UNIVERSITY OF PADOVA

**DEPARTMENT OF INFORMATION ENGINEERING**
**Ph.D. Course on Information Engineering**
**Curriculum: Bioengineering**
**Series: XXXII**

# Open-loop insulin dosing personalization in type 1 diabetes using continuous glucose monitoring data and patient characteristics

**Course director:**
Andrea Neviani
**Advisor:**
Giovanni Sparacino
**Co-advisor:**
Andrea Facchinetti

**Ph.D. candidate:**
Giacomo Cappon

A thesis submitted
for the degree of
PhilosophiæDoctor (Ph.D.)

# Acknowledgements

A very special thanks goes to my girlfriend Francesca and my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

I would like to deeply thank my advisor Giovanni Sparacino and my co-advisor Andrea Facchinetti for the invaluable commitment, the constant encouragement, and the precious suggestions, that beside supporting my Ph.D. study and related research, helped me to grow up as a person.

In addition, I would like to thank the reviewers of this thesis, Prof. Francesco Di Nardo and Prof. Jorge Bondia, for their precious suggestions.

Finally, I extend my thanks to all my colleagues who shared with me this incredible experience for the stimulating discussions and all the fun we had in the last three years. Their were always available whenever I ran into a trouble spot or had a question about my research or writing. They consistently allowed this work to be my own, but steered me in the right direction whenever I needed it.

# Abstract

Patients with type 1 diabetes (T1D) require lifelong insulin therapy in order to maintain their blood glucose (BG) concentration within the euglycemic range preventing long-term complications associated with hyperglycemia and avoiding dangerous episodes of hypoglycemia. To achieve proper glycemic control, people with T1D need to perform a constant learning process about how daily conditions (e.g. insulin administrations, meals schedule and composition, physical activity, and illness) affect BG levels. More than 500,000 operations can be needed during the lifetime of a T1D patient to manage the therapy. For this reason, management of diabetes is burdensome for patients, and results in deteriorating their quality of life. One of the major issues in the daily management of T1D concerns with the amount of insulin that has to be administered, by a subcutaneous bolus injection, in order to compensate the increase of BG associated with meals. So far, a standard simple mathematical formula (SF), designed by clinical investigators on an empirical basis, is commonly used by patients to calculate the size of insulin boluses. SF leverages on the current BG level obtained from self monitoring blood glucose (SMBG) samples, the estimated amount of carbohydrates (CHOs) present in the meal, and patient specific therapy parameters. While the SF is well-established in clinical practice, the insulin amount determined through its use could be suboptimal due to several reasons, including the error patients make in estimating CHO, the intrinsical sparseness of SMBG, and the inability of accounting for many important factors such as patients' intra-/interday variability.
Margins of improvement over the SMBG-based SF emerged in the past decade, when diabetes management has been transformed by the introduction of minimally invasive continuous glucose monitoring (CGM) sensors, which have been recently approved by regulatory agencies, such as the Food and Drug Administration (FDA), to be usable to make treatment decisions, such as insulin dosing. Of course, CGM provides an increased amount of available features on BG, such as the rate of change (ROC), that could be exploited to improve

insulin standard therapy. As a matter of fact, several attempts have been proposed in the literature to account for CGM-derived information and adjust SF accordingly, but unfortunately, they fall short in personalizing such an adjustment patient-by-patient.

In this thesis we propose new methodologies for determining a dose of insulin bolus able to effectively account for the "dynamic" information on BG provided by the ROC and patient characteristics, the final aim being to personalize the standard insulin therapy and eventually improve the glycemic control. In particular, to identify the possible margins of improvement, in the first part of the thesis we assess and analyze the criticalities of three popular literature techniques that exploit the ROC magnitude and direction to adjust the insulin bolus amount computed through SF. To such a scope, we designed ad-hoc in silico clinical trials implemented using a popular powerful simulation tool, i.e. the UVa/Padova T1D Simulator. Then, in the second part, we propose two novel machine learning based algorithms that, being fed by information on current patient status and characteristics, provide patients with new tools to adjust SF in a personalized manner. Finally, in the third part of the thesis, we abandon the idea of using the insulin bolus provided by SF as a sort of initial estimate to be simply adjusted, and we design a brand new formula for insulin bolus determination that naturally takes into account for CGM-derived information and current patient status and characteristics. This represents an innovation in the literature because no insulin bolus formulae specifically designed for use with CGM have been proposed yet.

The thesis is organized in six chapters. In Chapter 1, after introducing T1D therapy, the importance of designing new personalized insulin dosing guidelines leveraging on CGM is discussed. Then, some insulin dosing techniques proposed in the literature to adjust SF according to ROC are reviewed. The chapter ends with the statement of the aim of the thesis.

In Chapter 2, we assess in depth, by means of data of 100 virtual subjects generated by the UVa/Padova T1D Simulator, strengths and the limitations of the above cited literature techniques for the adjustment of SF according to ROC. In particular, we demonstrate that these techniques fall short in universally improve the glycemic control in the T1D population due to their inability of personalizing the adjustment of SF at the patient level.

In Chapter 3, we develop a new method based on a neural network fed by CGM information and easily accessible patient parameters to improve the performance, in term of glycemic control, over the available literature methods.

The new technique consists in using the SF parameters, ROC, body weight, insulin pump basal infusion rate and insulin sensitivity as features to train a neural network to "optimally" adjust SF. Using the UVa/Padova T1D Simulator, we generated data of 100 subjects. Specifically, we simulated each subject several times to analyze different meal conditions in terms of carbohydrate intakes, preprandial BG and ROC, and determine, by trial-and-error, the optimal adjustment, in terms of units of insulin, that would have been needed to be applied to SF to obtain optimal glycemic control. Finally, using the consolidated machine learning paradigm, a subset of data associated to 80 subjects are extracted to be used for training and validation of the neural network while data associated to the remaining 20 subjects are extracted and used to assess the model performance. Results demonstrate that the proposed model is able to significantly improve glycemic control.

Given the encouraging results obtained in Chapter 3, in Chapter 4, a different approach is developed which still use tools popular in the machine learning field. In particular, CGM data, together with commonly recorded data, i.e. CHO intake and insulin bolus infusion recordings, are used to train a gradient boosted tree model that allows to classify, at meal-time, the postprandial glycemic status (i.e., BG concentration being too low, too high, or within the target range). Then such an outcome is used to reduce or increase the corresponding meal bolus dose provided by SF accordingly. Results obtained on data of 100 simulated subjects show that, when used to adjust, in real-time, meal insulin boluses obtained with a bolus calculator, the proposed approach improves glycemic control.

The results we obtained in Chapter 3 and Chapter 4 showed that our strategy, i.e. leveraging on machine learning to adjust SF, allows to improve glycemic control. However, possible margins of improvement rose by the possibility of abandoning the idea of using the insulin bolus computed through SF and design a new formula that naturally include patient characteristics and status to compute the insulin dose. As such, in Chapter 5, different types of linear regression models have been analysed and applied in order to design new formulae for insulin meal bolus calculation. Results, obtained on data of 100 subjects generated by the UVa/Padova T1D Simulator, show that, when compared with state-of-art insulin dosing guidelines, the proposed formulas improve glycemic control over the population encouraging further development of these promising methods.

Finally, the major findings of the work carried out in this thesis, possible

applications and margins of improvements are summarized in Chapter 6.

# Sommario

I pazienti affetti da diabete di tipo 1 (T1D) necessitano, durante l'intera durata della loro vita, di una terapia basata su amministrazioni di insulina esogena al fine di mantenere la propria concentrazione di glucosio nel sangue (BG) all'interno del range euglicemico e quindi prevenire complicazioni a lungo termine dovute ad episodi di iperglicemia ed evitare al contempo pericolosi episodi ipoglicemici. Per assicurare un adeguato controllo glicemico, i soggetti affetti da T1D necessitano di un continuo processo di apprendimento al fine di capire come, condizioni quotidiane, che includono, ad esempio, amministrazioni di insulina esogena, dieta, attivita' fisica, e malesseri, impattano il livello di BG. Durante la vita di un paziente affetto da T1D, si stima possano essere necessarie piu' di 500,000 azioni (es. assunzioni di boli correttivi di insulina, misurazioni di BG tramite dispositivi pungidito). Per questa ragione, il pesante carico dovuto alla gestione del diabete si traduce in una ridotta qualita' della vita. Uno dei problemi piu' importanti della gestione quotidiana del T1D riguarda l'ammontare di insulina esogena da amministrare, tramite l'iniezione di un bolo sottocutaneo, al fine di compensare l'incremento di BG associato ai pasti. Ad oggi, una semplice formula matematica standard (SF), sviluppata da ricercatori clinici su base empirica, viene comunemente utilizzata dai pazienti per il calcolo della dimensione del bolo di insulina. SF sfrutta il livello di BG corrente, misurata da campioni di sangue ottenuti tramite dispositivi di automonitoraggio capillare (SMBG), la quantita' stimata di carboidrati (CHO) assunti, e parametri di terapia paziente specifici. Anche se SF e' ben affermata nella pratica clinica, l'ammontare di insulina calcolato tramite SF puo' risultare subottimo a causa di diversi fattori come l'errore commesso dai pazienti nella stima di CHO, l'intrinseca sparsita' delle misure SMBG, e l'incapacita' di tenere conto di importanti fattori come, per esempio, la variabilita' intra/inter-giornaliera della sensibilita' insulinica dei pazienti. Negli ultimi anni, diverse possibilita' di miglioramento sono emerse grazie all'introduzione, nella terapia per il diabete, di sensori min-

imamente invasivi per il monitoraggio in continua del glucosio (CGM), i quali sono stati recentemente approvati dai vari enti regolatori, come la Food and Drug Administration (FDA), per essere utilizzati al fine di formulare decisioni terapeutiche come il dosaggio di insulina. I sensori CGM offrono molta piu' informazione su BG, come, per esempio, il valore dell'attuale trend glicemico (ROC), che idealmente puo' essere sfruttato per migliorare la terapia standard del diabete. A tale fine, in letteratura, diversi sono stati i tentativi di integrare l'informazione su BG derivata dai sensori CGM al fine di aggiustare l'ammontare di insulina calcolato tramite SF. Sfortunatamente, questi tentativi falliscono nel personalizzare tale correzione a livello del singolo paziente.

In questa tesi, allo scopo di personalizzare la terapia standard del diabete e migliorarne i benefici in termini di controllo glicemico, proponiamo nuove metodologie per il dosaggio di insulina capaci di tenere conto in maniera efficace dell'informazione "dinamica" su BG descritta dalla ROC e dalle caratteristiche peculiari del singolo paziente. In particolare, con lo scopo di individuarne i possibili margini di miglioramento, la prima parte della tesi vertera' sulla valutazione e l'analisi delle criticita' di tre popolari tecniche stato dell'arte che, sfruttando il modulo e il segno della ROC, correggono l'ammontare di insulina calcolato tramite SF. A tal fine, abbiamo progettato, in silico, trial clinici ad-hoc implementati tramite l'utilizzo di un tanto diffuso quanto potente strumento di simulazione: il simulatore di T1D UVa/Padova. Quindi, nella seconda parte della tesi, proponiamo due algoritmi innovativi basati sul machine learning che, utlizzando come input l'informazione sullo stato corrente del paziente e le sue caratteristiche, rappresentano nuovi strumenti per la correzione personalizzata di SF. Infine, nella terza parte, abbandoniamo l'idea di utilizzare il bolo di insulina ottenuto da SF come una sorta di punto di partenza da correggere, e proponiamo una formula completamente nuova per la determinazione del bolo di insulina che integra, nella sua definizione, l'informazione aggiuntiva portata dall'utilizzo dei sensori CGM, lo stato corrente del paziente e le sue caratteristiche. Quest'ultimo punto rappresenta un'innovazione nella letteratura dato che, ad oggi, nessuna formula, che sia stata progettata nello specifico per essere utilizzata con i sensori CGM, e' stata ancora proposta.

La tesi e' organizzata in sei capitoli. Nel Capitolo 1, dopo aver descritto come si presenta, ad oggi, la terapia standard per il T1D, viene discussa l'importanza di progettare nuove linee guida per il calcolo della dose di insulina capaci di fare leva sulle potenzialita' offerte dai sensori CGM. Successi-

vamente, vengono presentate alcune tecniche di letteratura per la correzione di SF tramite ROC. Il capitolo termina con la dichiarazione formale dell'obiettivo di questa tesi.

Nel Capitolo 2, valutiamo approfonditamente, utilizzando dati di 100 pazienti virtuali generati tramite il simulatore di T1D UVa/Padova, i punti di forza e di debolezza, delle tecniche di letteratura sopra citate per la correzione di SF tramite ROC. In particolare, dimostriamo come queste tecniche falliscono nel migliorare universalmente il controllo glicemico nella popolazione virtuale a causa della loro incapacita' di personalizzare tale correzione di SF a livello del singolo paziente.

Nel Capitolo 3, sviluppiamo un nuovo metodo per la correzione di SF basato su una rete neurale con lo scopo di migliorare la performance in termini di controllo glicemico ottenute con i metodi di letteratura stato dell'arte sopra citati. Al fine di correggere in maniera "ottima" SF, tale rete neurale viene allenata tramite l'utilizzo dell'informazione derivata da CGM e alcuni parametri paziente specifici facilmente accessibili che sono: i parametri di SF, la ROC, il peso del paziente, il rate di infusione della pompa insulinica, e la sensibilita' insulinica del paziente. Con il simulatore di T1D UVa/Padova abbiamo generato dati di 100 pazienti virtuali. Nello specifico, abbiamo simulato ogni soggetto piu' volte per analizzare diverse condizioni al pasto in termini di CHO, BG e ROC preprandiale, e abbiamo determinato, con una procedura trial-and-error, il valore di correzione ottimo, in termini di unita' di insulina, che sarebbe stato necessario applicare a SF al fine di ottenere il controllo glicemico ottimo. Infine, utilizzando il consolidato paradigma del machine learning, un subset di dati associati a 80 pazienti e' stato utilizzato per allenare e validare la rete neurale, mentre i dati associati ai rimanenti 20 pazienti sono stati utilizzati per valutare le performance del modello. I risultati ottenuti dimostrano come il modello proposto sia capace di migliorare significativamente le performance glicemiche.

Dati i risultati incoraggianti ottenuti nel Capitolo 3, nel Capitolo 4, utilizzando ancora comuni strumenti propri del machine learning, proponiamo un diverso approccio al problema. In particolare, dati CGM e dati comunemente raccolti dal paziente, come l'ammontare di boli di insulina e di carboidrati assunti, sono stati utilizzati per allenare un albero decisionale con lo scopo di classificare, al momento del pasto, lo stato glicemico postprandiale (ossia se il livello di BG postprandiale e' troppo basso, troppo elevato, o nell'intervallo target). Successivamente, l'output dell'albero decisionale e' stato utilizzato

come indicazione per ridurre o aumentare la dose di insulina calcolata tramite SF. I risultati ottenuti su dati di 100 soggetti virtuali mostrano come il nuovo approccio sia capace di migliorare il controllo glicemico.

I risultati che abbiamo ottenuto nel Capitolo 3 e nel Capitolo 4 mostrano come la strategia da noi proposta, ossia sfruttare algoritmi di machine learning per correggere SF, sia capace di migliorare il controllo glicemico. Ad ogni modo, nuovi margini di miglioramento possono essere individuati nella possibilita' di abbandonare l'idea di utilizzare la dose di insulina calcolata tramite SF e progettare una nuova formula per il dosaggio di insulina che integra naturalmente la caratteristiche del paziente e il suo stato glicemico. A tale scopo, nel Capitolo 5, diversi tipi di modelli di regressione lineare sono stati analizzati e applicati per costruire nuove formule per il calcolo della dose di insulina. I risultati ottenuti su dati di 100 pazienti generati tramite il simulatore di T1D UVa/Padova, dimostrano che, quando confrontati con le linee guida stato dell'arte per il dosaggio di insulina, le formule proposte migliorano il controllo glicemico nella popolazione diabetica virtuale suggerendo quindi di proseguire lo sviluppo di tali metodologie.

Infine, nel Capitolo 6 vengono riassunti i principali risultati ottenuti in questo lavoro di tesi, le applicazioni pratiche, e i possibili margini di miglioramento da perseguire nel futuro.

# Contents

# Abbreviations

**AUROC** Area under the receiver operator characteristic curve

**BC** Bolus calculator

**BG** Blood glucose

**BGRI** Blood glucose risk index

**BU** Method to correct SF from Buckingham et al.

**BW** Body weight

**CBR** Case-based reasoning

**CF** Correction factor

**CGM** Continuous glucose monitoring

**CHO** Carbohydrates

**COB** Carbohydrate-on-board

**CR** Carbohydrate-to-insulin ratio

**CSII** Continuous subcutaneous insulin infusion

**DSS** Decision support system

**FDA** Food and Drug Administration

$\mathbf{G}_C$ Current glucose level

$\mathbf{G}_T$ Target glucose level

**GBT** Gradient boosted tree

**HBGI** High blood glucose index

**IG** Interstitial glucose

**IMB** Insulin meal bolus

**INS** Insulin infusion recordings

**IOB** Insulin-on-board

**I**$_P$ Plasmatic insulin concentration

**ISCT** In silico clinical trial

**LASSO** Least absolute shrinkage and selection operator

**LBGI** Low blood glucose index

**LR** Linear regression

**MDI** Multiple daily injections

**MLR** Multiple linear regression

**MSE** Mean-squared error

**NN** Neural network

**NNC** Neural network based corrector

**PCC** Pearson's correlation coefficient

**PE** Method from Pettus and Edelman

**R2R** Run-to-run

**R**$_a$ Glucose rate of appearance in the plasma

**ROC** Rate of change

**RSS** Residual sum of squares

**S2013** 2013 version of the UVa/Padova T1D Simulator

**SC** Method to correct SF from Scheiner

**SD** Standard deviation

**SF** Standard formula

**SMBG** Self-monitoring of blood glucose

**T1D** Type 1 diabetes

**T1D-DM** Type 1 diabetes decision-making

**TDI** Total daily insulin

**TPE** Tree Parzen estimator

**TSS** Total sum of squares

**XGB** Extreme gradient boosted tree

**XGB-IMB** XGB based BC

# Chapter 1

# Type 1 diabetes and open loop control of blood glucose concentration

## 1.1 Type 1 diabetes (T1D) mellitus and standard insulin therapy

Type 1 diabetes (T1D) is a chronic autoimmune disease which consists of the destruction of the pancreatic beta cells that are responsible for insulin production [1] [2]. As a consequence, patients affected by T1D can be treated only by exogenous administrations of insulin to avoid their blood glucose (BG) concentration to exceed the safe range, i.e. 70-180 mg/dL, and resulting in hyperglycemic events that, in the long term, lead to neuropathy, retinopathy, and serious micro-/macrovascular heart diseases. On the other hand, excessive administrations of exogenous insulin are even more dangerous, since they trigger hypoglycemic events that, in the short term, result to dizziness, lightheadedness, fainting, and, in extreme cases, to coma or even death [3].
T1D standard therapy is an "open-loop" control approach that aims at maintaining BG in the safe range during the day and consists in injecting exogenous insulin either via multiple daily injections (MDI) (Figure 1.1, panel A), where fast-acting insulin is used to counterbalance glucose fluctuations due to meal carbohydrate intakes, and long-acting insulin is used once or twice per day to keep BG to its basal level, or via continuous subcutaneous insulin infusion (CSII) by means of an insulin pump (Figure 1.1, panel B). Insulin dosing is an extremely delicate and challenging task. Usually, each injection is tuned ac-

cording, but not limited to, diet [4], physical activity [5], and BG concentration, that is commonly obtained through the use of self monitoring of blood glucose (SMBG) meters [6], i.e. portable systems that allow measuring BG concentration, usually 3-4 times per day, on a drop of capillary blood as shown in Figure 1.2.

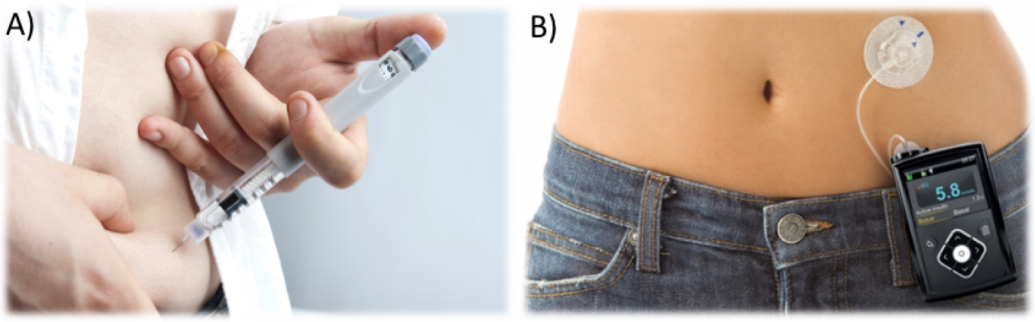



**Figure 1.1:** Insulin delivery via multiple daily injections (A) or via insulin pump (B). (Source: [7], [8])


**Figure 1.2:** SMBG procedure: Patients with T1D collect a blood sample from their finger and measure the BG concentration by a portable sensor device. (Source: [9])

## 1.2 The standard formula (SF) used in clinical practice to compute the insulin dose

So far, in order to maximize the benefits of the open-loop insulin therapy, an empirical simple standard formula (SF), that estimates the insulin bolus thought to be able to maintain BG within the safe range, is commonly used in clinical practice [10]. Such a formula is defined as follows:

$$IMB = \frac{CHO}{CR} + \frac{G_C - G_T}{CF} - IOB \tag{1.1}$$

where $IMB$ (U) is the estimated insulin bolus amount.
SF consists of three terms. The first term compensates for the carbohydrate amount planned to be eaten and it is computed as the ratio between CHO ($g$), i.e. the estimated carbohydrate intake, and the carbohydrate-to-insulin ratio ($CR$) ($g/U$), i.e. a therapy parameter that specifies the number of grams of carbohydrate covered by each unit of insulin [11]. The second term consists in increasing or reducing the needed amount of insulin according to how much the patient is currently far from its target BG level ($G_T$). It is calculated as the ratio between the difference between the current BG concentration ($G_C$) and $G_T$, and the correction factor $CF$ ($mg/dL/U$), i.e. a therapy parameter that quantifies the drop in blood glucose level caused by the administration of 1 unit of insulin [12]. Finally, in order to avoid insulin overdosing, the third term corrects the total insulin amount $IMB$ by subtracting the so-called insulin-on-board ($IOB$), i.e. an estimate of the amount of insulin in the body that has been previously injected and that has not been assimilated by the organism yet.
To help patients in computing SF, it is possible to adopt the so-called bolus calculators (BC) [13], i.e. simple software tools implementing SF that are commonly integrated in all major insulin pumps or available as standalone mobile applications (Figure 1.3 shows as example of BC mobile application in action). BCs resulted to be very useful in T1D management [14][15], indeed, for most people, bolus insulin equations are difficult and time consuming to solve by mental calculations. In this regard, many literature studies document a high frequency of poor numeracy skill among adults with T1D. For example, in a study conducted by Marden et al. [16], 201 adults with T1D were asked to manually compute the required amount of insulin needed to compensate for a specific amount of carbohydrate in a meal. It resulted that 57% of the subjects came to a wrong insulin dose computation. On the contrary, when provided with a BC, the proportion of wrong computations was reduced to 7% confirming the beneficial effects brought by the integration of BC in T1D therapy. Moreover, Gross et al. [14] demonstrated that BCs significantly reduced the number of insulin boluses required to contrast postprandial hyperglycemia (due to meal-insulin under-dosage) and the amount of rescue carbohydrates needed to recover from postprandial hypoglycemia (due to meal-insulin over-dosage).

