculator and the SBC. This work has also contributed in developing a novel methodology to generate a cohort of VPs focused on a specific group of people with T1D. The parameters in VPs have been varied in a day-to-day fashion to capture the daily changes in BG profile in real patients with T1D. The generated virtual cohort of VPs has been used for the validation of the developed insulin bolus calculator and the FAID.

4.1 Contributions

The mentioned contributions can be detailed further into more particular ones that have been achieved during the development of the thesis:

- **Method for the generation of VPs.** This study proposed a novel algorithm to generate a virtual cohort of people with T1D. The parameters reflecting insulin sensitivity and endogenous glucose production were optimized in a day-to-day manner to achieve the BG profile reflected by the clinical data. The clinical data was taken into account to modify the parameters, resulting in more realistic BG outcomes in terms of the generated VPs.The proposed algorithm has been developed for testing the controllers and therapeutic strategies developed for the management of people with T1D. Finally, it is based on the FDA-approved UVa/Padova simulator and can be utilized for the pre-clinical validation.

- **Novel insulin bolus calculator based on RL.** This thesis introduced an insulin bolus calculator independent of CHO estimation. To achieve this, a RL based insulin bolus dosing system was designed. The proposed algorithm utilized the Q-Learning methodology of RL. The main challenge faced was calculating insulin bolus with no information about the CHO content in meals. To solve this problem a multiple sub action spaces strategy was built based on the CGM value immediately before meal intake and the dosage was in accordance to the total daily insulin requirement of the patient. The proposed algorithm was compared to the SBC with and without CHO misestimation. Finally, the bolus insulin on board was also taken into account as a safety measure to minimize the occurrence of severe hypoglycemia.

- **Fully automatic insulin delivery system.** Finally, this thesis contributes a novel ML drvien FAID comprised of the integration of a CL PD controller, a UKF-based meal detector, and a DRL insulin bolus calculator. The DRL algorithm is based on DQNs and the feature of memory replay to calculate the insulin bolus without utilizing CHO content in meals and patient's CR, and CF thus, avoiding announcement of meals by patient. It was demonstrated that the FAID system showed comparable glycemic performance to that of the HAID without a significant increase in the severe hypoglycemia. The main

focus was to eliminate patient interventions in the CL system to eliminate the errors due to CHO misestimation and to relieve the management burden of the people with T1D.

4.2 Future Work

The studies presented in this thesis show encouraging results providing a realistic simulation platform for validating the strategies designed for the management of T1D. The results demonstrated for realizing the FAID are also in competition with the HIAD but additional research is needed to further improve the glycemic control performance specially in free living conditions. This section discusses potential future pathways that can be developed from the bases of the contributions presented in this work.

The methodology presented for the generation of virtual T1D cohort can be extended in two major directions. Firstly, two parameters were used for the modification of the VPs from the UVa/Padova simulator in the proposed approach. Other potential parameters could be used and optimized in a similar way to make the cohort more challenging and realistic. Secondly, other possibilities for modification pattern of the parameters can be explored. For example, modifying the parameters in an hourly fashion to capture the intraday variations in the glycemic profiles of real patients.

The proposed RL insulin bolus calculators and the FAID system exhibit one limitation in common and that was validation of the proposed methodologies in in-silico trials. The implementation in clinical settings could be challenging because of uncertain conditions in real-life scenarios. Five steps are suggested to follow for the evaluation of the proposed FAID in clinical trials in future study. 1) Generating a virtual cohort that exhibit the BG profile similar to the target cohort of people with T1D selected for the clinical trials. The methodology presented as a contribution in this work can be used for this purpose. 2) Train the DRL agents in a simulation environment as an initial step. 3) Perform the analysis of the meal detector on the clinical data from the real patients and tuning should be performed if required. 4)

Preclinical validation of the trained DRL agents under real-life scenarios with the clinical data from real patients. The CGM data of the real patients should be used by the DRL agents to recommend insulin bolus for meals in this phase. 5) Implementing the trained agents for real patients in clinical trials with safety precautions. The RL algorithm can be implemented on multiprocessing platforms for embedded applications such as Xilinx Zynq as reported in (Spano et al., 2019). CGM sensor can be interfaced with it and the entire FAID system need to be programmed in it. This could be a possible option to be the core of insulin pump but the hardware implementation possibilities for the FAID system need to be explored in future study.

BIBLIOGRAPHY

Ahmad, S., Beneyto, A., Contreras, I. and Vehi, J. (2022), 'Bolus insulin calculation without meal information. a reinforcement learning approach', *Artificial Intelligence in Medicine* **134**, 102436.