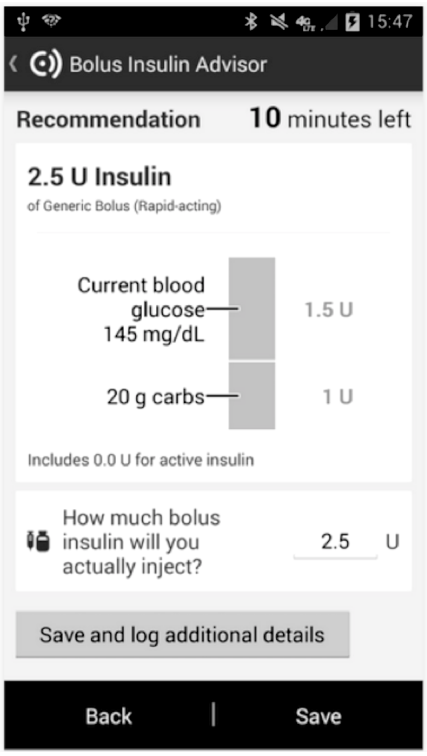



**Figure 1.3:** Example of a bolus calculator app in action (Accu-Check Connect [17]).

## 1.3 Open problems of current methodologies for insulin therapy

There are multiple disadvantages related to the standard insulin therapy which make it sub-optimal. First of all, the patient lacks motivation to collect SMBG measurements or does not have sufficient education on how to interpret them. Moreover, the standard insulin therapy is very demanding as it greatly impacts the quality of life of patients with T1D that are required to perform, for the entirety of their life, a huge number of tasks. Just to give an example, to effectively counterbalance the BG fluctuation due to a meal, patients are required to measure preprandial BG via SMBG, to inject insulin boluses to contrast prandial hyperglycemia and eventually to repeat the whole procedure until BG gets back within the target range. As a matter of fact, it can be estimated that a subject suffering from T1D can potentially perform more than 500,000 actions during its lifetime. Additionally, although BCs implementing SF have been proven to be clinically effective, they are still far from being optimal. First of all, SF relies on SMBG measurements which are sparse [18] [19], thus not able to properly capture both the shortcoming of hypo-/hyperglycemic events and the information on glucose trend. This can be particularly dangerous for

patients with impaired awareness of hypoglycemia, in whom the ability to perceive the onset of hypoglycemia is reduced or even absent [20]. In these patients, the incidence of severe low glucose episodes that requires external assistance for recovery is up to six fold higher than in subjects with normal awareness of hypoglycemia [21]. This limitation can be in part overtaken by resorting to the new continuous glucose monitoring sensors, as discussed later on in Subsection 1.4.1. Moreover, focusing on the SF definition, parameters $CR$ and $CF$ may vary during the day according to intra-/interday insulin sensitivity variability, physical activity, alcohol consumption and other factors that could affect the patient metabolism, such as hormone cycles, illness, and stress [22] [23]. Finally, effectiveness of SF strongly depends on the patients' skill of correctly estimating CHO intakes (inaccuracy of $\pm$20 g significantly affects postprandial glycaemia in children [24]).

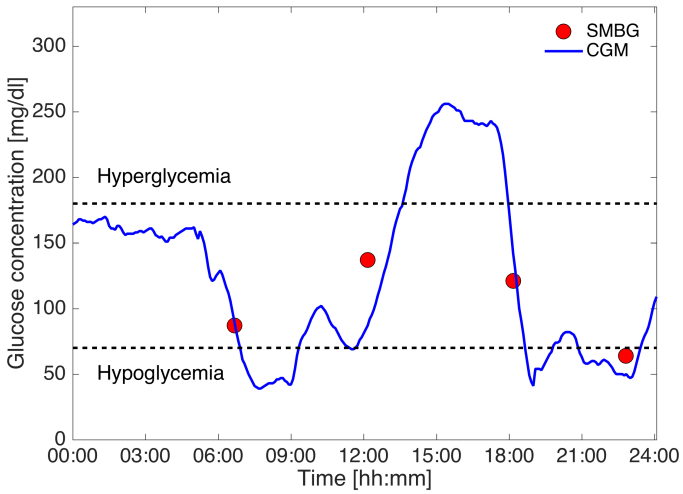



**Figure 1.4:** Representative BG monitoring data obtainable with SMBG (red dots) and with CGM (blue line).

## 1.4 New scenarios in T1D management

### 1.4.1 Continuous glucose monitoring (CGM)

Exciting perspectives in the management of T1D were opened in the last 20 years by the introduction of minimally invasive needle sensors for real-time continuous glucose monitoring (CGM) [25][26][27]. CGM delivers an almost continuous glucose trace providing a BG measurement every 1-5 minutes, mitigating the need of the SMBG, and greatly increasing the information on BG

fluctuations (see Figure 1.4 which shows that, using CGM, one can detect hypoglycemic and hyperglycemic episodes otherwise undetectable through SMBG use).

Since the first CGM prototype approved for commercialization by the US Food and Drug Administration (FDA) back in 1999 [28], CGM evolved rapidly. Indeed, until few years ago, CGM devices were only approved to be used in adjunct to SMBG (the so-called adjunctive use), i.e. before making therapy decisions such as insulin dosing, BG concentration provided by CGM needed to be confirmed by an SMBG due to CGM poor accuracy. As such, CGM measurements resulted to be unreliable and, more importantly, this problem limited the early adoption of CGM devices by both patients and clinicians who felt unsafe in integrating CGM in diabetes treatment. This completely changed in the last few years thanks to the constant efforts, both in terms of money and time, of CGM manufactures to push forward glucose sensing technologies. Indeed, many commercialized CGM sensors now achieve accuracies close to, or even within, the SMBG accuracy range [29][30]. This technological improvement led to the regulatory approval of CGM for the so-called non-adjunctive use [31], i.e. allowing CGM to be used to make treatment decision without the need of confirmatory fingersticks, whose safety have been proven by computer simulations and a randomized non-inferiority clinical trial [32].
The beneficial impact brought by CGM use has been proved through many other randomized clinical trials [33] [34] [35] [36] [37]. CGM resulted to be effective in improving glycemic control, mitigating hypoglycemic events, and reducing glycemic variability, not only for patients with T1D, but also subjects suffering of type 2 diabetes, and gestational diabetes just to mention a few. Clinical evidence is constantly growing in support of CGM encouraging and motivating its spreading across the diabetes population.

### Exploiting CGM to improve bolus calculators

Besides achieving good measurement accuracy, improving diabetes treatment, and guaranteeing all the basic requirements needed for long-term wearable biosensors [38], i.e. bio-compatibility, lifetime, safety, sensitivity, and specificity, state-of-the-art CGM systems provide users with a set of smart features to improve patient self-management. Indeed, a very interesting and appealing perspective lies in the possibility of designing and developing new algorithms for diabetes care exploiting CGM-derived information. Specifically, BC based

on SF can be targeted and improved by including the information provided by CGM sensors [39]. For example, in Herrero et al. [40], an approach to automatically adjust $CR$ and $CF$ exploiting the CGM measurements is presented and demonstrated to improve glycemic control. The method is based on a run-to-run (R2R) control technique, i.e., an iterative procedure in which the values of $CR$ and $CF$ are updated daily according to a performance metric, e.g., the distance between the minimum postprandial glucose concentration and the patient's target BG. Another work by Herrero et al. [41] improved the performance achieved by R2R by integrating case-based reasoning (CBR), i.e., an artificial intelligence technique that solves new problems by applying solutions learned from solving similar problems in the past.
All of the above cited methods rely on SF calculation, which takes into account only a single BG measurement. Intuitively, this might be suboptimal given that one can exploit the CGM measurements to obtain more information on the patients' current status. Indeed, in a non-adjunctive CGM scenario, SF can still be implemented and used by simply substituting the BG reading obtained using SMBG with the value provided by CGM at the time of the bolus calculation. However, this approach not only leaves the "static" nature of SF unaltered, but also does not use the additional information brought by CGM. Present CGM systems, such as the Dexcom G5 Mobile (Dexcom Inc, San Diego, California, USA) [42] or the Senseonics Eversense (Senseonics, Inc., Germantown, MD, USA) [43] system, provide users with the so-called rate-of-change (ROC) arrows (see Figure 1.5), i.e., a graphical indication of the current direction and velocity of changing glucose. It is natural to think of improving BCs outcomes by modulating the recommended insulin dosage according to the patient's current ROC. Indeed, such an information is intuitively helpful to better determine the amount of exogenous insulin to be administered. As such, new insulin dosing strategies are needed to fully benefit of non-adjunctive CGM use.

#### Adjustment of SF according to ROC

The ROC information provided by CGM can be usefully exploited to adjust SF. Intuitively, when ROC is negative, i.e. BG is decreasing, one would like to reduce the insulin bolus computed using SF to preventively avoid the possible shortcoming of dangerous hypoglycemic episodes. On the other hand, when


**Figure 1.5:** Example of CGM systems that provide ROC. To the left, the Dexcom G5 CGM system (taken from [42]). To the right, the Senseonics Eversense CGM system (taken from [43]). Both systems report the ROC magnitude and direction by mean of an arrow in the upper panel of their mobile app.

ROC is positive, one would think of increasing the insulin bolus amount provided by SF to prevent, or at least mitigate, hyperglycemia.
Following this concept, several methodologies have been proposed in the literature the aim being adjusting SF according to ROC direction and magnitude. The most popular methodologies include those of Buckingham et al. [44], Scheiner [45], and Pettus and Edelman [46], and provide patients with simple guidelines to integrate ROC in the insulin bolus computation. However, to the best of our knowledge, despite such guidelines are intuitive and are based on clinician's experience, no study has evaluated their impact on glycemic control either quantitatively nor qualitatively. The main reason is that a clinical study that would generate statistically-meaningful clinical data to support the superiority of one method over another, may not be feasible to conduct [47]. Indeed, testing new insulin dosing strategies may induce patients to risky scenarios, such as severe hypoglycemia, that are so varied and rare that would require a huge amount of data in order to be properly studied.

### 1.4.2 In silico clinical trials to assess new insulin therapy strategies

#### In vivo versus in silico clinical trials

Development of new strategies for diabetes treatment and their assessment on human subjects is a really time consuming, costly process. According to

the Tufts Center for the Study of Drug Development, the commercialization of new pharmaceutical products has been increasing exponentially mainly because of clinical assessment [48]. Clinical assessment requires the enrollment of a sufficient number of subjects in order to provide actual proof of a new product/treatment superiority. More importantly, clinical studies have to guarantee patient safety, but this is not easy as it strongly depends on which strategy one has to test. To overcome this issue, in silico clinical trials can be performed. In silico clinical trials allow the set up of computer simulation environment where virtual subjects are modeled by initializing a disease/intervention model with quantitative information either measured on an individual (subject specific model), or inferred from population distributions of those values (population specific model). Specifically to diabetes, they consist of using an individualized simulation model of glucose-insulin dynamics to preliminary design and test new methodologies for BG control in a safe and cost efficient environment. In silico clinical trials can be used to overcome the limitations of clinical trials. Indeed, they allow to explore the impact of new treatments both in a average low risk situation and in numerous high risk scenarios which are impossible, dangerous, and unethical to replicate in a real-life setting. Specifically, they allow to run multiple tests on the same virtual subject maintaining the same surrounding conditions, i.e. something impossible to replicate in real-life since patient's behaviour and physiological status are never the same. For this purpose, multiple simulation models have been proposed in the literature so far including those of Hovorka et al. [49], Willinska et al. [50], Kanderian et al. [51], and Haidar et al. [52]. In particular, one of the most used model of T1D physiology is the UVa/Padova T1D Simulator [53], firstly developed in 2008 by the University of Padova and University of Virginia and briefly described in the following Subsection.

**The UVa/Padova T1D Simulator**

The UVa/Padova T1D Simulator is a large-scale maximal model describing the glucose insulin regulatory system by means of 13 differential equations and 35 subject-specific parameters. The simulator is equipped with 300 "virtual" subjects (100 adults, 100 adolescents, 100 children) spanning the real T1D population variability observed in vivo. In addition, the simulator is equipped with models reproducing CGM sensor errors, e.g. virtual sensors that can be placed on the subjects to evaluate the impact BG measurement error during

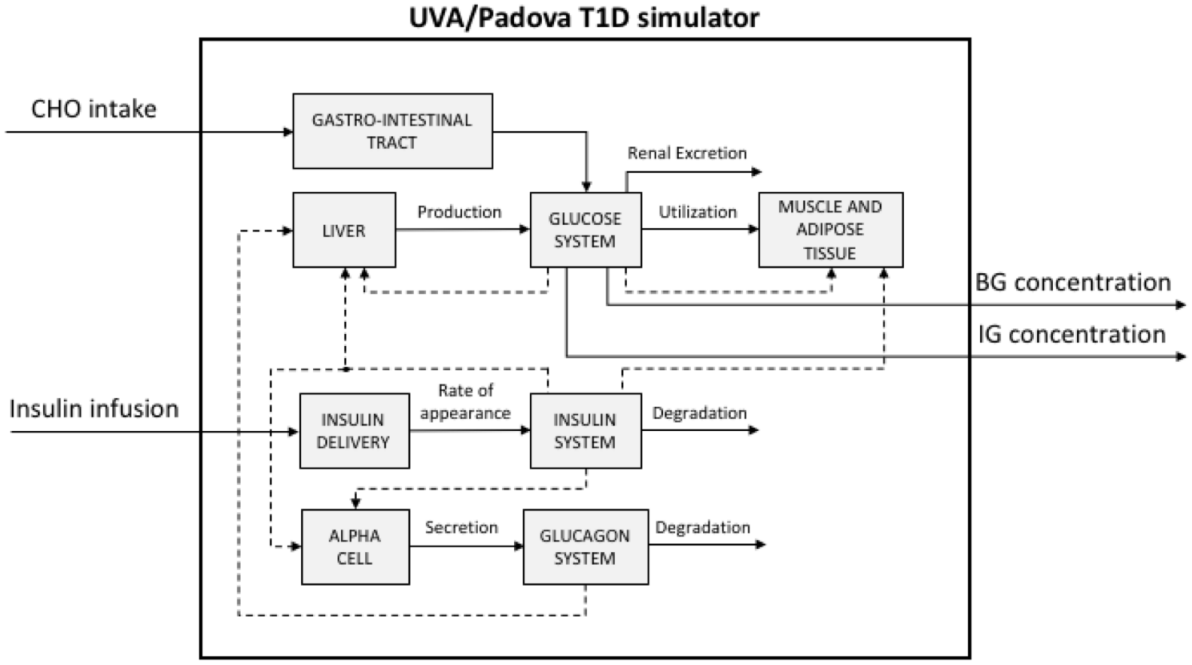



**Figure 1.6**: Schematic of the model included in the FDA-accepted T1D simulator (2013 version). Note: exogenous glucagon delivery is not shown in the scheme because not used in this thesis.

the in silico experiments. Subcutaneous insulin delivery is modeled as well, allowing to place virtual insulin pumps on the virtual subjects and assessed predefined insulin treatment strategies. Each of the virtual patient parameters have been derived by a unique data set of 204 normal individuals who underwent a triple tracer meal protocol that allowed to measure not only plasma glucose and insulin, but also crucial fluxes of the glucose-insulin system i.e. plasma rate of appearance of ingested CHO, endogenous glucose production, glucose utilization and insulin secretion [54]. Specifically, in order to obtain the model parameters for patients with T1D from those obtained for the healthy subjects, clinical relevant modifications were introduced. A schematic representation of the UVa/Padova T1D Simulator is shown in Figure 1.6. Briefly, it takes as input CHO intakes and exogenous insulin administrations to provide, as outputs, BG and interstitial glucose (IG) concentrations.

The UVa/Padova T1D Simulator was approved in January 2008 by the FDA as a substitute to animal trials for the preclinical testing of control strategies in artificial pancreas studies. This means that in silico clinical trials can produce reliable preclinical results in a fraction of the time and cost required for animal trials. Then, in 2013, a new more sophisticated version of the simulator

has been developed improving the model of glucose dynamics in the hypoglycemic range and incorporating new models of glucagon kinetics, secretion, and action. Moreover, a new refined way to generate the virtual population was adopted. As a result, the new version of the simulator showed better agreement with the real patient-specific parameters population. As the previous version, in 2013, the new simulator has been accepted by FDA as a substitute for preclinical trials. In 2013, the use of the UVa/Padova T1D simulator was furtherly extended from single meal to multiple days by incorporation of a model of intra- and inter-day variability of insulin sensitivity [55] derived from data collected in 20 T1D subjects studied with the triple-tracer method [56]. Finally, the 2013 release of the UVA/Padova T1D simulator with incorporation of the time-variability of insulin sensitivity was validated on 141 glucose traces collected in 47 T1D subjects [57] and have been approved by FDA as a substitute for preclinical trials as the previous version [58].
In conclusion, the UVa/Padova T1D Simulator is the right tool to perform in silico clinical trials aimed to preliminary assess the clinical safety and feasibility of new treatments for the manual control of BG in T1D.

## 1.5 Beyond SF: Toward a personalized insulin therapy

In the last few years, the increasing availability of patient data gathered through wearable biosensors, such as CGM devices, enabled the creation of rich datasets in which CGM measurements have been integrated with other data of different "nature", e.g. insulin infusion recordings, and physical activity history, e.g. the Ohio dataset [59]. This unlocked many possibilities, such as the development of new personalized decision making tools and applications that aim to fit each patient physiology around their personal needs [57][52]. These strategies have the potential of greatly improving the outcome of current standard therapy and at the same time allow to reduce the burden brought by diabetes care to healthcare agencies which can potentially design specific prevention plans to save money and resources and at the same time increase treatment quality.
Following this rationale, "smarter" BCs can be designed the aim being personalizing the insulin dosing both at the individual and population level. Indeed, CGM data can be used to discover daily, weekly, and even seasonal pat-

terns that can be exploited to modulate meal-insulin bolus accordingly. Furthermore, insulin therapy outcomes can be used to fed cluster algorithms to automatically divide patients into different insulin sensitivity categories and, again, tune the insulin bolus amount accordingly, i.e. the more the insulin sensitivity associated to a specific subject, the smaller the insulin dose.
Given the complexity of the problem at hand, the non linear nature of glucose-insulin dynamics and the necessity of exploiting information coming from different domains, suitable techniques, able to implement data driven techniques to deliver personalized medical treatment, must be devised. With such a scope, methodologies belonging to the field of machine learning [60][61], e.g. linear regressions, neural networks and random forests just to mention a few, could be fed not only with features extracted from the CGM data stream, such as, the ROC information, but also other (easily accessible) therapy parameters, such as $CR$, $CF$, and the insulin pump basal rate, that implicitly reflect and characterize the individual patient physiology and have influence on the glycemic outcome during a meal [62]. Therefore, they appear natural candidates to constitute the backbone of the methodologies for insulin dosing personalization proposed in the remainder of the present thesis.

## 1.6 Aim and structure of the thesis

The main goal of this thesis is to improve the glycemic control achievable with the standard T1D insulin therapy by developing new personalized techniques for insulin dosing. In particular, we will build new data driven methodologies models that exploit both CGM data and easily accessible patient characteristics. To do so, in Chapter 2 we review the current literature techniques for the personalization of SF using CGM information (previously mentioned in Subsection 1.4.1) and perform in silico clinical trials to evaluate them both quantitatively and qualitatively. In Chapter 3 we preliminary show how, by adopting a NN-based algorithm fed with CGM information and easily accessible parameters, it is possible to adjust SF and improve the glycemic control obtained with the current literature techniques, thus demonstrating the potential of integrating machine learning methods in T1D treatment tracing back to machine learning. In Chapter 4, a different personalized approach to insulin dosing is proposed. Specifically, a new model based on eXtreme Gradient Boosted trees is used in a two step procedure: first an eXtreme Gradient Boosted tree model classifies, at meal time, the postprandial glycemic status using CGM data to-

gether with carbohydrate intakes data and insulin infusion recordings, then its outcome is used to proactively improve glycemic control through the real time modulation of SF. Then, in Chapter 5 we design a new formula for insulin bolus calculation based on linear regression models. In particular, our idea is to release the hypothesis of using SF as a starting point and adjusting it according to patient status and characteristics, and provide patients with new rules that naturally take into account for this information. To conclude, final considerations on the work carried out in this thesis as well as its future developments are discussed in Chapter 6.

# Chapter 2

# Literature methodologies for insulin dosing accounting for glucose rate of change: review and assessment

[1] In this chapter, we first present three popular methodologies that have been proposed in the literature by Buckingham et al. [44], Scheiner [45], and Pettus and Edelman [46] to account for ROC to adjust the insulin meal bolus amount calculated using SF. Then, by using a simulation framework, based on the UVA/Padova T1D Simulator, we assess their relative performance. The analysis will suggest how to orient our strategy to the solution of the insulin dosing personalization problem described in Chapter 3-5.

## 2.1 Literature algorithms for the adjustment of SF according to ROC

The method developed by Buckingham et al., hereafter label as BU, was firstly proposed in 2008 and consisted in modulating the insulin bolus amount calculated using SF according to ROC magnitude and direction. In practice, it consists of reducing/increasing insulin bolus by 10% when CGM sensor indicates that BG is falling/rising by 1-2 mg/dL/min. On the other hand, when CGM indicates that BG is falling/rising by 2-3 mg/dL/min or more, BU suggests to decrease/increase insulin bolus by 20%.