Ahmad, S., Ramkissoon, C. M., Beneyto, A., Conget, I., Giménez, M. and Vehi, J. (2021), 'Generation of virtual patient populations that represent real type 1 diabetes cohorts', *Mathematics* **9**(11), 1200.

Ahsan, M. M. and Siddique, Z. (2022), 'Machine learning-based heart disease diagnosis: A systematic literature review', *Artificial Intelligence in Medicine* p. 102289.

Aleppo, G. (2021), 'Clinical application of time in range and other metrics', *Diabetes Spectrum* **34**(2), 109–118.

Association, A. D. (2014), 'Diagnosis and classification of diabetes mellitus', *Diabetes care* **37**(Supplement_1), S81–S90.

Association, A. D. (2022), 'Introduction: Standards of medical care in diabetes—2022'.

Banting, F. G., Best, C. H., Collip, J. B., Campbell, W. R. and Fletcher, A. A. (1922), 'Pancreatic extracts in the treatment of diabetes mellitus', *Canadian Medical Association Journal* **12**(3), 141.

Battelino, T., Danne, T., Bergenstal, R. M., Amiel, S. A., Beck, R., Biester, T., Bosi, E., Buckingham, B. A., Cefalu, W. T., Close, K. L. et al. (2019), 'Clinical targets for continuous

glucose monitoring data interpretation: recommendations from the international consensus on time in range', *Diabetes care* **42**(8), 1593–1603.

Battineni, G., Sagaro, G. G., Chinatalapudi, N. and Amenta, F. (2020), 'Applications of machine learning predictive models in the chronic disease diagnosis', *Journal of personalized medicine* **10**(2), 21.

Beneyto, A., Bertachi, A., Bondia, J. and Vehi, J. (2018), 'A new blood glucose control scheme for unannounced exercise in type 1 diabetic subjects', *IEEE Transactions on Control Systems Technology* **28**(2), 593–600.

Boden, M. (1980), 'Artificial intelligence and natural man', *Synthese* **43**(3).

Boughton, C. K. and Hovorka, R. (2021), 'New closed-loop insulin systems', *Diabetologia* **64**(5), 1007–1015.

Brown, S. A., Kovatchev, B. P., Raghinaru, D., Lum, J. W., Buckingham, B. A., Kudva, Y. C., Laffel, L. M., Levy, C. J., Pinsker, J. E., Wadwa, R. P. et al. (2019), 'Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes', *New England Journal of Medicine* **381**(18), 1707–1717.

Cameron, F. M., Ly, T. T., Buckingham, B. A., Maahs, D. M., Forlenza, G. P., Levy, C. J., Lam, D., Clinton, P., Messer, L. H., Westfall, E. et al. (2017), 'Closed-loop control without meal announcement in type 1 diabetes', *Diabetes technology & therapeutics* **19**(9), 527–532.

Chaki, J., Ganesh, S. T., Cidham, S. and Theertan, S. A. (2022), 'Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review', *Journal of King Saud University-Computer and Information Sciences* **34**(6), 3204–3225.

Clemens, A. (1977), 'The development of biostator, a glucose controlled insulin infusion system (gciis)', *Horm Metab Res* **7**, 23–33.

Control, D. and Group, C. T. R. (1993), 'The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus', *New England journal of medicine* **329**(14), 977–986.

Daskalaki, E., Diem, P. and Mougiakakou, S. G. (2013), 'An actor–critic based controller for glucose regulation in type 1 diabetes', *Computer methods and programs in biomedicine* **109**(2), 116–125.

Daskalaki, E., Diem, P. and Mougiakakou, S. G. (2016), 'Model-free machine learning in biomedicine: Feasibility study in type 1 diabetes', *PloS one* **11**(7), e0158722.

De Paula, M., Acosta, G. G. and Martínez, E. C. (2015), 'On-line policy learning and adaptation for real-time personalization of an artificial pancreas', *Expert Systems with Applications* **42**(4), 2234–2255.

De Paula, M., Avila, L. O. and Martinez, E. C. (2015), 'Controlling blood glucose variability under uncertainty using reinforcement learning and gaussian processes', *Applied Soft Computing* **35**, 310–332.

DeFronzo, R. A., Ferrannini, E., Groop, L., Henry, R. R., Herman, W. H., Holst, J. J., Hu, F. B., Kahn, C. R., Raz, I., Shulman, G. I. et al. (2015), 'Type 2 diabetes mellitus', *Nature reviews Disease primers* **1**(1), 1–22.

Doyle III, F. J., Huyett, L. M., Lee, J. B., Zisser, H. C. and Dassau, E. (2014), 'Closed-loop artificial pancreas systems: engineering the algorithms', *Diabetes care* **37**(5), 1191–1197.

Emerson, H., Guy, M. and McConville, R. (2023), 'Offline reinforcement learning for safer blood glucose control in people with type 1 diabetes', *Journal of Biomedical Informatics* **142**, 104376.

Farrington, C. (2018), 'Psychosocial impacts of hybrid closed-loop systems in the management of diabetes: a review', *Diabetic Medicine* **35**(4), 436–449.

Garg, S. K., Weinzimer, S. A., Tamborlane, W. V., Buckingham, B. A., Bode, B. W., Bailey, T. S., Brazg, R. L., Ilany, J., Slover, R. H., Anderson, S. M. et al. (2017), 'Glucose outcomes with the in-home use of a hybrid closed-loop insulin delivery system in adolescents and adults with type 1 diabetes', *Diabetes technology & therapeutics* **19**(3), 155–163.

Giménez, M., Conget, I. and Oliver, N. (2021), 'Automated insulin delivery systems: today, tomorrow and user requirements', *Journal of diabetes science and technology* **15**(6), 1252–1257.

Göke, B. (2008), 'Islet cell function: α and β cells–partners towards normoglycaemia', *International Journal of clinical practice* **62**, 2–7.

Group, J. D. R. F. C. G. M. S. (2008), 'Continuous glucose monitoring and intensive treatment of type 1 diabetes', *New England Journal of Medicine* **359**(14), 1464–1476.

Haidar, A., Wilinska, M. E., Graveston, J. A. and Hovorka, R. (2013), 'Stochastic virtual population of subjects with type 1 diabetes for the assessment of closed-loop glucose controllers', *IEEE Transactions on Biomedical Engineering* **60**(12), 3524–3533.

Holt, R. I., DeVries, J. H., Hess-Fischl, A., Hirsch, I. B., Kirkman, M. S., Klupa, T., Ludwig, B., Nørgaard, K., Pettus, J., Renard, E. et al. (2021), 'The management of type 1 diabetes in adults. a consensus report by the american diabetes association (ada) and the european association for the study of diabetes (easd)', *Diabetes Care* **44**(11), 2589–2625.

Huang, S., Yang, J., Fong, S. and Zhao, Q. (2020), 'Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges', *Cancer letters* **471**, 61–71.

Islam, M. N., Hasan, M., Hossain, M. K., Alam, M. G. R., Uddin, M. Z. and Soylu, A. (2022), 'Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from ct-radiography', *Scientific Reports* **12**(1), 11440.

Jafar, A., El Fathi, A. and Haidar, A. (2021), 'Long-term use of the hybrid artificial pancreas by adjusting carbohydrate ratios and programmed basal rate: A reinforcement learning approach', *Computer Methods and Programs in Biomedicine* **200**, 105936.

Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B. J., Jacobsen, L. M., Schatz, D. A. and Lernmark, Å. (2017), 'Type 1 diabetes mellitus', *Nature reviews Disease primers* **3**(1), 1–17.

Kaul, S. and Kumar, Y. (2020), 'Artificial intelligence-based learning techniques for diabetes prediction: challenges and systematic review', *SN Computer Science* **1**(6), 322.

Khan, A. and Zubair, S. (2022), 'An improved multi-modal based machine learning approach for the prognosis of alzheimer's disease', *Journal of King Saud University-Computer and Information Sciences* **34**(6), 2688–2706.

Kovatchev, B. P., Breton, M., Dalla Man, C. and Cobelli, C. (2009), 'In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes'.

Lawton, J., Blackburn, M., Rankin, D., Allen, J., Campbell, F., Leelarathna, L., Tauschmann, M., Thabit, H., Wilinska, M., Hovorka, R. et al. (2019), 'The impact of using a closed-loop system on food choices and eating practices among people with type 1 diabetes: a qualitative study involving adults, teenagers and parents', *Diabetic Medicine* **36**(6), 753–760.

LeCun, Y., Bengio, Y. and Hinton, G. (2015), 'Deep learning', *nature* **521**(7553), 436–444.