[1] This chapter contains material published in the article [63]

Formally, the computation of $IMB$ through BU can be formulated as:

$$IMB_{BU} = (\frac{CHO}{CR} + \frac{G_C - G_T}{CF} - IOB) \cdot (1 + f_{BU}(ROC)) \tag{2.1}$$

where $f_{BU}$ is a deterministic function depending on ROC. Specifically, in BU, $f_{BU}$ is defined as:

$$f_{BU}(ROC) = \begin{cases} 0.2 & ROC \geq 2 \text{ mg/dL/min} \\ 0.1 & 1 \leq ROC < 2 \text{ mg/dL/min} \\ 0 & -1 < ROC < 1 \text{ mg/dL/min} \\ -0.1 & -2 < ROC \leq -1 \text{ mg/dL/min} \\ -0.2 & ROC \leq -2 \text{ mg/dL/min} \end{cases} \tag{2.2}$$

A different approach is proposed by Scheiner (SC). Instead of modulating SF according to ROC, it suggests using ROC as a "predictor" to infer the future value of BG in the next 30-60 minutes under the assumption that such a ROC will be steady by the time the insulin starts to act. In practice, this translates in adjusting SF by substituting $G_C$ with its predicted value obtained by projecting $G_C$ 30-60 minutes ahead following the ROC direction and magnitude.
Specifically, SC suggests to subtract/add 25 mg/dL to $G_C$ when CGM indicates that BG is falling/rising by 1-3 mg/dL/min. Instead, if BG is falling/rising by more than 3 mg/dL/min, SC advise patients to subtract/add 50 mg/dL to the measured $G_C$.
Formally, the computation of $IMB$ through SC can be formulated as:

$$IMB_{SC} = \frac{CHO}{CR} + \frac{(G_C + f_{SC}(ROC)) - G_T}{CF} - IOB \tag{2.3}$$

where $f_{SC}$ is a deterministic function depending on ROC. Specifically, in SC, $f_{SC}$ is defined as:

$$f_{SC}(ROC) = \begin{cases} 50 & ROC \geq 2 \text{ mg/dL/min} \\ 25 & 1 \leq ROC < 2 \text{ mg/dL/min} \\ 0 & -1 < ROC < 1 \text{ mg/dL/min} \\ -25 & -2 < ROC \leq -1 \text{ mg/dL/min} \\ -50 & ROC \leq -2 \text{ mg/dL/min} \end{cases} \tag{2.4}$$

A strategy similar to that of SC is adopted in the method proposed by Pettus and Edelman, in the following denoted by PE. PE consists in subtracting/adding 50 mg/dL to $G_C$ if BG is falling/rising by 1 mg/dL/min. Moreover, PE subtracts/adds 75 mg/dL to the measured BG, when BG is changing by 2-3 mg/dL/min. Finally, PE suggests to decrease/increase the measured $G_C$ by 100 mg when CGM indicates that BG is falling/rising by more than 3 mg/dL/min.
Formally, the computation of $IMB$ through PE can be formulated as:

$$IMB_{PE} = \frac{CHO}{CR} + \frac{(G_C + f_{PE}(ROC)) - G_T}{CF} - IOB \tag{2.5}$$

where $f_{PE}$ is a deterministic function depending on ROC. Specifically, in PE, $f_{PE}$ is defined as:

$$f_{PE}(ROC) = \begin{cases} 100 & ROC \geq 2 \text{ mg/dL/min} \\ 50 & 1 \leq ROC < 2 \text{ mg/dL/min} \\ 0 & -1 < ROC < 1 \text{ mg/dL/min} \\ -50 & -2 < ROC \leq -1 \text{ mg/dL/min} \\ -100 & ROC \leq -2 \text{ mg/dL/min} \end{cases} \tag{2.6}$$

Such modifications of $G_C$, more aggressive than in SC, were motivated by recent studies investigating how T1D patients use CGM and ROC to make therapy decisions [64][65]. This study revealed that patients with T1D make make much larger corrections to insulin dose compared to both BU and SC. For example, when $G_C$ is 110 mg/dL and BG is rapidly rising (BG is increasing by more than 3 mg/dL/min) patients reported to increase the insulin dose amount computed using SF by 81% (instead of 20% suggested by BU).
Table 2.1 reports a summary of the methods specifics.

## 2.2 In silico assessment of the methodologies by the UVa/Padova T1D Simulator

### 2.2.1 Rationale

Assessing methodologies for determining insulin dosing is challenging. Indeed, as already discussed in Chapter 1, to fairly evaluate and compare the

**Table 2.1:** Comparison of the literature guidelines for the adjustment of SF based on ROC.

| ROC indication | BU | SC | PE |
|---|---|---|---|
| Constant: glucose is steady or is not decreasing/increasing by more than 1 mg/dL/min | $IMB^* = IMB$ | $G_C^* = G_C$ | $G_C^* = G_C$ |
| Slowly rising: glucose is rising by 1-2 mg/dL/min | $IMB^* = IMB + 10\%$ | $G_C^* = G_C + 25$ | $G_C^* = G_C + 50$ |
| Rising: glucose is rising by 2-3 mg/dL/min | $IMB^* = IMB + 20\%$ | $G_C^* = G_C + 25$ | $G_C^* = G_C + 75$ |
| Rapidly rising: glucose is rising by more than 3 mg/dL/min | $IMB^* = IMB + 20\%$ | $G_C^* = G_C + 50$ | $G_C^* = G_C + 100$ |
| Slowly falling: glucose is falling by 1-2 mg/dL/min | $IMB^* = IMB - 10\%$ | $G_C^* = G_C - 25$ | $G_C^* = G_C + 50$ |
| Falling: glucose is falling by 2-3 mg/dL/min | $IMB^* = IMB - 20\%$ | $G_C^* = G_C - 25$ | $G_C^* = G_C - 75$ |
| Rapidly falling: glucose is falling by more than 3 mg/dL/min | $IMB^* = IMB - 20\%$ | $G_C^* = G_C - 50$ | $G_C^* = G_C - 100$ |

effect of different insulin bolus adjustment strategies on glycemic control, it would be necessary to have data collected in the same patient, in the same time window, but with different insulin therapies, which is clearly impossible in real life. To circumvent this problem, simulations can be exploited to design an ad-hoc in silico clinical trial, where the same time window, with identical patient's physiological characteristics and behaviour can be reproduced.
Here, to evaluate the performance of BU, SC, and PE, and to compare them against SF, we resort to the 2013 version of the UVa/Padova T1D Simulator (hereafter indicated as S2013) [66], being, at the best of our knowledge the best available tool to perform such an analysis in a safe and cost efficient environment. In the following, details on the data and the simulation setup we used during the assessment are provided, as well as, the rationale that led us to the design of the in silico clinical trial.

### 2.2.2 Simulated dataset

Using S2013, we performed an in silico clinical trial in 100 virtual adults subjects (see Table 2.2 for a description of average and range of the key metabolic parameters of the virtual adult population enclosed in S2013). For each subject, we performed a single-meal simulation of 9-hours starting at 1:00 PM, where a meal of 50g of CHO was placed. In particular, each simulation was studied for different preprandial conditions in terms of BG and ROC values in order

**Table 2.2:** Main metabolic and demographic parameters of the virtual adult population of S2013 (taken from [66]).

| Parameter | Mean(SD) | Minimum | Maximum |
|---|---|---|---|
| Mean weight (kg) | 69.7 (12.4) | 46.7 | 106.1 |
| Total daily insulin (U/day/kg) | 0.61 (0.18) | 0.27 | 1.19 |
| CR (g/U) | 15.9 (5.3) | 7.2 | 29.5 |
| Fasting BG | 119.6 (6.7) | 107.5 | 137.8 |

to evaluate the impact of the adjustment of SF when different meal conditions are met. To do so, we tuned by trial-and-error the timing and amount of morning breakfast and snack and amount of the respective insulin boluses in the 6 hours preceding the start of the simulation to get the patient to the desired BG and ROC preprandial conditions. The resulting set of scenarios consists, for each patient, in 24 different preprandial conditions, i.e. the combination of four different ROC intervals (-3 to -2 mg/dL/min, -2 to -1 mg/dL/min, 1 to 2 mg/dL/min, and 2 to 3 mg/dL/min) and six BG intervals (60 $\pm$ 5, 70 $\pm$ 5, 80 $\pm$ 5, 100 $\pm$ 5, 150 $\pm$ 5, and 250 $\pm$ 5).

Regarding the other possible preprandial conditions, we decided not to generate scenarios having constant ROC (when glucose is steady or is not increasing/decreasing by more than 1 mg/dL/min) since in this particular case insulin bolus computed using SF is the same as BU, SC, and PE. Moreover, we did not generate scenarios having extreme ROC intervals (ROC lower than -3 mg/dL/min or greater than 3 mg/dL/min) since not possible to reproduce with S2013 with realistic manipulations of time, carbohydrate content, and insulin bolus of morning breakfast and snack.

To assess the performance of each BC method per se, we decided to limit the exposition of the study to possibly confounding factors (e.g., patient behavior in making treatment decisions to mitigate hyper/hypoglycemia, changes in individual insulin sensitivity, low or high preprandial insulin/carbohydrate on board amount, CGM errors and artifacts that can also alter the estimation of the ROC value) and run the simulations in a noise free environment, that is, using optimal therapy parameters, allowing neither postprandial correction boluses nor hypotreatments, without simulating errors in meal CHO counting, BG measurement and ROC estimation.

To better grasp the dataset composition, two simulated examples are described below. Figure 2.1 shows the BG and ROC traces obtained for the virtual subject adult#1 when the target preprandial conditions at 1:00 PM are BG 100 $\pm$

5 mg/dL and ROC -2 to -1 mg/dL/min. From visual inspection, it can be observed that a morning breakfast of 82.5 g of CHO (5.5 g of CHO equally distributed over 15 minutes) and an insulin bolus of 17 U have been placed at around 10:15 and 10:30 AM, respectively. These events allowed the BG concentration and trend to reach the desired target range (here highlighted in yellow) at lunch time, i.e. 1:00 PM. A different simulated scenario is shown in Figure

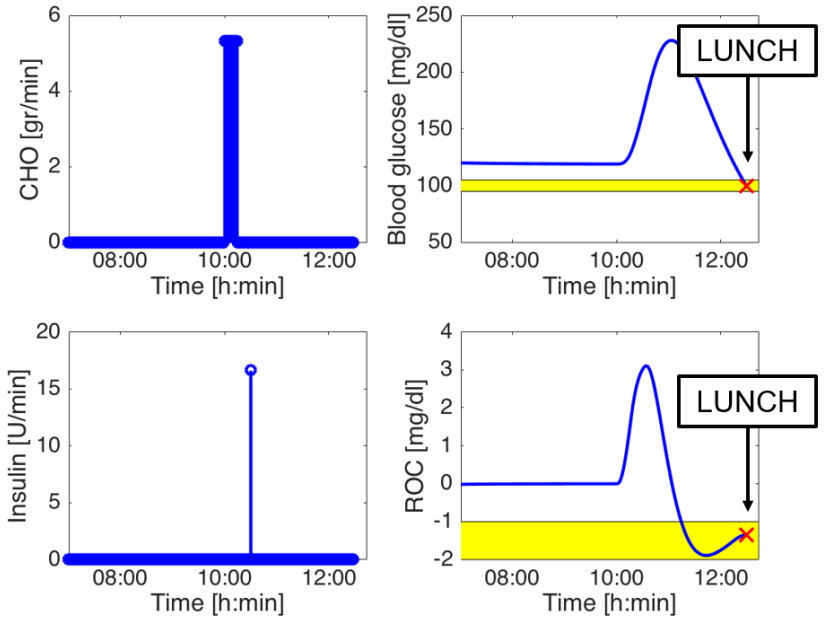



**Figure 2.1:** Representative simulated scenario for adult#1 when preprandial BG and ROC target are 100 mg/dL and -2 to -1 mg/dL/min. Left panels, show the CHO intake and insulin infusion placed within the morning time window in order to achieve the target preprandial conditions (here highlighted in yellow). Right panels show the obtained BG and ROC traces respectively. Red crosses represent the BG and ROC values at lunch time, i.e. 1:00 PM.

2.2. BG and ROC traces are shown for adult#1 when the target preprandial conditions at 1:00 PM are BG 70 $\pm$ 5 mg/dL and ROC 1 to 2 mg/dL/min. Here, a morning breakfast of 30 g CHO (2 g of CHO equally distributed over 15 minutes), a snack of 20 g of CHO (4 g of CHO equally distributed over 5 minutes) and insulin bolus of 9 U, have been placed at 7:00 AM, 11:30 AM and 7:10 AM, respectively, in order to meet the target BG and ROC values at lunch time.

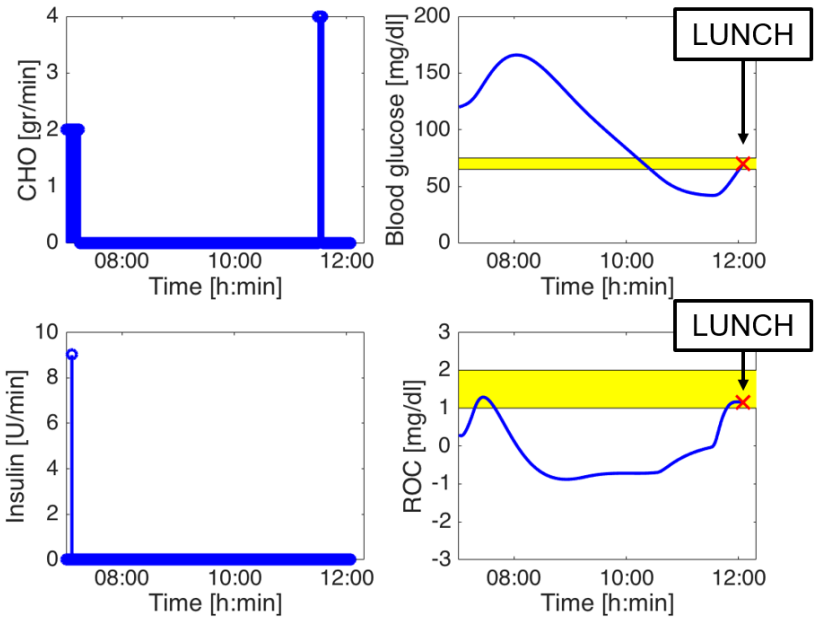



**Figure 2.2:** Representative simulated scenario for adult#1 when preprandial BG and ROC target are 70 mg/dL and 1 to 2 mg/dL/min. Left panels, show the CHO intake and insulin infusion placed within the morning time window in order to achieve the target preprandial conditions (here highlighted in yellow). Right panels show the obtained BG and ROC traces respectively. Red crosses represent the BG and ROC values at lunch time, i.e. 1:00 PM.

## 2.3 Criteria for performance evaluation and statistical analysis

### 2.3.1 Performance evaluation

For each subject and each (BG, ROC) pairs, we compared the BG profile obtained using the SF with the BG profiles obtained from the adoption of BU, SC, and PE. Note that, for each particular meal condition scenario, we discarded those virtual subjects whose insulin bolus computed with SF resulted to be zero or negative (e.g., because of too much elevated IOB values), since in such a situation the methods would not be comparable. The quality of glucose control was assessed, first of all, by calculating the blood glucose risk index (BGRI) in the postprandial 9-h time window. Briefly, this metric, thanks to the glucose scale symmetrization, is equally sensitive to hypoglycemia and hyperglycemia and condenses glycemic excursions in a single quantity, facilitating interpretation and comparison of the results. In details, the BGRI, first introduced by Kovatchev et al. [67], is composed by the sum of 2 terms: the Low Blood Glu-

cose Index (LBGI) and the High Blood Glucose Index (HBGI). The definition of LBGI and HBGI derives from a logarithmic transformation of the BG scale that balances the amplitude of hypo-/hyperglycemic ranges (enlarging the former and shrinking the latter) and makes the transformed data symmetric around zero and fitting a normal distribution [68]. This symmetrization is performed because in the standard BG scale, hypoglycemia (BG < 3.9 mmol/l) and hyperglycemia (BG > 10 mmol/L) have very different ranges, and euglycemia is not central in the entire blood glucose range (1.1-33.3 mmol/L). Consequently, the scale is not symmetric and its clinical center (6-7 mmol/L) is far from its numerical center (17 mmol/L). As a result, a logarithmic data transformation that matches the clinical and numerical center of the BG scale have been applied, thus making the transformed data symmetric [69]. If BG measurements are expressed in mg/dl, the transformed data $f(BG)$ are given by:

$$f(BG) = 1.509 \cdot [(\log BG)^{1.084} - 5.381] \tag{2.7}$$

and are used to define a BG risk function $r(BG)$ as:

$$r(BG) = 10 \cdot f(BG)^2 \tag{2.8}$$

that associates to each BG reading a measure of its risk, as expressed with a number in the 0 to 100 range. For the definition of this metric, the risk index of equation 2.8 is used and thus the LBGI and HBGI are computed and added in order to generate the overall BGRI as follows:

$$LBGI = \frac{1}{N}\sum_{i=1}^{N} r(BG(i))\mathbb{1}(f(BG(i)) < 0) \qquad [dimensionless] \tag{2.9}$$

$$HBGI = \frac{1}{N}\sum_{i=1}^{N} r(BG(i))\mathbb{1}(f(BG(i)) > 0) \qquad [dimensionless] \tag{2.10}$$

$$BGRI = LBGI + HBGI \qquad [dimensionless] \tag{2.11}$$

Finally, to give a picture of the relative performance of the three methods more focused on hypoglycemia, we computed, for each subject, LBGI. Specifically, we considered LBGI obtained using the methods BU, SC, and PE under comparison and the LBGI obtained using SF, and we calculated the respective average ratio.

### 2.3.2 Statistical analysis

To evaluate the between-methods differences, a two-tailed Wilcoxon signed-rank test with 1% significance was performed. In particular, we applied the Bonferroni correction for multiple comparisons, thus dividing the significance level by the total number of comparisons we made. To assess whether or not some patient characteristics can impact the outcome, for each scenario we computed the Spearman's correlation coefficient between body weight (BW), total daily insulin (TDI), basal insulin infusion rate ($I_b$), $G_T$, $CR$, $CF$, IOB, and the BGRI difference ($\Delta$BGRI) between BU versus SF, SC versus SF, and PE versus SF. The statistical significance of the obtained correlation was evaluated using as significance level $\alpha$ = 1%.

## 2.4 Results

Figure 2.3 presents two examples of BG traces obtained using both ROC-modified BC literature methods and SF. Top panel represents adult#6 with preprandial BG 100 $\pm$ 5 mg/dL and ROC -3 to -2 mg/dL/min. In this case adjusting the original insulin bolus amount according to ROC allows to achieve a better glucose control both during and after the meal, reducing and even zeroing (when using PE) time spent in hypoglycemia. In particular, PE correction is the one that performs best (BGRI(SF) = 6.39, BGRI(BU) = 5.00, BGRI(SC) = 3.84, BGRI(PE) = 1.81). The second example, reported in the bottom panel, represents adult 20 with preprandial BG 150 $\pm$ 5 mg/dL and ROC 2 to 3 mg/dL/min. This case presents worse glycemic outcomes compared to SF when applying either SC, PE, and BU correction (BGRI(SF) = 7.80, BGRI(BU) = 13.86, BGRI(SC) = 9.14, BGRI(PE) = 14.93) which drive the virtual subject to hypoglycemia. Below, for each of the considered ROC intervals, we analyze and compare the obtained glycemic outcomes resulting from the adoption of SF, BU, SC, and PE in the virtual population.

### 2.4.1 ROC -3 to -2 mg/dL/min

In Figure 2.4A, the distributions of the $\Delta$BGRI between BU versus SF, SC versus SF, and PE versus SF obtained in all the 100 virtual subjects when preprandial ROC is -3 to -2 mg/dL/min are shown via boxplot representation. Ac-

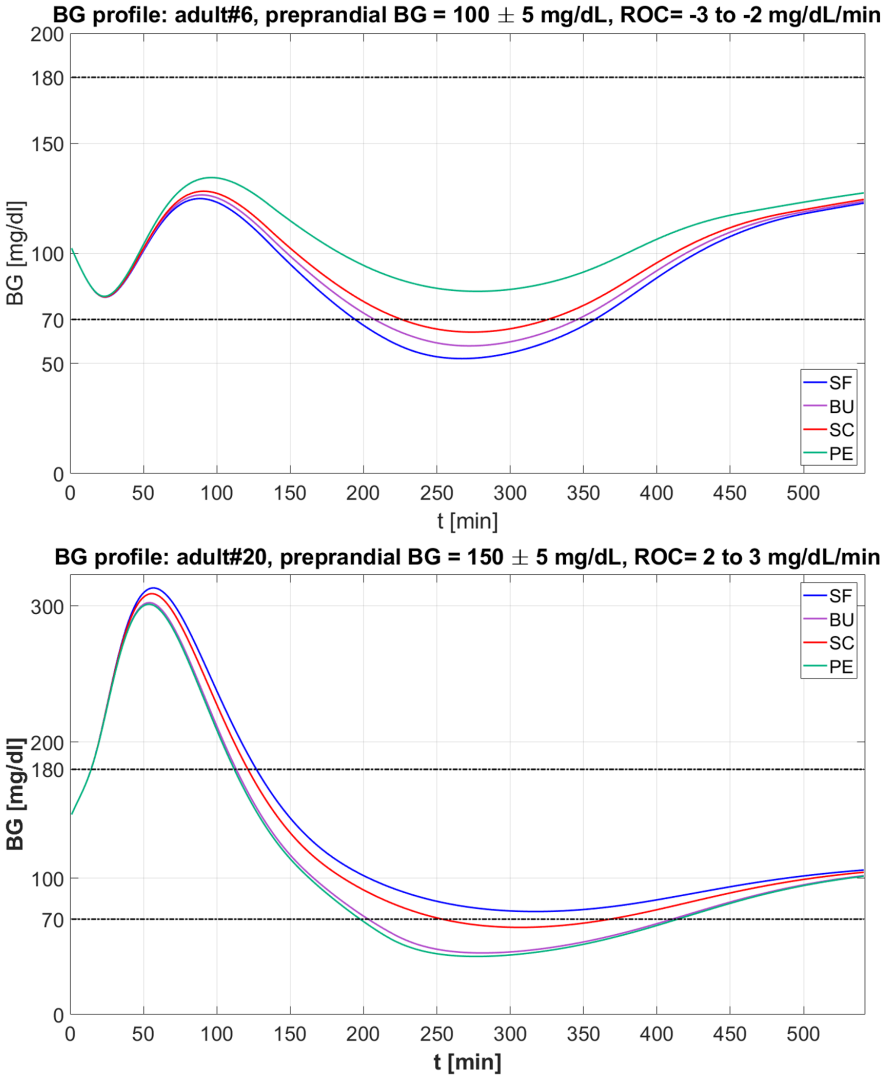



**Figure 2.3:** Example of BG trace obtained with SF (in blue), BU (in violet), SC (in red), PE (in green), and BU (in violet) methods. Top panel. BG profiles obtained for the virtual subject adult 6 when preprandial ROC is -3 to -2 mg/dL/min and preprandial BG is 100 ± 5 mg/dL. Bottom panel. BG profiles obtained for the virtual subject adult 20 when ROC is 2 to 3 mg/dL/min and preprandial BG is 150 ± 5 mg/dL.

cording to Figure 2.4A, all methods improve significantly the BGRI ($p < 0.01$) obtained with SF. This is visible also in Table 2.3, where we report the median/interquartile range of the BGRI distribution obtained using SF, BU, SC, and PE. However, the boxplots in Figure 2.4A also evidence that several subjects obtained worse glycemic outcomes when the dose is corrected according to either BU, SC, or PE compared to SF. In Figure 2.5A, we report the boxplot representation of the ΔBGRI distributions obtained by comparing SC versus BU, PE versus BU, and PE versus SC when preprandial ROC is -3 to -2 mg/dL/min. Indeed, except for preprandial BG 150 ± 5 mg/dL where all

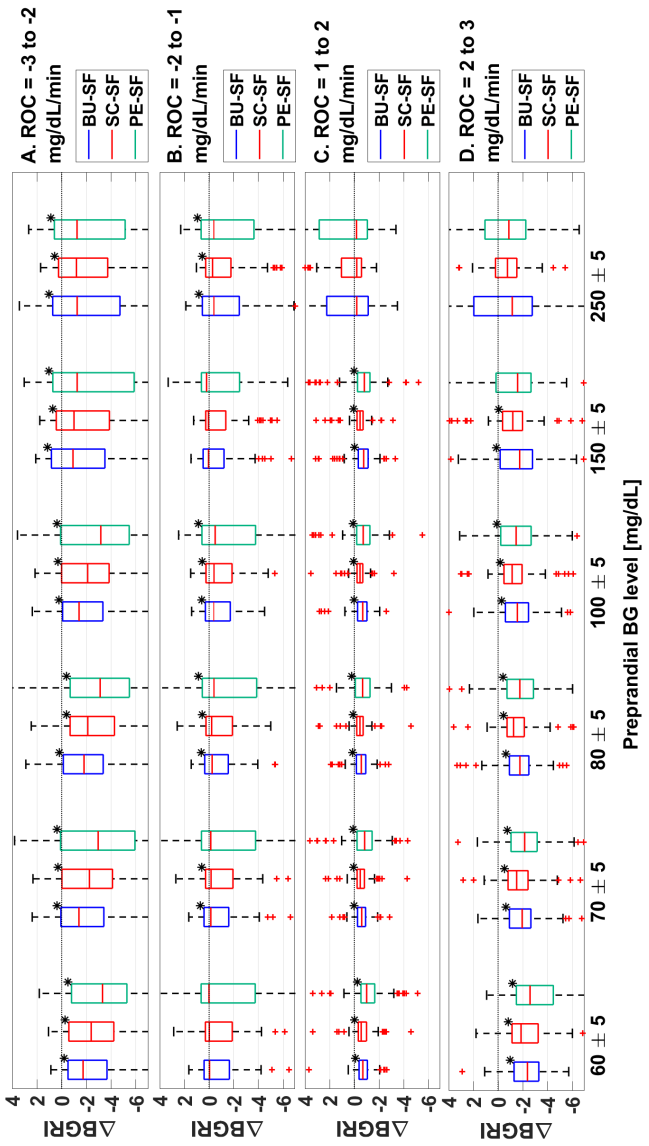



**Figure 2.4:** Boxplot representation of the distribution of ΔBGRI obtained comparing BU versus SF (in blue), SC vs SF (in red), PE versus SF (in green) when ROC is -3 to -2 mg/dL/min (panel A), -2 to -1 mg/dL/min (panel B), 1 to 2 mg/dL/min (panel C), and 2 to 3 mg/dL/min (panel D). Red horizontal lines represent median, boxes mark interquartile ranges, dashed lines are the whiskers, red crosses indicate outliers. Black stars indicate statistical significant between-distribution differences. Lower ΔBGRI means better glucose control quality.

methods achieve equivalent performance, PE obtained significantly ($p < 0.01$) better BGRI values compared to both BU and SC. Furthermore, as visible in Table 2.3A, comparing SC versus BU, better BGRI are achieved using BU. As far as the correlation analysis is concerned, comparing BU, SC, and PE with the SF, we find that, when ROC is -3 to -2 mg/dL/min, the respective $\Delta$BGRI are weakly but significantly correlated ($p < 0.01$) with $CR$ and IOB. No significant correlation has been found for the other considered patients' parameters. Table 2.4A reports the average ratio between the LBGI obtained using the considered methods and the LBGI obtained using the SF when ROC is -3 to -2 mg/dL/min. Results show that, as expected, the LBGI decreases using BU, SC, and PE, as the computed insulin bolus amount decreases as well. In particular, PE achieves lower ratios, being the most conservative.

### 2.4.2 ROC -2 to -1 mg/dL/min

Figure 2.4B shows the distributions of $\Delta$BGRI obtained by comparing the considered methods against SF when preprandial ROC is -2 to -1 mg/dL/min. In this case, the outcomes are strongly dependent on the specific preprandial BG level. Indeed, although all methods allowed obtaining statistically significant better performance than SF for preprandial BG 80 $\pm$ 5, 100 $\pm$ 5 and 250 $\pm$ 5 mg/dL, there are no statistically significant ($p < 0.01$) differences when comparing BU, SC, and PE versus SF for preprandial BG 60 $\pm$ 5 and 150 $\pm$ 5 mg/dL, and PE versus SF for preprandial BG 70 $\pm$ 5 mg/dL. Considering the median results (see Table 2.3), the best outcomes were achieved without adjusting the reference insulin bolus when preprandial BG is 150 $\pm$ 5 mg/dL and using PE for the other preprandial BG conditions. As shown in Figure 2.4B, the fact that the considered methods do not perform better than SF is due to the presence of a non-negligible number of subjects whose BGRI is higher when using BU, SC, and PE. Moreover, such a worsening of BGRI is more pronounced if larger corrections are applied to SF, like using the PE guideline. Indeed, the bigger the correction of SF, the higher the variability obtained in terms of BGRI. Figure 2.5B shows the $\Delta$BGRI distributions obtained comparing the considered methods between each other. In these scenarios we cannot detect any statistically significant difference between the methods. Furthermore, the correlation analysis obtains same qualitative results as ROC -3 to -2 mg/dL/min. Specifically, comparing BU versus SF, SC versus SF, and PE versus SF when ROC is -2 to -1 mg/dL/min, we find weak but significant correla-

**Table 2.3:** Median [interquartile range] BGRI results obtained for different preprandial BG level when ROC is -3 to -2 mg/dL/min (Panel A), -2 to -1 mg/dL/min (Panel B), 1 to 2 mg/dL/min (Panel C), and 2 to 3 mg/dL/min (Panel D).

| | | **Preprandial BG (mg/dL)** | | | | | |
|---|---|---|---|---|---|---|---|
| | | $60 \pm 5$ | $70 \pm 5$ | $80 \pm 5$ | $100 \pm 5$ | $150 \pm 5$ | $250 \pm 5$ |
| **(A) ROC = -3 to -2 mg/dL/min** | SF | 14.72 [6.85-27.47] | 12.84 [4.86-21.67] | 13.91 [5.09-31.16] | 12.88 [5.03-30.58] | 8.03 [5.12-21.45] | 10.91 [7.93-21.71] |
| | BU | 11.45 [5.35-22.88] | 8.63 [5.13-18.49] | 9.25 [4.84-26.15] | 9.10 [4.97-26.94] | 7.86 [5.17-15.92] | 10.53 [7.44-16.26] |
| | SC | 12.23 [5.54-23.84] | 7.95 [3.98-18.70] | 10.09 [4.58-26.19] | 9.01 [4.97-26.66] | 7.77 [5.04-17.85] | 10.29 [7.28-17.67] |
| | PE | 10.95 [5.18-20.86] | 7.55 [4.00-17.00] | 9.18 [4.49-24.59] | 8.36 [5.01-25.16] | 7.70 [4.97-15.14] | 9.80 [7.22-15.28] |
| **(B) ROC = -2 to -1 mg/dL/min** | SF | 5.51 [3.77-13.07] | 5.88 [3.74-13.72] | 5.70 [3.75-15.57] | 6.50 [3.95-20.64] | 5.92 [4.61-9.62] | 10.56 [7.95-15.06] |
| | BU | 5.39 [3.96-10.09] | 5.60 [3.77-10.85] | 5.68 [3.62-13.08] | 6.75 [4.08-18.15] | 6.26 [4.23-8.96] | 9.96 [7.92-14.51] |
| | SC | 5.20 [4.03-10.02] | 5.68 [3.87-11.50] | 5.61 [3.65-12.41] | 6.23 [3.90-17.35] | 6.07 [4.31-8.61] | 10.31 [7.81-14.00] |
| | PE | 5.15 [4.03-9.70] | 5.31 [3.89-9.61] | 5.31 [3.81-10.40] | 5.74 [4.11-13.95] | 6.13 [4.34-8.38] | 9.90 [7.95-13.67] |
| **(C) ROC = 1 to 2 mg/dL/min** | SF | 9.15 [6.81-12.28] | 8.07 [6.27-10.99] | 7.52 [5.82-10.95] | 8.98 [6.56-11.35] | 10.16 [8.25-12.60] | 14.30 [11.19-20.77] |
| | BU | 8.45 [6.36-11.78] | 7.54 [5.57-10.15] | 7.21 [5.57-9.65] | 8.51 [6.21-10.59] | 9.79 [7.57-12.93] | 14.26 [11.19-20.77] |
| | SC | 8.38 [6.42-12.05] | 7.66 [5.84-10.38] | 7.32 [5.66-10.03] | 8.59 [6.19-10.98] | 9.85 [7.77-13.56] | 14.49 [10.86-21.82] |
| | PE | 8.17 [6.21-11.28] | 7.63 [5.52-9.82] | 7.40 [5.27-9.93] | 8.53 [6.02-10.68] | 9.67 [7.63-13.39] | 14.28 [11.47-21.63] |
| **(D) ROC = 2 to 3 mg/dL/min** | SF | 16.99 [11.85-22.01] | 13.11 [9.81-17.89] | 12.34 [9.24-17.47] | 11.03 [8.40-16.21] | 13.65 [9.52-18.71] | 16.89 [12.70-20.80] |
| | BU | 14.58 [10.70-19.21] | 11.48 [8.49-15.42] | 10.66 [8.15-15.00] | 9.93 [7.57-14.98] | 12.88 [8.99-18.21] | 16.43 [12.24-21.39] |
| | SC | 14.72 [10.56-19.05] | 11.24 [8.59-15.46] | 10.37 [8.37-15.18] | 9.89 [7.82-15.77] | 13.54 [8.78-19.38] | 16.54 [12.81-21.65] |
| | PE | 14.14 [10.36-18.09] | 11.08 [8.41-15.69] | 10.94 [8.08-15.98] | 9.96 [7.60-15.25] | 13.11 [9.38-19.55] | 15.95 [12.58-23.27] |

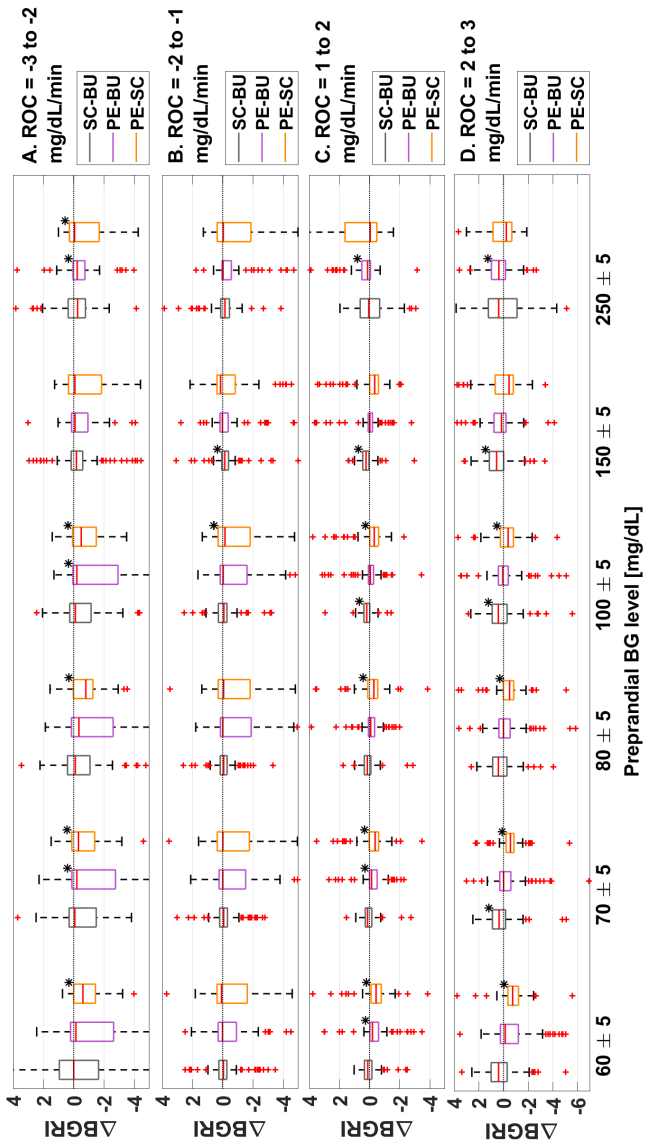



**Figure 2.5:** Boxplot representation of the distribution of ΔBGRI obtained comparing SC versus BU (in gray), PE versus BU (in violet), and PE versus SC (in green) when ROC is -3 to -2 mg/dL/min (panel A), -2 to -1 mg/dL/min (panel B), 1 to 2 mg/dL/min (panel C), and 2 to 3 mg/dL/min (panel D). Red horizontal lines represent median, boxes mark interquartile ranges, dashed lines are the whiskers, red crosses indicate outliers. Black stars indicate statistical significant between-distribution differences. Lower ΔBGRI means better glucose control quality.

**Table 2.4**: Mean (SD) ratio between the LBGI obtained using the considered methods (BU,SC, and PE) and the LBGI obtained using the SF. Results are obtained for different preprandial BG level when ROC is -3 to -2 mg/dL/min (Panel A), -2 to -1 mg/dL/min (Panel B), 1 to 2 mg/dL/min (Panel C), and 2 to 3 mg/dL/min (Panel D).

| | | **Preprandial BG (mg/dL)** | | | | | |
|---|---|---|---|---|---|---|---|
| | | $60 \pm 5$ | $70 \pm 5$ | $80 \pm 5$ | $100 \pm 5$ | $150 \pm 5$ | $250 \pm 5$ |
| **(A) ROC = -3 to -2 mg/dL/min** | BU-SF | 0.84 (0.16) | 0.83 (0.19) | 0.83 (0.20) | 0.85 (0.18) | 0.78 (0.22) | 0.70 (0.24) |
| | SC-SF | 0.83 (0.16) | 0.82 (0.18) | 0.81 (0.20) | 0.84 (0.17) | 0.80 (0.19) | 0.73 (0.22) |
| | PE-SF | 0.77 (0.20) | 0.77 (0.22) | 0.76 (0.24) | 0.79 (0.21) | 0.74 (0.24) | 0.65 (0.27) |
| **(B) ROC = -2 to -1 mg/dL/min** | BU-SF | 0.88 (0.14) | 0.87 (0.14) | 0.87 (0.14) | 0.87 (0.14) | 0.83 (0.17) | 0.76 (0.18) |
| | SC-SF | 0.88 (0.15) | 0.87 (0.15) | 0.87 (0.14) | 0.87 (0.14) | 0.83 (0.17) | 0.81 (0.14) |
| | PE-SF | 0.79 (0.23) | 0.78 (0.23) | 0.78 (0.22) | 0.79 (0.21) | 0.73 (0.25) | 0.68 (0.23) |
| **(C) ROC = 1 to 2 mg/dL/min** | BU-SF | 1.13 (0.21) | 1.17 (0.26) | 1.21 (0.32) | 1.21 (0.30) | 1.33 (0.40) | 1.56 (0.61) |
| | SC-SF | 1.13 (0.24) | 1.16 (0.27) | 1.19 (0.31) | 1.19 (0.32) | 1.25 (0.40) | 1.25 (0.32) |
| | PE-SF | 1.44 (0.84) | 1.52 (0.93) | 1.61 (1.06) | 1.62 (1.10) | 1.78 (1.48) | 1.64 (0.90) |
| **(D) ROC = 2 to 3 mg/dL/min** | BU-SF | 1.19 (0.49) | 1.27 (0.54) | 1.42 (0.76) | 1.53 (0.83) | 1.93 (1.34) | 2.73 (2.16) |
| | SC-SF | 1.21 (0.55) | 1.29 (0.70) | 1.40 (0.84) | 1.47 (0.95) | 1.72 (1.47) | 1.62 (0.86) |
| | PE-SF | 1.50 (1.20) | 1.75 (1.75) | 1.97 (1.12) | 2.10 (1.55) | 2.52 (2.18) | 2.21 (1.89) |

tion ($p < 0.01$) between $\Delta$BGRI and $CR$ and IOB. Table 2.4B reports the average ratio between the LBGI obtained using the considered methods and the LBGI obtained using the SF when ROC is -2 to -1 mg/dL/min. Results show same qualitative outcomes as ROC -3 to -2 mg/dL/min. Specifically, average LBGI decreases when BU, SC, and PE are adopted.

### 2.4.3 ROC 1 to 2 mg/dL/min

Figure 2.4C reports, via boxplot representation, the $\Delta$BGRI distributions obtained comparing BU, SC, and PE versus SF. In this case, adjusting insulin bolus according to ROC we almost always obtained better results in terms

of glycemic control. In particular, best median BGRI values (see Table 2.3) are achieved using PE for preprandial BG 60 $\pm$ 5 and 150 $\pm$ 5 mg/dL, BU for preprandial BG 70 $\pm$ 5, 80 $\pm$ 5, and 250 $\pm$ 5 mg/dL and SC for preprandial BG 100 $\pm$ 5 mg/dL. However, when preprandial BG is 250 $\pm$ 5 mg/dL there is no statistically significant difference between BU, SC, and PE and SF, due to a substantial degradation of the glucose control performance in almost half of the considered subjects when a correction is applied to SF. Focusing on the between-methods comparison (see Figure 2.5C), results suggest that, when preprandial BG is 150 $\pm$ 5 mg/dL, the best outcomes are achieved by the PE indication. On the other hand, when preprandial BG is 250 $\pm$ 5 mg/dL, the best results are obtained using BU. Notably, SC presents worse BGRI results compared to both BU and PE, in all of the considered scenarios. Furthermore, as reported in Table 2.4C, comparing SC versus BU, better BGRI are achieved using the latter. Concerning the correlation analysis results, comparing BU versus SF, SC versus SF, and PE versus SF when ROC is 1 to 2 mg/dL/min, we find that $\Delta$BGRI is significantly correlated ($p < 0.01$) with BW and $G_T$. No significant correlation has been found for the other considered patients' parameters. Finally, in Table 2.4C we report the average ratio between the LBGI obtained using the considered methods and the LBGI obtained using the SF when ROC is 1 to 2 mg/dL/min. Again, these results are not surprising, since BU, SC, and PE increase the insulin bolus amount when ROC > 0 mg/dL/min. Specifically, higher ratios are obtained using the PE method, since it implements the most aggressive adjustment of the SF.

### 2.4.4 ROC 2 to 3 mg/dL/min

Figure 2.4D shows the $\Delta$BGRI distributions obtained by comparing the considered correction methods against SF. Results show that adjusting the reference insulin bolus according to ROC significantly ($p < 0.01$) improved, on average, the glycemic outcomes. In particular, best median BGRI results (see Table 2.3) are achieved using PE for preprandial BG 60 $\pm$ 5, 70 $\pm$ 5, and 250 $\pm$ 5 mg/dL, SC for preprandial BG 80 $\pm$ 5 and 100 $\pm$ 5 mg/dL and BU for preprandial BG 150 $\pm$ 5 mg/dL. However, there are no statistically significant differences comparing BU, SC, and PE versus SF when preprandial BG is 250 $\pm$ 5 mg/dL, since correcting SF leads to a degradation of the glycemic outcomes in a consistent part of the population. In Figure 2.5D we show the $\Delta$BGRI distributions of SC versus BU, PE versus BU, and PE versus SC via boxplot representation.

Results show that BU and PE get almost the same performance and outperform SC when preprandial BG is 150 $\pm$ 5 mg/dL. On the other hand, when preprandial BG is 250 $\pm$ 5 mg/dL, PE and SC are not significantly different from each other and the best results are achieved using BU. Finally, as far as the correlation analysis is concerned, we find that, comparing BU, SC, and PE with the SF when ROC is 2 to 3 mg/dL/min, $\Delta$BGRI is significantly correlated ($p < 0.01$) with BW, $G_T$ and $CR$. Table 2.4D reports the average ratio between the LBGI obtained using the considered methods and the LBGI obtained using the SF when ROC is 2 to 3 mg/dL/min. Results show the same qualitative results as ROC 1 to 2 mg/dL/min.

## 2.5 Outcome of the assessment and indications for new developments

Our results showed that, overall, none of the considered approaches clearly prevails on the others, since the best modulation of the insulin bolus results strongly related to preprandial conditions. In the scenarios in which preprandial BG is rapidly decreasing (ROC -3 to -2 mg/dL/min), we observed the best performance for PE which allowed reducing the insulin bolus amount more than the other methods, thus resulting in better glycemic outcome. When BG is slightly decreasing (ROC -2 to -1 mg/dL/min), none of the three "dynamic" methods is superior for all the preprandial BG levels tested. The same comment applies to positive ROC scenarios, in which in particular no statistically significant BGRI improvement was obtained by either SC, BU and PE compared to SF when BG was 250 $\pm$ 5 mg/dL. Finally, our results showed that none of the methods to correct SF according to ROC perform well in all the subjects. Indeed, even if all the methods achieved significantly better results in median BGRI compared to SF, this improvement was not achieved for all the subjects, with several individuals obtaining worse glycemic outcomes when either BU, SC, or PE were applied. In particular, our results show an important limitation of BU, SC, and PE. In fact, when ROC is positive, the risk for hypoglycemia systematically increases even if, following the intuition, the increased amount of administered insulin should compensate the positive ROC without inducing hypoglycemia.

Our results also demonstrated the existence of a weak, but significant, correlation between the obtained glycemic outcomes and some of the patient's

specific parameters. Specifically, our analysis suggests that, when ROC is negative, the insulin bolus dose should be modulated exploiting also $CR$ and the IOB amount instead of just ROC. On the other hand, when ROC is positive, results indicates that BW, $G_T$ and $CR$ should also be accounted in the adjustment of the SF.