Lee, S., Kim, J., Park, S. W., Jin, S.-M. and Park, S.-M. (2020), 'Toward a fully automated artificial pancreas system using a bioinspired reinforcement learning design: In silico validation', *IEEE Journal of Biomedical and Health Informatics* **25**(2), 536–546.

Lehmann, E. and Deutsch, T. (1992), 'A physiological model of glucose-insulin interaction in type 1 diabetes mellitus', *Journal of biomedical engineering* **14**(3), 235–242.

Lim, M. H., Lee, W. H., Jeon, B. and Kim, S. (2021), 'A blood glucose control framework based on reinforcement learning with safety and interpretability: In silico validation', *IEEE Access* **9**, 105756–105775.

Mastrototaro, J. (1999), 'The minimed continuous glucose monitoring system (cgms)', *J Pediatr Endocrinol Metab* **12**(13), 751–758.

Melendez-Ramirez, L. Y., Richards, R. J. and Cefalu, W. T. (2010), 'Complications of type 1 diabetes', *Endocrinology and Metabolism Clinics* **39**(3), 625–640.

Mujahid, O., Contreras, I. and Vehi, J. (2021), 'Machine learning techniques for hypoglycemia prediction: trends and challenges', *Sensors* **21**(2), 546.

Nordhaug Myhre, J., Tejedor, M., Kalervo Launonen, I., El Fathi, A. and Godtliebsen, F. (2020), 'In-silico evaluation of glucose regulation using policy gradient reinforcement learning for patients with type 1 diabetes mellitus', *Applied Sciences* **10**(18), 6350.

Nwokolo, M. and Hovorka, R. (2023), 'The artificial pancreas and type 1 diabetes', *The Journal of Clinical Endocrinology & Metabolism* p. dgad068.

Orozco-López, O., Rodríguez-Herrero, A., Castañeda, C. E., García-Sáez, G. and Hernando, M. E. (2020), 'Method to generate a large cohort in-silico for type 1 diabetes', *Computer Methods and Programs in Biomedicine* **193**, 105523.

Peacock, S., Frizelle, I. and Hussain, S. (2023), 'A systematic review of commercial hybrid closed-loop automated insulin delivery systems', *Diabetes Therapy* **14**(5), 839–855.

Pickup, J. C. (2012), 'Insulin-pump therapy for type 1 diabetes mellitus', *New England Journal of Medicine* **366**(17), 1616–1624.

Pickup, J. C. (2019), 'Is insulin pump therapy effective in type 1 diabetes?', *Diabetic Medicine* **36**(3), 269–278.

Pintaudi, B., Gironi, I., Nicosia, R., Meneghini, E., Disoteo, O., Mion, E. and Bertuzzi, F. (2022), 'Minimed medtronic 780g optimizes glucose control in patients with type 1 diabetes mellitus', *Nutrition, Metabolism and Cardiovascular Diseases* **32**(7), 1719–1724.

Ramkissoon, C. M., Herrero, P., Bondia, J. and Vehi, J. (2018), 'Unannounced meals in the artificial pancreas: detection using continuous glucose monitoring', *Sensors* **18**(3), 884.

Renard, E. (2022), 'Automated insulin delivery systems: from early research to routine care of type 1 diabetes', *Acta Diabetologica* pp. 1–11.

Renard, E., Ikegami, H., Daher Vianna, A. G., Pozzilli, P., Brette, S., Bosnyak, Z., Lauand, F., Peters, A., Pilorget, V., Jurišić-Eržen, D. et al. (2021), 'The sage study: global observational analysis of glycaemic control, hypoglycaemia and diabetes management in t1dm', *Diabetes/metabolism research and reviews* **37**(7), e3430.

Resalat, N., El Youssef, J., Tyler, N., Castle, J. and Jacobs, P. G. (2019), 'A statistical virtual patient population for the glucoregulatory system in type 1 diabetes with integrated exercise model', *PloS one* **14**(7), e0217301.

Roberts, A., Fried, L., Dart, J., de Bock, M., Fairchild, J., King, B., Ambler, G. R., Cameron, F., McAuley, S. A., Keech, A. C. et al. (2022), 'Hybrid closed-loop therapy with a first-generation system increases confidence and independence in diabetes management in youth with type 1 diabetes', *Diabetic Medicine* p. e14907.

Röder, P. V., Wu, B., Liu, Y. and Han, W. (2016), 'Pancreatic regulation of glucose homeostasis', *Experimental & molecular medicine* **48**(3), e219–e219.