In conclusion, the assessment of BU, SC, and PE suggests that using CGM trend information for insulin dosing requires particular attention. Indeed, although having information about ROC can potentially improve the current SF for insulin bolus calculation, the rules for integrating ROC information in BC must be carefully devised and comprehensively assessed to ensure their safety. Therefore, further investigations are needed to design new BCs able to effectively take advantage of the "dynamic" information on BG available from CGM. In particular, new methodologies should personalize the correction of SF taking into account also the subject's characteristics.

Following the lessons learnt in the present chapter, we decided to explore novel insulin bolus adjustment approaches based on tools such as neural networks, extreme gradient boosted trees, and multiple linear regression, tracing back to machine learning in order to develop new methodologies able to take into account also patient's specific parameters and current physiological status.

# Chapter 3

# Development of a neural network based algorithm to personalize the insulin dosing

[1] In the following, to surpass the limitations of existing literature discussed in the previous chapter, we propose a novel strategy based on a neural network (NN) for the adjustment of the insulin bolus dose provided by SF that, instead of accounting for ROC only as in BU, SC, and PE, exploit additional information regarding patient current physiological status.

## 3.1 Rationale

As presented in Chapter 2, BU, SC, and PE, consist in increasing or reducing the insulin bolus computed with SF as a function of ROC. In particular, focusing on SC and PE, both methods integrate SF with a correction based on a prediction of future BG by computing B as:

$$IMB = \frac{CHO}{CR} + \frac{(G_C + f(\mathbf{X})) - G_T}{CF} - IOB \tag{3.1}$$

where $f(\cdot)$ is a deterministic function ranging from -100 to 100 mg/dL depending on a set of variables $\mathbf{X}$ (here $\mathbf{X}$ = {ROC}). Notably, the main limitation of these two methods is that the ROC-based adjustment is equal for all individuals and for all preprandial BG level. A possible margin of improvement is to develop methods to design a function $f(\cdot)$, in Equation 3.1, able to take into ac-

[1] This chapter contains material published in the article [70]

count simultaneously some individual parameters of the patient (i.e., $CF$, $CR$, insulin pump basal infusion rate, $G_T$, body weight, insulin sensitivity) as well as the state (i.e., $G_C$, $CHO$, ROC, IOB) at the time of the bolus.
Therefore, the leading idea here is to develop a NN-based "corrector", hereafter labeled as NNC, to determine how to optimally correct the insulin bolus amount calculated with SF by exploiting the information on patient's CHO intake, preprandial conditions and the aforementioned individual parameters. The NNC will be developed and assessed by using a simulated dataset described in the following section.

## 3.2 Simulated dataset

To develop and assess NNC, we used the same dataset simulated for the assessment of BU, SC, and PE, which consist in data of 100 virtual adult subjects each simulated several times in a single meal noise-free scenario with different conditions at meal time in terms of BG and ROC where, for the sake of simplicity and practicality, ROC values at meal time are discretized as in Figure 3.1. See Section 2.2.2 for more details on how the dataset was built.

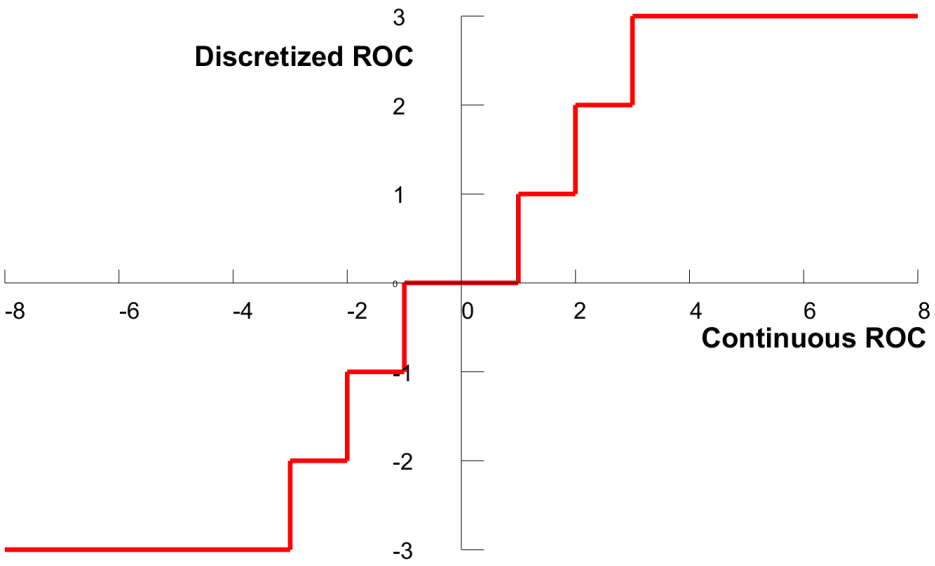



**Figure 3.1:** ROC discretization. Specifically, ROC was set to -3 if ROC $\leq$ -3 mg/dL/min, -2 if -3 < ROC $\leq$ -2 mg/dL/min, -1 if -2 < ROC $\leq$ -1 mg/dL/min, 0 if -1 < ROC < 1 mg/dL/min, 1 if 1 $\leq$ ROC < 2 mg/dL/min, 2 if 2 $\leq$ ROC < 3, and 3 if ROC $\geq$ 3.

In order to determine how to determine how to correct insulin bolus amount calculated by SF, it is important to have an ideal reference. To provide such a reference, let's consider a specific patient $p$ and a meal condition $m$ (that is, meal amount, preprandial BG and ROC values). The problem is to identify the "optimal" $f(\cdot) = Y_{pm}$ correction value to be applied to SF such that it allows to achieve the best glycemic outcome, quantified here by BGRI. For this purpose, given $p$ and $m$, we analyzed multiple simulations where we computed the meal insulin bolus using Equation 3.1 with different values of $Y_{pm}$ chosen from an equally spaced manually specified grid of values $F := \{-200 + 10 \cdot k \quad | \quad k = 0, ..., 40\}$. As a result, for each $p$ and $m$, we obtained a total of 41 BG profiles that we quantitatively evaluate by calculating the respective BGRI, i.e. $BGRI_{pm}$. Then, we set the "optimal" correction $Y_{pm}$ as the value of $F$ associated to the minimum $BGRI_{pm}$. In particular, if two or more $Y_{pm}$ allow to get the minimum $BGRI_{pm}$, $Y_{pm}$ is defined as the most "conservative" correction $f(\cdot)$, that is, the closer value to 0. Notably, some $Y_{pm}$ are equal to $\pm$ 200 due to the fact that $F$ is finite. We decided to remove those cases from the dataset to get it loose from the $F$ definition. As a result, the final dataset is composed by a total of 9963 records, identified by $p$, $m$ and the respective $Y_{pm}$ . It is important to remark that the set of target $Y_{pm}$ values together with Equation 3.1 represent a sort of "optimal" bolus calculator, hereafter labeled as OPT. The OPT correction will be used both in the training of the NNC and, as a reference, in the evaluation of the methods.
Figure 3.2 shows the optimal correction $Y_{pm}$ values, as a function of preprandial BG and ROC values, obtained for a selected patient (adult#1) and meal CHO intake (50 g). It is possible to observe that the optimal correction value is strongly dependent from the specific preprandial condition and far from the adjustments proposed by both SC and PE.

## 3.3 The new NN-based insulin bolus calculator formula

The chosen NNC structure is summarized in Figure 3.3A. See Appendix A for a detailed description on neural networks, how to choose their structure, as well as their training procedure. It consists of a feed-forward fully connected NN composed of three hidden layers. Network structure has been chosen through a preliminary investigation on the training set data, following the rationale

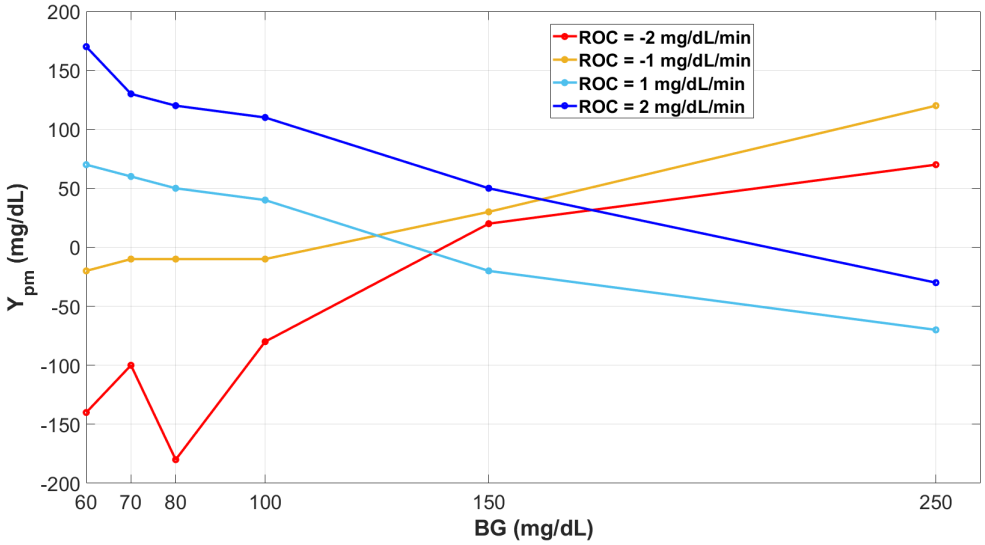
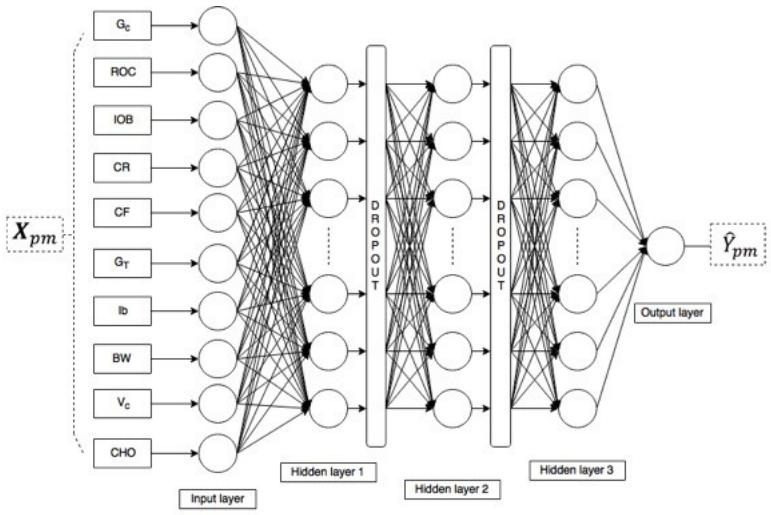



**Figure 3.2:** Optimal correction $Y_{pm}$ values as a function of preprandial BG and ROC values obtained for patient adult#1 when meal CHO intake is 50 g.

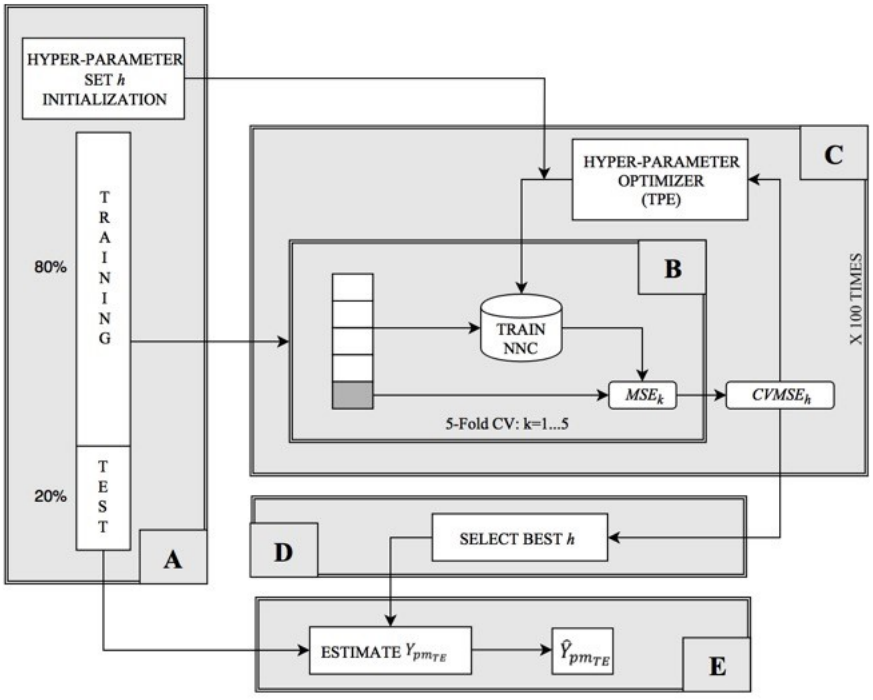
of obtaining a compromise between capability of fitting the training data and ability to generalize. In addition, two dropout layers have been interposed between each hidden layer to prevent overfitting and improve the generalization capability of NNC [71]. For each dataset record, we build a record consisting of 10 patient-related characteristics, hereafter labeled as $\mathbf{X}_{pm}$. Among those, a subset of three features are related to the patient preprandial status: $G_C$, ROC, and IOB. Then, we also considered four patient-specific therapy parameters: CR, CF, $G_T$ and the insulin pump basal infusion rate, $I_b$. In addition to those, we stored two physiology related features, that is, the body weight, BW, and the interday insulin sensitivity variability profile class, $V_C$ (as defined in Visentin et al. [55]). Finally, we memorized the meal carbohydrate amount, that is, CHO.

The NNC input layer consists of 10 neurons, each of which associated to a feature of $X_{pm}$, while the output layer is composed of one single neuron that combines the outputs of the last fully connected layer to produce an estimate of $Y_{pm}$ , that is, $\hat{Y}_{pm}$.

Training of NNC is performed through gradient descent training algorithm applied in a mini-batch mode [72]. In particular, we build a software framework (schematized in Figure 3.3B) to solve two problems: first, we want to tune the NNC hyper-parameters and structure, second we want to automatize the model selection and training procedure. In detail, block A splits the abovementioned dataset to define training and test data, assigning 80% of the




(a)




(b)

**Figure 3.3:** (a) Structure of the proposed NNC neural network. (b) Scheme of the software framework implemented to tune NNC hyper-parameters *h*: block A randomly initializes *h* values and splits the dataset to define test and training set; block B assesses the performance of h in a 5-fold CV setting over the training set; block C implements TPE to optimize *h*; block D selects the best h set and finally; block E evaluates the performance of NNC on the test set.

available ($\mathbf{X}_{pm}$, $Y_{pm}$) pairs to the training set, that is, ($\mathbf{X}_{pm_{TR}}$, $Y_{pm_{TR}}$), and the remaining 20% to the test set, that is, ($\mathbf{X}_{pm_{TE}}$, $Y_{pm_{TE}}$). Moreover, block A is in charge of randomly initializing the 9 NNC hyper-parameters that we identify as critical and required to be tuned: the number of hidden units and the activation function type of each hidden layer, the dropout percentage of each dropout layer, and the mini batch size.
Then, for a given set of hyper-parameters $h$, block B assesses its performance in a 5-fold cross-validation setting. In practice, the training set is divided into five folds, then four folds are used for training the NNC and the fifth one is used for validation and evaluation. Finally, block B computes the average intrafold mean squared error (MSE) defined as:

$$CVMSE_h = \frac{1}{5}\sum_{k=1}^{5} MSE_k \tag{3.2}$$

where subscript $h$ indicates the hyper-parameter set at hand and $MSE_k$ stands for the MSE computed by considering the *k-th* fold as validation set:

$$MSE_k = \frac{1}{N_k}\sum_{k=1}^{N_k}(Y_{pm_k} - \hat{Y}_{pm_k})^2 \tag{3.3}$$

where $N_k$ is the cardinality of the $k-th$ fold, $Y_{pm_k}$ are the target corrections associated to the *k-th* fold and $\hat{Y}_{pm_k}$ are the respective estimates obtained by the trained NNC.
To solve the hyper-parameter optimization task, block C is iterated 100 times to implement the tree-structured Parzen estimator (TPE) technique [73], that is, an sequential model-based global optimization algorithm where new observation (a set of hyper-parameters) is collected and analyzed at the end of each iteration to decide which set of hyper-parameters will be tried next (see Appendix A for a detailed description of TPE). Finally, in block D, we select the final set of hyper-parameters $h$ by which we obtained the minimum $CVMSE_h$ and, in block E, we eventually obtain the estimated optimal corrections associated to the test set, that is, $\hat{Y}_{pm_{TE}}$.

## 3.4 Assessment of glycemic outcomes

For each test set scenario in ($\mathbf{X}_{pm_{TE}}$, $Y_{pm_{TE}}$), we compared the BG profile obtained using SF, with the BG profiles obtained from the adoption of Equation 3.1. In particular, Equation 3.1 is calculated using different definitions of $f(\cdot)$,

that is, $f(\cdot) = f_{SC}(ROC)$ (see Equation 2.4), $f(\cdot) = f_{PE}(ROC)$ (see Equation 2.6), $f(\cdot) = Y_{pm_{TE}}$ and $f(\cdot) = \hat{Y}_{pm_{TE}}$. The quantitative assessment of postprandial glucose control has been performed by calculating BGRI.
By mean of these comparisons we want to verify whether machine learning based methodologies have room to perform better than both the gold standard, that is, SF, and if they improve the performance of currently available methodologies for computing Equation 3.1, that is, SC and PE (here we decide not consider BU since its definition is not expressible in terms of Equation 3.1). Moreover, comparing the estimated corrections, $\hat{Y}_{pm_{TE}}$, against their target value, $Y_{pm_{TE}}$, we want to assess how good the NNC is in predicting the optimal value of $f(\cdot)$ and, if not, discuss which are the possible causes. Finally, to statistically evaluate the between-methods differences, we performed a non-parametric multi-way ANOVA test, explicitly considering the virtual patient and the associated correction method as factors, followed by a multi-comparison using Dunn's post hoc test with 0.1% significance level. The reason for choosing such a statistical test is that the test set could contain traces coming from the same virtual patient (same physiology, but different initial conditions). Therefore, we need to perform the appropriate statistical analysis to be able to consider dependencies between observations within subjects.

## 3.5 Performance of NNC against literature techniques

### 3.5.1 Representative example

Figure 3.4 shows two examples of BG profiles obtained with the considered methods. In the top panel case (adult#26, $G_C$ = 250 $\pm$ 5, ROC = -1 mg/dL/min and CHO = 100 g) the NNC achieves significantly better results compared to SF, SC and PE being able to avoid hypoglycemia. Moreover, comparing the glycemic outcomes obtained using Equation 3.1 with OPT ($Y_{pm}$ = -170.00 mg/dL) versus its estimate provided by NNC ($\hat{Y}_{pm}$= -167.65 mg/dL), it is possible to observe how the NNC provides a good approximation of the target correction value. In Figure 3.4, the BG profiles have been zoomed in to highlight traces around minimum level. Notably, while SF, SC and PE lead adult#26 to severe hypoglycemia, NNC allows to keep BG greater than 70 mg/dL significantly reducing BGRI. In detail, the obtained BGRI values are: BGRI(SF) = 14.27, BGRI(OPT) = BGRI(NNC) = 12.10, BGRI(SC) = 13.59, and BGRI(PE) = 13.07. Qualitatively equal results are obtained in the bottom panel case

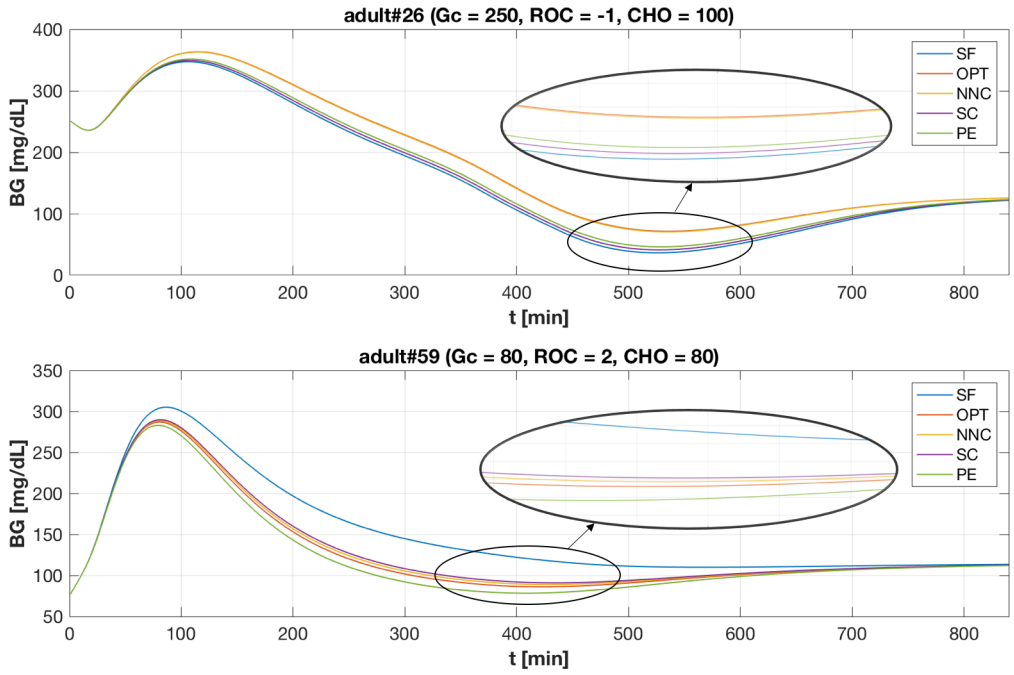



**Figure 3.4:** Example of BG trace obtained with SF (in blue), OPT (in red), NNC (in yellow) SC (in violet), and PE (in green) methods. X-axis have been truncated at 840 min since, thereafter, the BG traces coincides. Top panel. BG profiles obtained for the virtual subject adult#26 when $G_C$ = 250 $\pm$ 5 mg/dL, ROC = -1 mg/dL/min and CHO = 100 g. Bottom panel. BG profiles obtained for the virtual subject adult#59 when $G_C$ = 80 $\pm$ 5 mg/dL, ROC = 2 mg/dL/min and CHO = 80 g. Profiles have been zoomed in to highlight BG traces around minimum level.

(adult#59, $G_C$ = 80 $\pm$ 5 mg/dL, ROC = 2 mg/dL/min, and CHO = 80 g). In particular, the NNC estimated $f(\cdot)$ value ($\hat{Y}_{pm}$ = 54.36 mg/dL) achieves better BGRI compared with SF, SC and PE. Again, comparing the BG traces using the optimal $f(\cdot)$ value ($Y_{pm}$ = 60.00 mg/dL) versus its estimate provided by the NNC, the error introduced by NNC does not lead to significant differences. In particular, the obtained values of BGRI are: BGRI(SF) = 4.01, BGRI(OPT) = BGRI(NNC) = 3.26, BGRI(SC) = 3.27, and BGRI(PE) = 3.39.