Rodríguez-Sarmiento, D. L., León-Vargas, F. and García-Jaramillo, M. (2022), 'Artificial pancreas systems: experiences from concept to commercialisation', *Expert Review of Medical Devices* **19**(11), 877–894.

Roze, S., Smith-Palmer, J., Valentine, W., de Portu, S., Nørgaard, K. and Pickup, J. (2015), 'Cost-effectiveness of continuous subcutaneous insulin infusion versus multiple daily injections of insulin in type 1 diabetes: a systematic review', *Diabetic Medicine* **32**(11), 1415–1424.

Samadi, S., Rashid, M., Turksoy, K., Feng, J., Hajizadeh, I., Hobbs, N., Lazaro, C., Sevil, M., Littlejohn, E. and Cinar, A. (2018), 'Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system', *Diabetes technology & therapeutics* **20**(3), 235–246.

Schapire, R. E. and Freund, Y. (2012), 'Foundations of machine learning'.

Shah, V. N., Laffel, L. M., Wadwa, R. P. and Garg, S. K. (2018), 'Performance of a factory-calibrated real-time continuous glucose monitoring system utilizing an automated sensor applicator', *Diabetes technology & therapeutics* **20**(6), 428–433.

Song, L., Liu, C., Yang, W., Zhang, J., Kong, X., Zhang, B., Chen, X., Wang, N., Shen, D., Li, Z. et al. (2020), 'Glucose outcomes of a learning-type artificial pancreas with an unannounced meal in type 1 diabetes', *Computer methods and programs in biomedicine* **191**, 105416.

Spano, S., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Matta, M., Nannarelli, A. and Re, M. (2019), 'An efficient hardware implementation of reinforcement learning: The q-learning algorithm', *Ieee Access* **7**, 186340–186351.

Steiner, S. S., Li, M., Hauser, R. and Pohl, R. (2010), 'Stabilized glucagon formulation for bihormonal pump use'.

Sun, Q., Jankovic, M. V., Budzinski, J., Moore, B., Diem, P., Stettler, C. and Mougiakakou, S. G. (2018), 'A dual mode adaptive basal-bolus advisor based on reinforcement learning', *IEEE journal of biomedical and health informatics* **23**(6), 2633–2641.

Sutton, R. S. and Barto, A. G. (2018), *Reinforcement learning: An introduction*, MIT press.

Tejedor, M., Woldaregay, A. Z. and Godtliebsen, F. (2020), 'Reinforcement learning application in diabetes blood glucose control: A systematic review', *Artificial intelligence in medicine* **104**, 101836.

Torrent-Fontbona, F. and López, B. (2018), 'Personalized adaptive cbr bolus recommender system for type 1 diabetes', *IEEE journal of biomedical and health informatics* **23**(1), 387–394.

Tyler, N. S. and Jacobs, P. G. (2020), 'Artificial intelligence in decision support systems for type 1 diabetes', *Sensors* **20**(11), 3214.

Viroonluecha, P., Egea-Lopez, E. and Santa, J. (2022), 'Evaluation of blood glucose level control in type 1 diabetic patients using deep reinforcement learning', *Plos one* **17**(9), e0274608.

Wang, Y., Dassau, E. and Doyle, F. J. (2009), 'Closed-loop control of artificial pancreatic β-cell in type 1 diabetes mellitus using model predictive iterative learning control', *IEEE Transactions on Biomedical Engineering* **57**(2), 211–219.

Wang, Y., Zhang, J., Zeng, F., Wang, N., Chen, X., Zhang, B., Zhao, D., Yang, W. and Cobelli, C. (2017), '"learning" can improve the blood glucose control performance for type 1 diabetes mellitus', *Diabetes technology & therapeutics* **19**(1), 41–48.

Wang, Z. and Hong, T. (2020), 'Reinforcement learning for building controls: The opportunities and challenges', *Applied Energy* **269**, 115036.

Wilinska, M. E. and Hovorka, R. (2008), 'Simulation models for in silico testing of closed-loop glucose controllers in type 1 diabetes', *Drug Discovery Today: Disease Models* **5**(4), 289–298.

Zhu, T., Li, K., Herrero, P. and Georgiou, P. (2020), 'Basal glucose control in type 1 diabetes using deep reinforcement learning: An in silico validation', *IEEE Journal of Biomedical and Health Informatics* **25**(4), 1223–1232.

Zhu, T., Li, K., Kuang, L., Herrero, P. and Georgiou, P. (2020), 'An insulin bolus advisor for type 1 diabetes using deep reinforcement learning', *Sensors* **20**(18), 5058.