### 3.5.2 Assessment of the methods in terms of BGRI

In Figure 3.5, the distributions of the BGRI difference ($\Delta$BGRI) between SF versus OPT, SF versus NNC and OPT versus NNC are shown via boxplot representation. Numerical values (median and interquartile range) are reported in the first three columns of Table 3.1. SF versus OPT distribution represents the best achievable improvement by mean of Equation 3.1. In particular, using

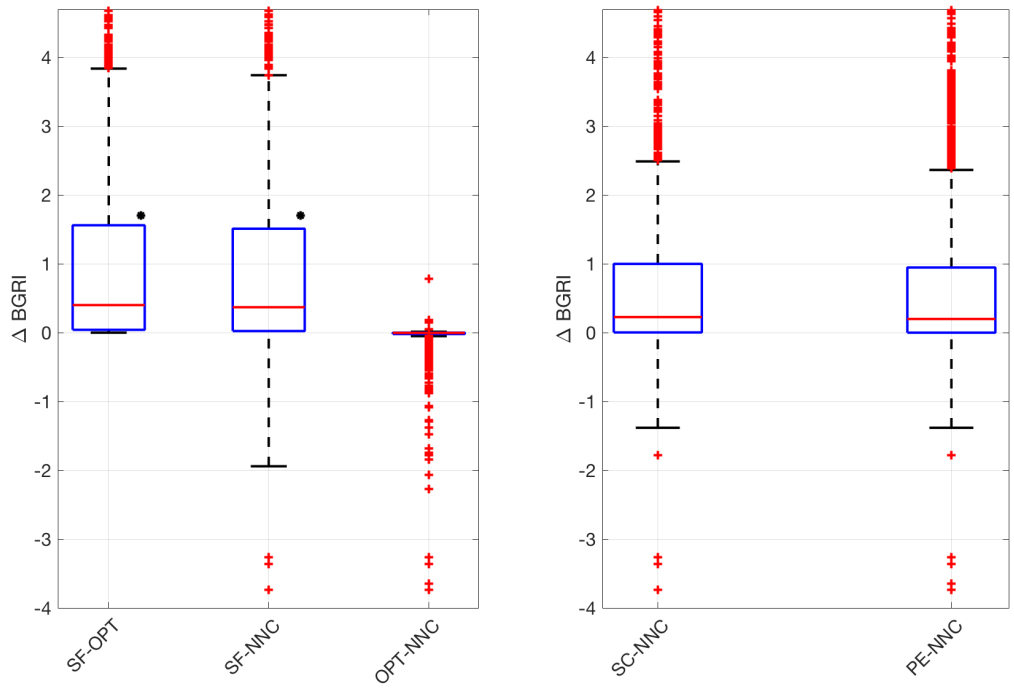



**Figure 3.5:** Boxplot representation of the distribution of ΔBGRI obtained comparing (a) SF versus OPT, SF versus NNC, OPT versus NNC, SC versus NNC, and PE versus NNC. Red horizontal lines represent median, boxes mark interquartile ranges, dashed lines are the whiskers, red crosses indicate outliers. Black dots indicate statistical significant between-distributions differences.

the optimal correction, we obtained a statistically significant ($p < 0.001$) median BGRI improvement of 0.40 (see Table 3.1). Focusing on SF versus NNC, as for SF versus OPT, the obtained results are better in NNC (see Table 3.1). Notably, the median ΔBGRI and interquartile range is almost the same as in SF versus OPT, meaning that the NNC estimates the optimal target values with good precision. As further proof of the capability of the NNC in estimating the optimal correction values and improving the glycemic outcomes, in Figure 3.5 we report the distribution of the difference between BGRI in OPT and NNC. No statistically significant difference between the two distributions is

**Table 3.1:** Median [interquartile range] results of ΔBGRI (first row) obtained comparing SF versus OPT, SF versus NNC, OPT versus NNC, SC versus NNC and PE versus NNC and p-values obtained using the Dunn's post hoc test (second row).

| | **SF-OPT** | **SF-NNC** | **OPT-NNC** | **SC-NNC** | **PE-NNC** |
|---|---|---|---|---|---|
| ΔBGRI | 0.40 [0.04, 1.56] | 0.37 [0.02, 1.51] | 0 [-0.02, 0.01] | 0.23 [0.01, 0.99] | 0.20 [0.01, 0.95] |
| p-value | <0.001 | <0.001 | 0.30 | <0.001 | <0.001 |

found (median difference equal to 0, p = 0.30). However, Figure 3.5 points out that, in several scenarios, the NNC error is relevant and, compared to SF, lead to worse BGRI values. In particular, BG control worsens when using NNC to estimate $f(\cdot)$ due to the error the model makes in approximating OPT. This shortcoming represents a limit of the NNC and future work will be performed to properly investigate how to deal with this drawback to guarantee safety and effectiveness in all conditions. Figure 3.5 shows the ΔBGRI distribution and median/interquartile results obtained comparing the NNC with SC and PE. Numerical values (median and interquartile range) are reported in the last two columns of Table 3.1. Overall, according to Figure 3.5, NNC introduces a small but statistical significant improvement ($p < 0.001$) equal to 0.23 and 0.20 if compared with SC and PE, respectively. However, as above, boxplots show that, in several scenarios, NNC lead to worse glycemic outcomes due to the error in the estimate of the optimal $f(\cdot)$ value.

## 3.6 Summary of the outcome and indications for the next steps

In this chapter, the development of a new NN-based methodology to personalize insulin bolus calculation exploiting $G_C$, ROC, IOB, $CR$, $CF$, $I_b$, $G_T$, BW, $V_C$, and CHO has been described. Both intuition and evidence provided by the obtained results suggest that the optimal modulation of insulin bolus is strongly related to preprandial conditions and individual parameters of the patient.
An in silico study performed in 100 virtual subjects in noise-free conditions, showed that, in terms of glycemic outcomes measured as BGRI, the new method outperforms the literature approaches SF, SC, and PE, and is close to the optimum determined by exhaustive search. Although the quantitative improvements might seem minor, these preliminary results encourage further investigations on machine learning based methodologies to provide patients with decision support tools able to ease their daily insulin therapy routine. With regard to this aspect, it is important to stress that other non linear machine learning techniques (e.g., kernel support vector machines or regression trees) could be considered for the scope as well. Implementation of alternative strategies for the adjustment of SF will be discussed in Chapter 4.
Finally, the effectiveness of the optimal correction $Y_{pm}$ to be applied to the SF is, not surprisingly, related to the structure of the SF itself. Indeed, an accurate

analysis of the results of our simulations showed that building a tool able to correct optimally SF seems possible only if, during the training phase, all the possible physiological classes/typologies of the patients at hand are available. While this highlights a restriction of the domain of validity of the NNC presented here, it also suggests that limitations in predicting the optimal correction $Y_{pm}$ are intrinsically related to the strict constraints related to the original structure of SF. Indeed, we found that there is no correlation between $Y_{pm}$ and the actual insulin bolus amount. To better illustrate the point, let us consider two patients having the same meal conditions, that is, same $G_C$, CHO and IOB, whose optimal insulin boluses are both 0. Since, in general, two patients have different therapy parameters, that is, $G_T$, CR and CF, this will result in different optimal $Y_{pm}$ values. This consideration suggests that correcting SF by a fixed values as defined by $f_{SC}(ROC)$ and $f_{PE}(ROC)$ can be, in general, suboptimal. A possible margin of improvement, hence, lies in releasing the hypothesis of using the insulin bolus amount provided by SF as a starting point and devising new methods that naturally take into account for CGM-derived information and current patient status and characteristics to compute the meal insulin bolus. This will be extensively discussed in Chapter 5 where we develop new dosing rules based on linear regression models.

# Chapter 4

# Classification of postprandial meal status with application to insulin dosing

[1]

As showed in the previous chapter, machine learning based algorithms, thanks to their ability of capturing and exploiting non linear relationships, such as glucose-insulin dynamics, seem to be viable candidates to personalize insulin dosing with the aim of improving T1D therapy and achieving better glycemic control. In view of these findings, in this chapter we explore an approach based on eXtreme Gradient Boosted (XGB) trees [75] aiming at classifying, at meal time, the postprandial glycemic status (i.e., BG concentration being too low, too high, or within the target range). Then such an outcome is used to reduce or increase the corresponding meal bolus dose provided by SF accordingly.

## 4.1 Method overview

As depicted in Figure 4.1, the XGB-based method implements a two step procedure:

- Step 1: given a set of input features, XGB is used to predict, at meal time, the future glycemic status, i.e. to determine a classification of BG concentration as too low, too high, or within target range, in the postprandial

[1] This chapter contains material published in the article [74]

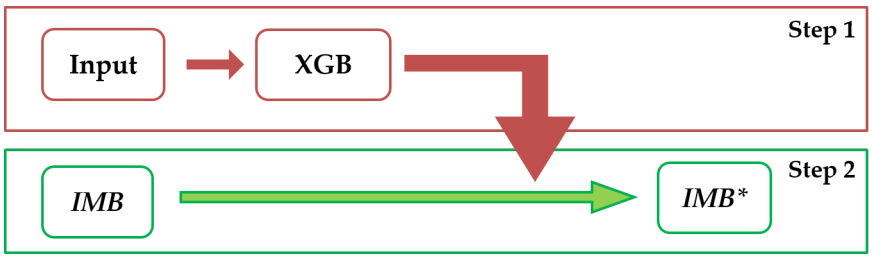



**Figure 4.1:** Schematic representation of the method rationale. In Step 1, input features are used to train XGB and predict the future postprandial status. Then, in Step 2, the XGB outcome is used to modulate $IMB$ accordingly and obtain the "adjusted dose", i.e. $IMB^*$.

time window, i.e. between 2h and 6h after meal, exploiting the knowledge of preprandial CGM measurements, carbohydrate intake estimates and insulin infusion recordings. In particular, to define the glycemic status in the postprandial time window, i.e. if after the meal the individual glycemia will be in hypoglycemia, hyperglycemia, or within a target range, the CGM-based metric described by Herrero et al. [40] was employed. It consists in computing the minimum postprandial CGM ($G_{min}$) level within a predefined postprandial time window as:

$$G_{min} = \min_{t \in [t_m+2h, t_m+6h]} CGM(t) \tag{4.1}$$

where, $t_m$ is the time of the meal. The rationale for choosing a postprandial time window from two to six hours after meal time is that we are interested in computing the minimum postprandial glucose after the glucose peak time. Note that a typical glucose peak after meals is around 70 min but can potentially be longer for low-absorption meals. We then chose to end the window at six hours post meal to make sure that the effect of the short-acting meal insulin bolus is over [76]. Finally, the assumption that basal insulin (i.e., long-acting insulin) is correctly adjusted is made. Then, the classification target Y consists of three classes defined as follows:

$$Y = \begin{cases} C_1 & \text{if } G_{min} < th_{hypo} \\ C_2 & \text{if } th_{hypo} \leq G_{min} \geq th_{hyper} \\ C_3 & \text{if } G_{min} > th_{hyper} \end{cases} \tag{4.2}$$

where $th_{hypo}$ and $th_{hyper}$ are the hypoglycemia and hyperglycemia thresholds, respectively. Here, these thresholds were set to $th_{hypo}$ = 70 mg/dL

and $th_{hyper}$ = 140 mg/dL. Figure 4.2 shows a graphical representation of the used classification criterion.

- Step 2: the future glycemic status produced as XGB outcome is used to improve glycemic control by a suitable real time modulation of SF. Specifically, the $IMB$ computed through SF will be increased if the XGB outcome indicates a future postprandial hyperglycemia. On the other hand, $IMB$ will decreased if XGB forecast an hypoglycemic event.

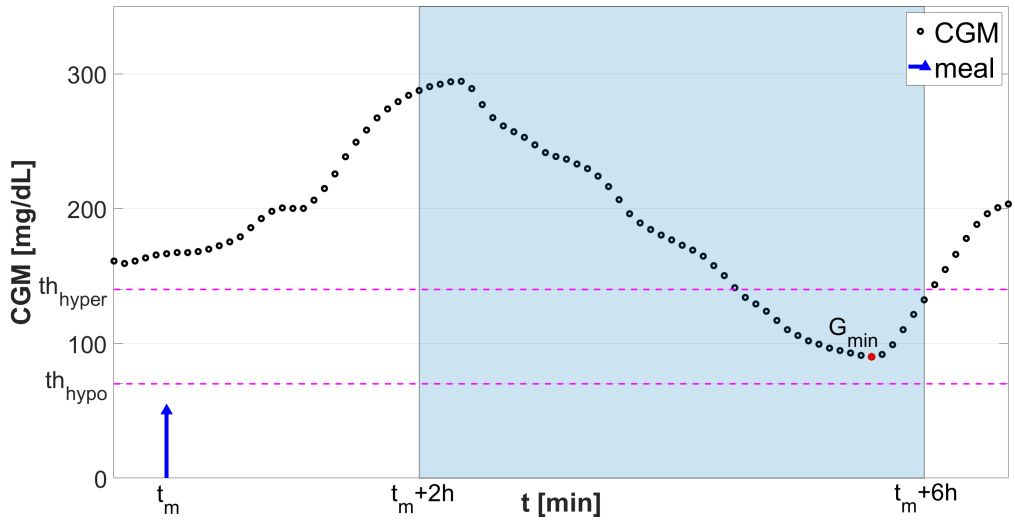



**Figure 4.2:** Graphical representation of the classification scheme. Black dots are the CGM samples. Blue stem indicates the meal event. Red dot is the minimum glucose level $G_{min}$ reached in the fixed postprandial window $[t_m + 2h, t_m + 6h]$ (highlighted in light blue). Horizontal magenta lines denote the thresholds used to discretize $G_{min}$ into the classification target $Y$.

## 4.2 Classification of future postprandial glycemic status: method description and assessment

### 4.2.1 eXtreme Gradient-Boosted tree model

XGB is an highly effective and efficient implementation of gradient boosted trees (GBT), i.e. supervised learning algorithms that have been shown to provide state-of-the-art results in many classification tasks [75]. See Appendix B for a more detailed description of GBT. It is particularly well-known for being

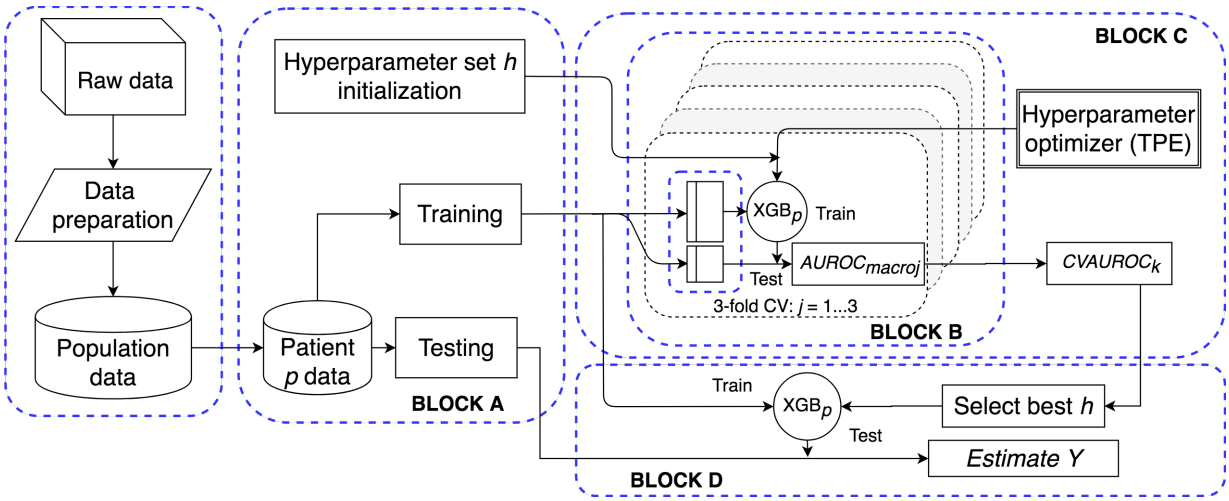



**Figure 4.3:** Structure of the proposed $XGB_p$ software framework. After data preparation, for patient $p$, block A initializes $h$ and splits the data in training and test; block B computes the performance of the hyperparameter set in a 3-fold CV setting over the training set; block C implements a Tree-structured Parzen Estimator to optimize $h$; block D selects the best $h$ set and evaluates the performance of $XGB_p$ on the test set.

robust against outliers, to work well for relatively small datasets, and to automatically handle feature selection; thus, it well fit the purpose of this work. Moreover, XGB associates each input to the probability of belonging to each one of the three considered target classes, i.e. $p(C_1)$, $p(C_2)$, $p(C_3)$. This feature will result very useful, as described later in Section 4.3. The adoption of XGB, for the classification of future postprandial glycemic status, is furtherly substantiated by the fact that, if compared to other machine-learning approaches, such as NNs, it naturally provides a model that is interpretable by associating a certain level of importance to each of the input features.
For each studied subject $p$, we trained a different XGB model, $XGB_p$, the aim being accounting for inter-subject variability and personalizing the methodology at the individual level. For this purpose, a software framework (schematized in Figure 4.3) was built to automatically manage data preparation, model tuning, and testing.

**Features vector and data preparation** The selected input (features vector) ($X_{ip}$) for the XGB model is composed by the following entities:

- Estimated amount of ingested carbohydrates ($CHO_i$);
- Meal insulin bolus ($IMB_i$);
- Two binary indicators denoting whether there was a hypo/hyperglycemic

event in the last three hours. This feature allows to capture the physiological response to hypoglycemia (e.g. secretion of glucagon) or the ingestion of rescue carbohydrates;

- The hour-of-day of $t_{m_i}$ and three binary indicators representing the meal type (i.e. breakfast, lunch, or dinner), which are used to capture the subject's intra-day variability (e.g. circadian rhythms);

- Two features describing the time elapsed since the last insulin bolus and meal intake, respectively. This feature might help to capture specific patient behaviors, such as using multiple boluses to treat the same meal and/or snacking pattern;

- CGM data within the time window $[t_{m_i} - 1h, t_{m_i}]$;

In addition, data were preprocessed in order to obtain additional features. In detail, for each ingested meal at time $t_{m_i}$, CGM data, the estimated amount of carbohydrates (CHO), and insulin data (INS), were considered within the time window $[t_{m_i} - 1h, t_{m_i}]$. Then, such data were processed as follows:

- CGM was used to obtain the corresponding glucose rate of change, static risk (SR), and dynamic risk (DR) [77] time series, which empower the model with additional features that capture the dynamics of the CGM signal (e.g. glycemic variability);

- CHO was used to calculate the rate of glucose appearance in the blood (Ra) within $[t_{m_i} - 1h, t_{m_i}]$ through the use of a gastrointestinal model [49] that describes carbohydrate digestion and glucose absorption as follows:

$$Ra(t) = \frac{CHO(t) \cdot CHO_{BIO} \cdot t \cdot exp(-t/t_{MAXG})}{t^2_{MAXG}} \tag{4.3}$$

  where $CHO_{BIO}$ (dimensionless) is the carbohydrate bioavailability, $t_{MAXG}$ (min) is the time of maximum appearance rate of glucose. Here, Ra is generated setting $t_{MAXG}$ and $CHO_{BIO}$ to population values, i.e. 50 min and 0.8, respectively;

- INS data were transformed into two continuous signals representing an estimate of plasma insulin concentration ($I_P$) [78] and IOB [79] to account for insulin absorption and clearance as follows:

$$\begin{cases} \dot{C}_1(t) = INS(t) - \frac{C_1(t)}{t_{DIA}} \\ \dot{C}_2(t) = \frac{(C_1(t) - C_2(t))}{t_{DIA}} \\ \dot{I}_P(t) = -k_e I_P(t) + \frac{C_2(t)}{t_{DIA}} \\ IOB(t) = C_1(t) + C_2(t) \end{cases} \tag{4.4}$$

where $C_1$ and $C_2$ are the subcutaneous insulin absorption compartments, $t_{DIA}$ (min) is the duration of insulin action, and $k_e$ (min$^{-1}$) is the decay rate. Here, $t_{DIA}$ and $k_e$ have been set to population values, i.e. 120 min and 0.98 min$^{-1}$. As per the Ra signal, $I_P$ and IOB was estimated within $[t_{m_i} - 1h, t_{m_i}]$ assuming no additional insulin infusion in that period.

**Model tuning and testing** For each subject, training of XGB$_p$ was performed through gradient descent algorithm [60]. Figure 4.3 schematizes the model tuning and testing procedure. In detail, referring to Figure 4.3, block A splits the dataset $(X_{ip}, Y_{ip})$ into training and testing data. Block A was also in charge of initializing the hyper-parameters to random values. These hyper-parameters are: the number of trees, the maximum depth of each tree, the subsample ratio of the training instances, the L2 regularization term on weights, and the learning rate. Then, in block B, training data were used to tune the model and, given a set of hyper-parameters $h$, performance was assessed in a three-fold cross validation setting. Specifically, training data were split in three folds, of which two folds were used for training, and the third one was used to validate and evaluate the model. Then, performance of the $k$-th hyper-parameter set was quantified in terms of intra-folds macro-average area under the receiver operator characteristic curve (AUROC) [80]:

$$CVAUROC_k = \frac{1}{3} \sum_{j=1}^{3} AUROC_{macro_j} \tag{4.5}$$

where $AUROC_{macro_j}$ denotes the intra-folds macro-average AUROC computed using the $j$-th fold as validation set:

$$AUROC_{macro_j} = \frac{1}{3} \sum_{c=1}^{3} AUROC_{C_C} \tag{4.6}$$

where $AUROC_{C_C}$ stands for the AUROC of the $c$-th class computed in a one-vs-all fashion.

Hyper-parameters were optimized by iterating block C using the TPE technique [73]. Finally, block D selected the set of hyper-parameters k associated to the maximum $CVAUROC_k$ and it obtained the predicted classes on the test set.

### 4.2.2 Simulated dataset used for the assessment

To evaluate the proposed XGB model, data of 100 virtual adult subjects were generated by simulation over two months using a state-of-the-art mathematical model of a T1D subject decision-making (T1D-DM) [81]. This model is an evolution of S2013, which in addition to the glucose-insulin dynamics, includes a model of intraday variability of insulin sensitivity [55], models of glucose sensors error [82][83], and a behavioral model for a T1D population. We decided to use a more sophisticated version of the UVa/Padova T1D Simulator given that, to train XGB in order to challenge the algorithm to classify the postprandial glycemic status in the presence of errors and uncertainty, and we did not necessitate of noise-free data since here we did not compare XGB against a sort of ideal algorithm as for NNC. The obtained dataset consists, for each individual, of one continuous time series, i.e. CGM; and two impulsive signals, i.e. INS and CHO, respectively. The first month of recordings was used to train and tune the model, while the second month was used for testing purposes. In particular, this choice allowed making a trade-off between model performance and the possibility of deploying such a model in real life. In fact, using more training data tends to improve model performance, but at the same time it requires more time to collect them, which might jeopardize patient's adherence to the therapy.

### 4.2.3 Classification results

Figure 4.4 shows the AUROC distributions computed in the test set for class $C_1$ , $C_2$ , $C_3$ , and their macro-average obtained in the population. On average, good performance is achieved. Specifically, the obtained average [interquartile range] AUROC values are: AUROC($C_1$) = 0.89 [0.86-0.93], AUROC($C_2$) = 0.76 [0.69-0.82], AUROC($C_3$) = 0.86 [0.80-0.91], AUROC(MACRO-AVERAGE) = 0.84 [0.78-0.87]. Particularly, $C_2$ results to be the most difficult class to predict. This is due to the discretization applied to $G_{min}$. Specifically, since each $G_{min}$ is hard-assigned to a single class, the classification error is likely higher when

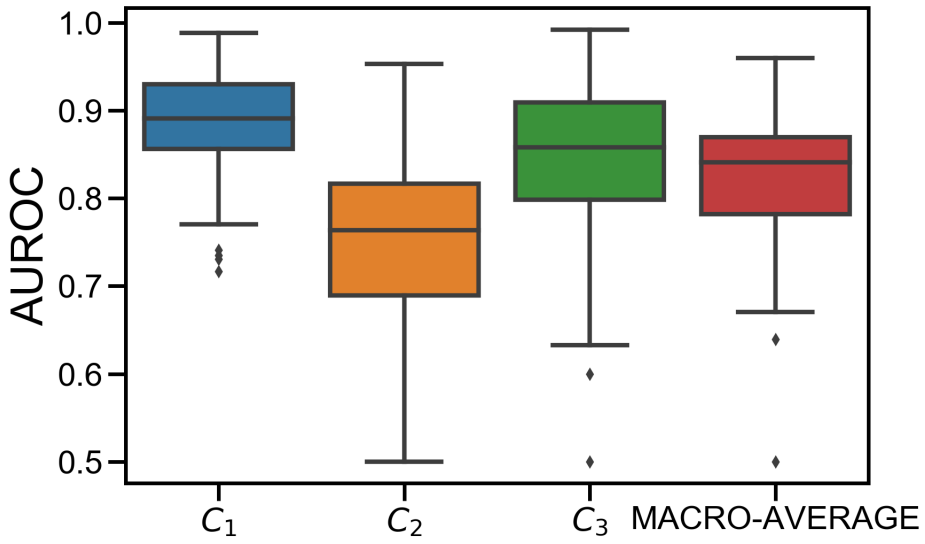



**Figure 4.4:** Boxplot representation of the distribution of AUROC obtained for $C_1$ (in blue), $C_2$ (in orange), $C_3$ (in green), and their macro-average (in red) obtained in the population. Black horizontal line represents median, the black box marks the interquartile range, vertical black lines are the whiskers and black diamonds indicate outliers.

the original $G_{min}$ was close to the $th_{hypo}/th_{hyper}$ thresholds before being discretized. So, given that $C_2$ is both upper ($th_{hyper}$) and lower ($th_{hypo}$) bounded, it is more likely that the error increases because of this discretization.
Notably, best performance is obtained for $C_1$, meaning that the model is accurate in detecting hypoglycemia. Finally, several critical outliers are present in $C_2$ and $C_3$. However, they are strictly higher than 0.5, meaning that XGB always behaves better than the "random" classifier [60].
Finally, Figure 4.5 (left panel) shows the ROC curves obtained for a representative subject (adult#1). The model achieves very good performance. In particular, AUROC computed for class $C_1$, $C_2$, and $C_3$ are 0.97, 0.83, and 0.95, respectively. Figure 4.5 (right panel) reports the corresponding confusion matrix. Moreover, analyzing the classification error that XGB makes when it predicts the wrong class, it can be seen that only picks adjacent classes. This is very important, since it avoids dangerous counter-actions by the patient, i.e. classifying an actual hypoglycemia as a hyperglycemia could lead the patient to increase the severity of the episode.

### 4.2.4 Potential utility of the XGB-based estimates

In a real-time setting, the developed $XGB_p$ classifier can be applied at meal-time to forecast the post-prandial glycemic status (i.e., hyperglycemia, euglycemia or hypoglycemia). Such information on the future glycemic status can be used for multiple purposes. For instance, it can be used to generate smart alerts when future adverse events are forecasted; temporarily suspend basal insulin delivery, or suggest carbohydrate intake, to prevent hypoglycemia; and, the application which occupies us in this thesis, to recommend to modulate the insulin dose obtained with SF to be delivered at meal time.
In the next Section, XGB application is shown, the aim being improving glycemic control by using its outcome in real-time to adjust the meal insulin bolus amount computed using SF.

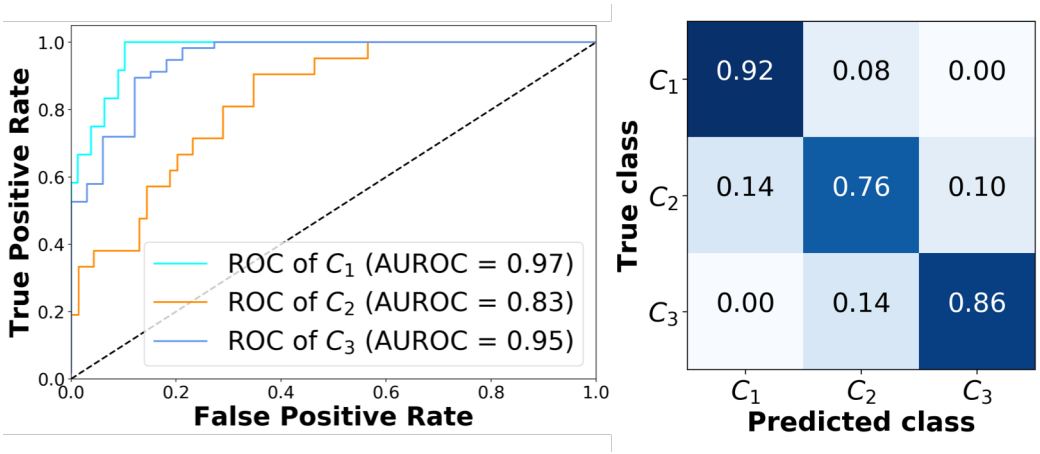



**Figure 4.5:** Classification results obtained with XGB corresponding to adult#1. Left panel. ROC curves obtained with XGB for $C_1$, $C_2$, and $C_3$. Right panel. The corresponding confusion matrix.

## 4.3 Using the XGB classifier to adjust meal insulin bolus: method description and assessment

### 4.3.1 Meal insulin dose adjustment strategy

Here, we present a proof of concept for a simple empirical strategy to adjust the meal insulin bolus according to the real-time post-prandial glucose classification provided by the XGB model at meal-time. Specifically, at meal time $t_m$, if XGB classified future $G_{min}$ belonging to $C_i$, $IMB$ computed using SF is

adjusted using Equation 4.7 hereafter labeled as XGB-IMB:

$$IMB^*(t_m) = IMB(t_m)(1 + p(C_i)f_i(IMB(t_m))) \tag{4.7}$$

where $p(C_i)$ is obtained by XGB and denotes the probability of $G_{min}$ belonging to class $C_i$; and $f_i(\cdot)$ is an empirical modulation function associated to $C_i$ which depends on the original $IMB$ amount. In particular, $f_i(\cdot)$ is defined in Equation 4.8:

$$f_1(IMB(t_m)) = \begin{cases} f_{1_{LOW}} & \text{if } IMB(t_m) < 5U \\ f_{1_{LOW}} + \frac{1}{3}\Delta f_1 & \text{if } 5 \leq IMB(t_m) < 10U \\ f_{1_{LOW}} + \frac{2}{3}\Delta f_1 & \text{if } 10 \leq IMB(t_m) < 15U \\ f_{1_{HIGH}} & \text{if } IMB(t_m) \geq 15U \end{cases}$$

$$f_2(IMB(t_m)) = 0 \tag{4.8}$$

$$f_3(IMB(t_m)) = \begin{cases} f_{3_{LOW}} & \text{if } IMB(t_m) < 5U \\ f_{3_{LOW}} + \frac{1}{3}\Delta f_3 & \text{if } 5 \leq IMB(t_m) < 10U \\ f_{3_{LOW}} + \frac{2}{3}\Delta f_3 & \text{if } 10 \leq IMB(t_m) < 15U \\ f_{3_{HIGH}} & \text{if } IMB(t_m) \geq 15U \end{cases}$$

where $\Delta f_i = (f_{i_{HIGH}} - f_{i_{LOW}})_{i=1,\dots,3}$. Intuitively, if $C_2$ is predicted, no adjustment of $IMB$ is performed. Moreover, by multiplying the modulation function by $p(C_i)$, the adjustment of $IMB$ became directly proportional to how much XGB was "certain" about assigning the current input to class $C_i$. This makes the adjustment robust to outliers, since in these cases, $p(C_i)$ results smaller, along with the $IMB$ increment/reduction. Finally, the choice of using different adjustment depending on the original $IMB$ amount allows avoiding too conservative, or too aggressive, adjustments when $IMB$ is small, or big.

### 4.3.2 Simulated dataset and statistical criteria used for the assessment

XGB-IMB was tested in silico on a new dataset generated from the 100 virtual adult subjects of T1D-DM. Note that such dataset is different from the one used to train and test the XGB model. In particular, two scenarios were created. In scenario A, 20 out of 100 subjects were used to tune the four XGB-IMB parameters ($f_{1_{LOW}}$, $f_{1_{HIGH}}$, $f_{3_{LOW}}$, and $f_{3_{HIGH}}$) over one-week simulation. Specifically,

the parameter set was chosen by using a grid-search strategy, which uses a cost function that minimizes BGRI of the studied population. In scenario B, the remaining 80 subjects were evaluated over one-month simulation with the aim of evaluating the performance of XGB-IMB strategy (with parameters obtained from scenario A). Finally, a comparison with the same metrics obtained by the use of SF was performed.

For each subject of scenario B, performance of XGB-IMB and SF were compared in terms of: mean ($MEAN_{BG}$) and standard deviation of glucose concentration ($SD_{BG}$); BGRI, percentage time in hypoglycemia (<70 mg/dL) ($\%T_{HYPO}$); percentage time in hyperglycemia (>180 mg/dL) ($\%T_{HYPER}$); percentage time in glucose target range ([70-180] mg/dL) ($\%T_{TARGET}$); and percentage time in tight glucose target range [90-140] mg/dL ($\%T_{TTARGET}$). These metrics are commonly used to assess glycemic outcomes in T1D, and follow a recent consensus report on outcome measures for artificial pancreas clinical trials evaluation [84].
These population metrics are reported as mean ($\pm$ standard deviation) for Gaussian distributed metrics or median [interquartile range]. For this purpose, non-Gaussianity of each distribution was checked by mean of the Lilliefors test with a 5% confidence level. Finally, assessment of statistical significance between-method differences was performed, with a 1% confidence level, using a paired t-test or the Wilcoxon rank sum test if the compared distributions were both Gaussian or not, respectively.

### 4.3.3 Performance of XGB-IMB in terms of glycemic control

**Scenario A** The four XGB-IMB parameters obtained from scenario A are $f1_{LOW}$ = -0.3, $f1_{HIGH}$ = -0.1, $f3_{LOW}$ = 0.8, $f3_{HIGH}$ = 0.5. Note that, while XGB-IMB reduces $IMB$ up to 30/10% if future hypoglycemia ($C_1$) is detected, it is quite aggressive at increasing $IMB$ if future hyperglycemia is predicted ($C_3$), i.e. up to 80/50% if the original $IMB$ is smaller/bigger than 5/15 units.

**Scenario B** Figure 4.6 presents an example of BG traces obtained for adult#1 using SF and XGB-IMB in scenario B. It can be observed that adjusting the original $IMB$ amount allows to achieve a better glucose control both during and after the meal, reducing hyperglycemia and allowing to avoid one out of

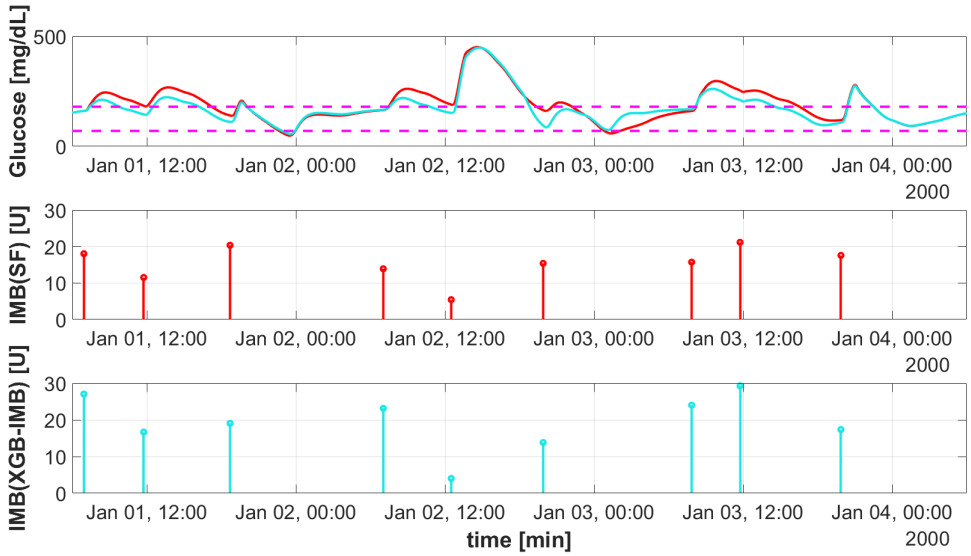



**Figure 4.6:** Top panel: Example of BG trace obtained in the first three days of simulation in adult#1 when SF (in red) and XGB-IMB (in light blue) are adopted to compute meal insulin dose in scenario B. Center panel: $IMB$ doses computed using SF. Bottom panel: $IMB$ doses computed with XGB-IMB.

two hypoglycemic episodes. In details, obtained results of SF vs. XGB-IMB are $\%T_{HYPO}$ = 1.25% vs. 1.10%, $\%T_{HYPER}$ = 41.47% vs. 30.98%, $\%T_{TARGET}$ = 57.29% vs. 67.92%, $\%T_{TTARGET}$ = 26.81% vs 29.07%, BGRI = 9.92 vs. 8.16, $MEAN_{BG}$ 175.43 mg/dL vs. 166.42%, and $SD_{BG}$ = 58.29 mg/dL vs. 53.24 mg/dL.

In Figure 4.7, the distributions of $\%T_{HYPO}$, $\%T_{HYPER}$, $\%T_{TARGET}$, $\%T_{TTARGET}$, BGRI, $MEAN_{BG}$, and $SD_{BG}$ obtained when SF and XGB-IMB are adopted to compute $IMB$ in scenario B are shown via boxplot representation. Numerical values are reported in Table 4.1. Overall, XGB-IMB improves glycemic control across the population. In particular, when comparing SF vs. XGB-IMB, a statistically significant (p<0.01) reduction of the overall glucose control in terms of $MEAN_{BG}$ and $SD_{BG}$ is achieved, i.e. 6.07 mg/dL and 4.95 mg/dL, respectively. Additionally, XGB-IMB lowers the glycemic risk in terms of BGRI (p>0.01) by 1.36. Consistently, percentage time metrics show better results when XGB-IMB is used. In detail, a significant improvement (p<0.01) in $\%T_{HYPER}$ (5.34%, from 35.18% to 29.84%), $\%T_{TARGET}$ (5.02%, from 61.98% to 67.00%) and $\%T_{TTARGET}$ (2.95%, from 28.22% to 31.17%) was observed, while a non-significant improvement (p = 0.34) of $\%T_{HYPO}$ (0.11%, from 1.93% to 1.82%) was observed.

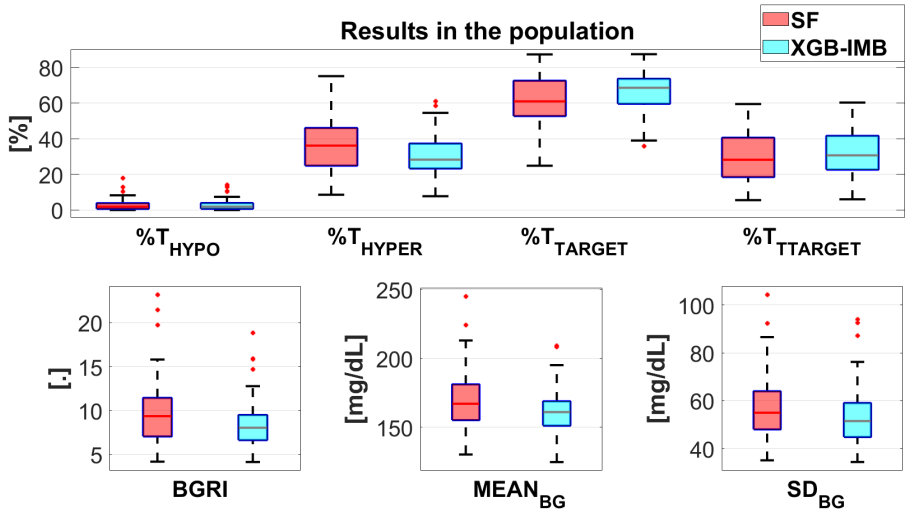



**Figure 4.7:** Boxplot of $\%T_{HYPO}$, $\%T_{HYPER}$, $\%T_{TARGET}$, $\%T_{TTARGET}$, BGRI, $MEAN_{BG}$, and $SD_{BG}$ obtained when SF (in red) and XGB-IMB (in light blue) are adopted to compute IMB in scenario B.

## 4.4 Summary of the obtained results and ideas for new insulin dosing strategies

In this chapter, a new methodology based on a state-of-the-art machine learning model, i.e. the XGB model, is used to predict at meal-time postprandial glycemic status. Results obtained with in silico data generated using a state-of-the-art T1D simulator, show that XGB is accurate at discriminating between the selected target classes (i.e., hypoglycemia, hyperglycemia, and euglycemia). Here we showed how the prediction for the postprandial glycemic status at meal-time provided by XGB can be used to adjust the meal insulin bolus dose computed by SF. Results obtained through a simple set of heuristic rules to adjust the meal insulin bolus, confirm that the proposed technique has the potential to improve post-prandial glycemic control in a T1D population. Moreover, from a practical point of view, the proposed XGB model-based technique can be easily integrated in currently available insulin pumps, or implemented in a stand-alone mobile application.

The results we obtained in both Chapter 3 and Chapter 4 showed that our strategy, i.e. leveraging on machine-learning to adjust SF, allows to improve glycemic control. However, a question rise: what if we get rid of SF and we design a brand new formula for insulin bolus dosing? Indeed, possible margins

**Table 4.1**: Median [interquartile range] is reported for $MEAN_{BG}$, $SD_{BG}$, BGRI, $\%T_{HYPO}$, $\%T_{TTARGET}$; mean ($\pm$ standard deviation) are reported for $\%T_{HYPER}$, $\%T_{TARGET}$.
*: Statistically significant difference between XGB-IMB and SF using the Wilcoxon rank sum test.
**: Statistically significant difference between XGB-IMB and SF using the t-test.

| | **SF** | **XGB-IMB** | **p-value** |
|---|---|---|---|
| $\mathbf{MEAN}_{BG}$ | 167.12 [155.28-181.16] | 161.05 [151.74-169.33] | <0.01* |
| $\mathbf{SD}_{BG}$ | 54.97 [48.03-63.95] | 51.02 [44.92-59.97] | <0.01* |
| **BGRI** | 9.36 [7.06-11.43] | 8.00 [6.60-9.83] | <0.01* |
| $\mathbf{\%T}_{HYPO}$ | 1.93 [0.07-3.81] | 1.82 [0.09-3.81] | 0.34 |
| $\mathbf{\%T}_{HYPER}$ | 35.18 ($\pm$ 14.06) | 29.84 ($\pm$ 11.50) | <0.01** |
| $\mathbf{\%T}_{TARGET}$ | 61.98 ($\pm$ 13.89) | 67.00 ($\pm$ 11.54) | <0.01** |
| $\mathbf{\%T}_{TTARGET}$ | 28.22 [18.54-40.56] | 31.17 [24.49-42.60] | <0.01** |

of improvement rose by the possibility of computing the $IMB$ through new formulae that naturally include patient characteristics and status. As such, in the next chapter, different types of linear regression models have been analysed and applied in order to design new formulae for insulin meal bolus calculation.

# Chapter 5

# New linear regression models for insulin dosing

[1] In the previous chapters, we demonstrated that a possible approach to improve insulin dosing consists in abandoning the idea of correcting SF and compute $IMB$ by adopting new formulas that naturally include ROC and patient features. In the present chapter, we propose a new linear regression modelling approach for the scope. Here, we chose linear regression models as they guarantee interpretability (by quantifying the contribution of each of the input features with a coefficient) and, at the same time, are computationally efficient and easily embeddable into standard BC. Development and test is performed by a suitably devised procedure that leverages data generated by means of S2013. Specifically, we firstly evaluated how precise are the developed formulae in approximating the optimal $IMB$, i.e. the $IMB$ guaranteeing the optimal glycemic control. Secondly, performance of each formula has been assessed by computing proper glycaemic control indices.

## 5.1 Rationale

Developing new models for meal insulin dosing is far from trivial. Indeed, huge amount of data are necessary in order to cover all the possible scenarios one can face at meal time. Specifically, the problem is twofold:

1. for each patient, different preprandial scenarios have to be studied. In-

[1]This chapter contains material of an article submitted for publication: G. Noaro, G. Cappon, S. Del Favero, G. Sparacino, and A. Facchinetti,âĂIJNew machine learning based formulas for insulin dosing in type 1diabetes leveraging continuous glucose monitoring data,âĂİIEEE Trans Biomed Eng, 2019.

deed, it is necessary to analyze the effect of a specific insulin bolus on patient's glycemia for different preprandial BG values and trends, but also for different meal CHO amount;

2. for each preprandial scenario, the "optimal" insulin bolus, i.e. the target amount, has to be identified.

Of course, as previously mentioned in Chapter 2, ad-hoc clinical studies aiming to gather these kind of datasets are both unfeasible and dangerous since, in order to identify the optimal insulin dose, they would require, for a given meal condition, to administer at the same time a multitude of insulin boluses, that is clearly impossible in real life. Moreover, is very difficult to easily observe a specific preprandial scenario in a free-living condition as it would require the exact chain of events in a preprandial time window in terms of CHO intake and insulin bolus amount and timing.
To overcome these limitations, as we did in the previous chapter 2, we resorted to S2013. Indeed, in silico simulations allow to safely and efficiently generate the desired preprandial scenarios in a virtual population of diabetic subjects and, more importantly, to identify by trial-and-error which insulin bolus would have been needed in order to optimally control each patient postprandial BG.
However, having access to such a dataset alone is not enough. In fact, one has to properly design the structure of the new formula in order to be interpretable and easy to compute. These two requirements are both equally necessary in order to let the formula be accepted by clinicians and adopted by patients as the standard of care. To do so, as already mentioned and demonstrated later in the present chapter, linear regression models can be used given their peculiar characteristic of being highly interpretable, computationally efficient, and easily embeddable into currently commercialized BC.

## 5.2 Creation of the simulated dataset for the model development and assessment

To develop and assess new models for insulin dosing, the dataset used in Chapter 2 is not sufficient. In fact, to build reliably such new methodologies the availability of "all the possible" preprandial scenarios in terms of BG and

ROC is necessary. As a result, in the following we describe how we generated a new, more complete, dataset for the scope.

S2013 has been used to generate data of 100 virtual adult subjects. For each subject, a simulated scenario of 12 hours (from 7 am to 7 pm) has been generated in a noise-free set up. As in Chapter 2, noise-free means that we did not introduced any confounding factors, allowing no corrective postprandial boluses and hypotreatment, without simulating errors in meal CHO counting, BG measurement and ROC estimation.
Specifically, simulations are conceptually divided in two parts:

- Pre-meal: it consists of the first 6 hours of simulation (from 7 AM to 1 PM), and it is used to bring the patient desired preprandial meal scenario in terms of preprandial BG and ROC;
- Post-meal: a meal was placed at 1PM and the respective meal insulin bolus was administered. Then, the remaining 6 hours of simulation (from 1 PM to 7 PM) were used to evaluate glycemic control and quantify the insulin bolus "performance".

In the following, a detailed description of the two parts is provided.

### 5.2.1 Generation of the preprandial meal scenarios

For each patient, the aim being obtaining specific preprandial BG and ROC at 1 PM, three events have been placed between 7 AM and 1 PM, namely two CHO intakes and one insulin bolus. Both CHO intakes and the insulin bolus were described by two parameters: the time of the event, and its "amount" (expressed in grams of CHO or units of insulin, accordingly). The resulting six parameters, i.e. $\hat{\boldsymbol{p}} = (t_{CHO_1}, CHO_1, t_{CHO_2}, CHO_2, t_{IMB_1}, IMB_1)$, were obtained by solving the following minimization problem:

$$\hat{\boldsymbol{p}} = argmin_{\hat{\boldsymbol{p}}} f(\hat{\boldsymbol{p}}) \tag{5.1}$$

with:

$$f(\hat{\boldsymbol{p}}) = (BG_T - BG(\hat{\boldsymbol{p}}))^2 + K \cdot (ROC_T - ROC(\hat{\boldsymbol{p}}))^2 \tag{5.2}$$

where $BG_T$ and $ROC_T$ are the target preprandial BG and ROC, respectively, $BG(\hat{\boldsymbol{p}})$ and $ROC(\hat{\boldsymbol{p}})$ are the BG and ROC value at 1 PM resulting from $\hat{\boldsymbol{p}}$, and, $K = 1000$ is a factor that normalizes the magnitude of the two quantities in order to make them weight the same.

For each patient, we generated, a total of 108 preprandial scenarios, resulting from the combination of 12 possible preprandial BG values, i.e. {70 $\pm$ 5, 80 $\pm$ 5,90$\pm$ 5,. . . ,170 $\pm$ 5,180 $\pm$ 5} mg/dL, and 9 possible preprandial ROC values, i.e. {-2 $\pm$ 0.25, -1.5 $\pm$ 0.25,. . . , 1.5 $\pm$ 0.25, 2 $\pm$ 0.25} mg/dL/min. Note that an error of $\pm$ 5 mg/dL and $\pm$ 0.25 mg/dL/min is allowed for preprandial BG and ROC, respectively, due to the difficulty to bring certain patients to specific preprandial conditions with realistic manipulations of $\hat{\boldsymbol{p}}$. Anyway, this error guarantees that no overlapping scenarios are present in the dataset. Finally, we decided to limit the possible preprandial BG value within the euglycemic range and the possible preprandial ROC value between -2 and 2 since, at this stage, we only want to cover scenarios where patients can normally eat and administrate insulin bolus without any need of additional rescue treatments.

Figure 5.1 shows the BG and ROC traces obtained for the virtual subject

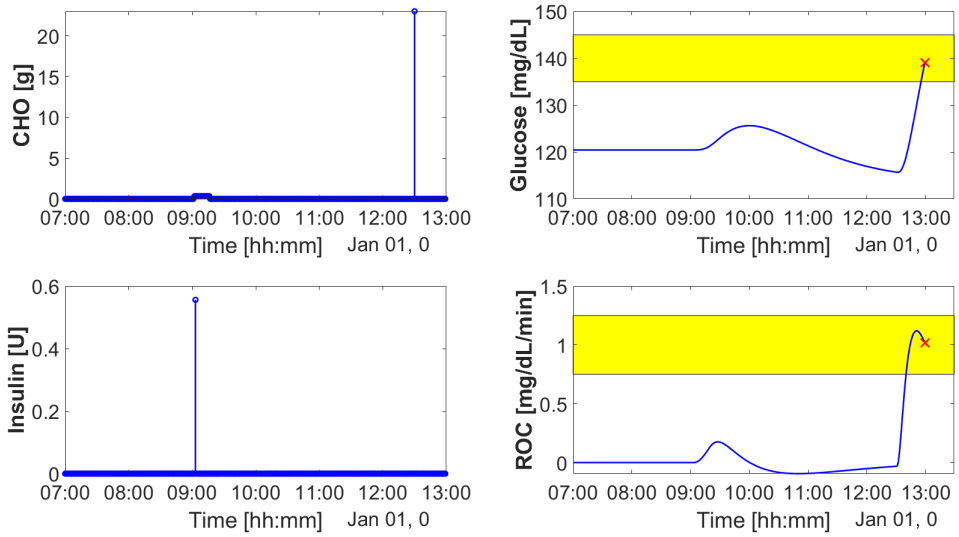



**Figure 5.1**: Representative simulated scenario for adult#1 when preprandial BG and ROC target are 140 $\pm$ 5 mg/dL and 1 $\pm$ 0.25 mg/dL/min, respectively. Left panels, show the CHO intake and insulin infusion placed within the morning time window in order to achieve the target preprandial conditions (here highlighted in yellow). Right panels show the obtained BG and ROC traces respectively. Red crosses represent the BG and ROC values at lunch time, i.e. 1:00 PM.

adult#1 in a representative preprandial scenario, i.e. when the target preprandial conditions at 1:00 PM are BG 140 $\pm$ 5 mg/dL and ROC 1 $\pm$ 0.25 mg/dL/min. Estimated $\hat{\boldsymbol{p}}$ is (9:00 AM, 5.00 g, 00:30 PM, 23.00 g, 9:00 AM, 0.55 U).
A different simulated scenario is shown in Figure 5.2. BG and ROC traces are shown for adult#1 when the target preprandial conditions at 1:00 PM are BG 100 $\pm$ 5 mg/dL and ROC -2 $\pm$ 0.25 mg/dL/min. Here, estimated $\hat{\boldsymbol{p}}$ is (10:00

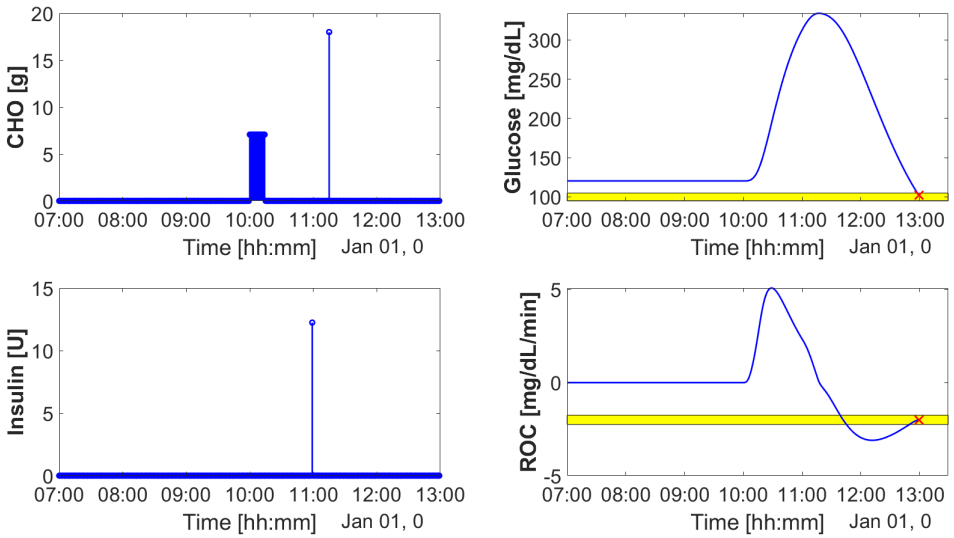



**Figure 5.2:** Representative simulated scenario for adult#1 when preprandial BG and ROC target are 100 $\pm$ 5 mg/dL and -2 $\pm$ 0.25 mg/dL/min, respectively. Left panels, show the CHO intake and insulin infusion placed within the morning time window in order to achieve the target preprandial conditions (here highlighted in yellow). Right panels show the obtained BG and ROC traces respectively. Red crosses represent the BG and ROC values at lunch time, i.e. 1:00 PM.

AM, 105.00 g, 11:15 AM, 18.00 g, 11:00 AM, 12.20 U). In both examples, it can be observed how the estimated $\hat{\boldsymbol{p}}$ allowed to bring the virtual patient to the desired preprandial scenario.

To sum up, we simulated for each adult 108 preprandial scenarios, i.e. the combination of the nine different ROC intervals and eleven BG intervals, in which, at 1 PM, each subject presents specific BG and ROC relying on CHO intake and insulin bolus between 7 AM and 1 PM.

### 5.2.2 Identification of the optimal insulin bolus dose to control post-meal glycaemia

In our simulation framework, the transition between pre-meal and post-meal phases happens at 1 PM, when subjects consumed a total of 15 different possible CHO intakes, i.e. {10, 20,..., 140, 150}. Then, for each meal the optimal insulin bolus to control BG, i.e. $IMB_{OPT}$, was defined as:

$$\hat{IMB}_{OPT} = argmin_{IMB} BGRI(IMB) \tag{5.3}$$

i.e. the insulin bolus $IMB$ that allows to obtain the minimum BGRI in the post-prandial time window defined between 1PM to 7PM.
We chose to minimize BGRI since it balances hypo- and hyperglycemic events, despite their original scale are not symmetric. Through the minimization of BGRI function, we derived the target needed to develop the new formula for insulin bolus dosing. Hereafter, we will label as OPT, the "optimal" BC, i.e. the BC that, for a given preprandial condition gives $IMB_{OPT}$ as output.
It is important to remark that, the obtained dataset represent an expanded version of the dataset used in Chapter 2. Indeed, since the same simulator and the same setup are used to find the optimal insulin bolus value, for a given preprandial condition in terms of BG and ROC values, and meal CHO of 50g, the optimal value obtained in Chapter 2 is the same as the one obtained here by solving the optimization problem.

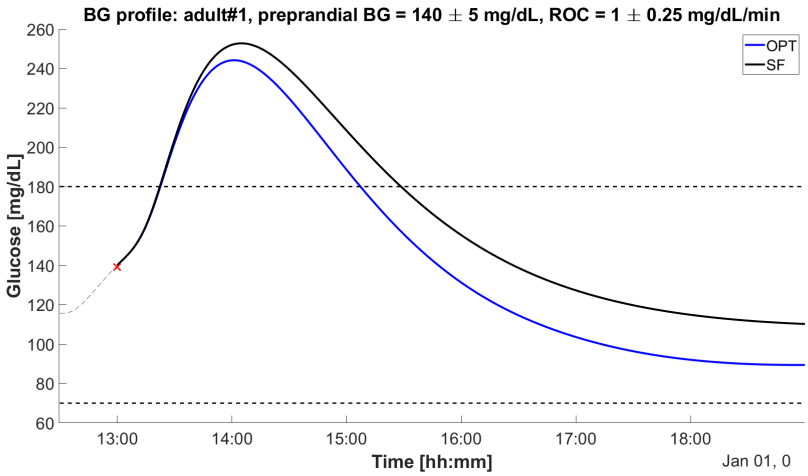



**Figure 5.3:** Representative BG curve obtained for adult#1 using SF and OPT when preprandial BG and ROC target are 140 ± 5 mg/dL and 1 ± 0.25 mg/dL/min, respectively.

Two representative scenarios are shown in Figure 5.3 and Figure 5.4 where adult#1 assumes 50 g of CHO at 1:00 PM. In particular, Figure 5.3 shows the BG curves obtained using SF and OPT when preprandial BG and ROC target are 140 ± 5 mg/dL and 1 ± 0.25 mg/dL/min, respectively. From visual inspection, it is possible to observe that using OPT it is possible to mitigate the time spent in the hyperglycemic range without inducing any hypoglycemia. In Figure 5.4, the BG curves obtained using SF and OPT when preprandial BG and ROC target are 140 ± 5 mg/dL and 1 ± 0.25 mg/dL/min, are shown. Again, compared to SF, OPT allows to tighten the BG fluctuation due to the CHO intake.

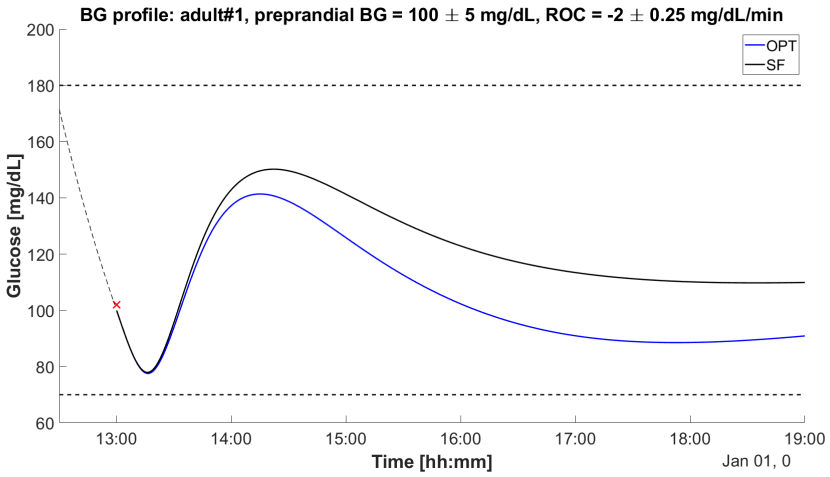



**Figure 5.4**: Representative BG curve obtained for adult#1 using SF and OPT when preprandial BG and ROC target are 100 ± 5 mg/dL and -2 ± 0.25 mg/dL/min, respectively.

### 5.2.3 Choice of the features to be included in the LR models

The aim being to develop a methodology for insulin dosing relying on easily accessible parameters only, we chose 11 features that are already well-established in current standard insulin therapy. This could make the potential integration of the new model in diabetes standard of care easier. In particular, features defining patients' characteristics and preprandial physiological condition for each preprandial scenario consists of 11 parameters:

- $CR$: the patient's carbohydrate-to-insulin ratio;
- $CF$: the patient's correction-factor;
- $G_T$: the patient's target glucose level;
- $I_b$: the patient's insulin pump basal infusion rate;
- $BW$: the patient's body weight;
- $BG$: the patient's preprandial BG at 1 PM;
- $ROC$: the patient's preprandial ROC at 1 PM;
- $CHO$: the meal carbohydrate amount the patient assume at 1 PM;
- $COB$: the patient's preprandial carbohydrate-on-board at 1 PM;

- $IOB$: the patient's preprandial insulin-on-board at 1 PM;
- $SF$: the insulin bolus amount that would have been administered using the SF described in Equation 1.1.

### 5.2.4 Final dataset preparation

The resulting dataset used in this study is composed of 162000 records, i.e. the possible combinations of 108 preprandial scenarios, 15 meal CHO, and 100 virtual subjects, and the respective $IMB_{OPT}$. Each record contains the 11 features chosen in the previous section.
Data have been divided into training and testing set, in a way that the training set is composed of independent subjects from the ones of testing set. Therefore, data of 80 subjects have been assigned to the training set, while the testing has the remaining 20. In particular, each subject has been assigned either to the training set or to the testing set in order to provide an unbiased evaluation of a final model fit on the training dataset. Because of the different measurement units among the 11 features, the variables have been standardized, removing the mean and scaling to unit variance. The pre-processing was performed in the same way for training and testing set.

## 5.3 Pre-analysis: Correlation between features and optimal bolus

Before starting with the creation of the regression models, it has been checked whether the selected features and the optimal bolus are correlated or not. This was done in order to both analyze the relationship of the target variable with all the other variables, and check whether the optimal bolus amount is predictable or not.
To quantify the linear correlation between two general features $X$ and $Y$, the sample Pearson correlation coefficient (PCC) has been computed:

$$PCC_{XY} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \tag{5.4}$$

where $N$ is the sample size, $X_i$ and $Y_i$ are the i-th samples of feature $X$ and $Y$, respectively, and $\bar{X}$ and $\bar{Y}$ are the sample mean of feature $X$ and $Y$, respectively.

Figure 5.5 shows the absolute values of PCC obtained for each of the dataset

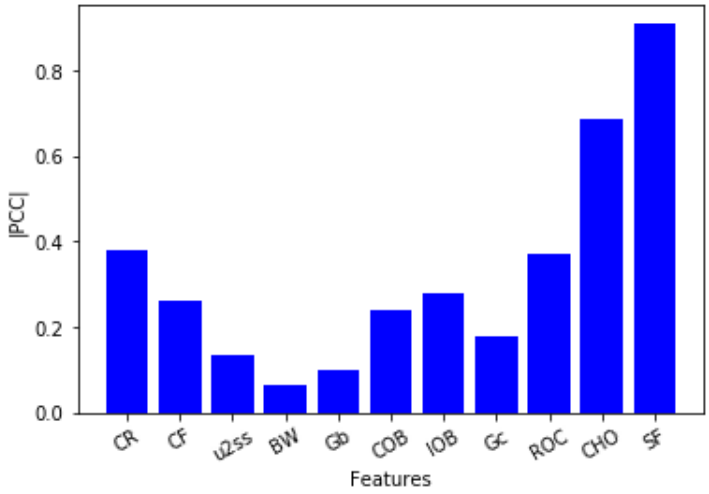



**Figure 5.5:** Bar plot of the absolute value of PCC between dataset features and target $B_{OPT}$.

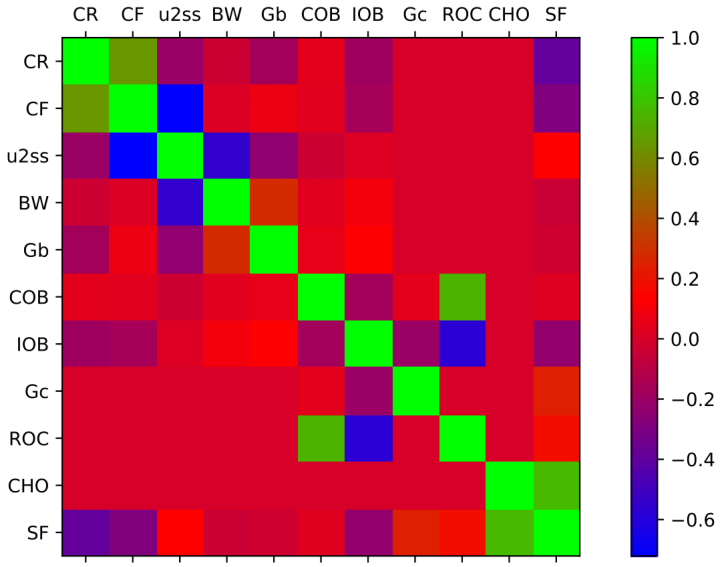



**Figure 5.6:** Heatmap of PCC between dataset features.

feature against the target variable, i.e. $IMB_{OPT}$ through a bar plot. It can be clearly seen that the most correlated variables with the target is SF, with PCC = 0.91, followed by CHO (PCC = 0.69). Also CR and ROC result correlated with the target (both PCC = 0.37), although less compared to CHO and SF.
In addition, PCC between each feature has been computed, so as to state if multicollinearity exists. In regression, multicollinearity refers to predictors

that are correlated with other predictors, indeed multicollinearity occurs when the model includes multiple factors that are correlated not just to your target variable, but also to each other. In other words, multicollinearity is associated with features being redundant. In Figure 5.6 it can be noted that several variables are correlated with each other, but some are more correlated than others, in particular SF and CHO having PCC = 0.76, while CR and CF have PCC = 0.65. The ROC is highly correlated with COB, showing a correlation coefficient of 0.74, while CF and u2ss are negative correlated (PCC = -0.72). All the other correlation coefficient are lower compared to the previous mentioned ones, but in general all the variables have a correlation, although low, with at least another variable.
In conclusion, significant information can be extracted from this correlation analysis, i.e. the target is highly correlated with several variables, suggesting that these variable (CHO, SF, CR and ROC) are relevant regressors for the models. Moreover, it can be stated that multicollinearity is present in the dataset. In this scenario, the coefficient estimates of the multiple regression may change in an irregular manner in response to small changes in the model or the data, as such, particular attention will be paid when we will analyze the trained models in order to assess their coherence with the target $IMB_{OPT}$ value. Specifically, we will check whether each of the coefficient signs associated to each model feature is coherent, e.g. CHO should be associated to a positive coefficient since the higher the CHO th higher should the $IMB$ be.

## 5.4 The three candidate linear regression (LR) models

### 5.4.1 Multiple linear regression model (MLR)

MLR has the following structure:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \tag{5.5}$$

where $y_i$ is the i-th sample of the target variable, i.e. $IMB_{OPT}$, $x_{ij}$ is the i-th sample of the j-th variable of the dataset, $\beta_j$ is the coefficient of each variable which quantifies the association between the j-th variable and the target, $\beta_0$ is the intercept of the model, and $p$ refers to the number of features.

Here, $\beta_j$ are unknown and must be estimated. Indeed, given estimates $\hat{\beta}_j$ one can compute (5.5).
Let's define the residual sum of square (RSS) as:

$$RSS = e_1^2 + e_2^2 + \cdots + e_N^2 \tag{5.6}$$

where $e_i^2$ refers to the error the model makes in estimating the i-th target value $y_i$, i.e. $e_i = y_i - \hat{y}_i$.
The unknown parameters $\beta_j$ can be estimated using a least square approach, i.e. choosing the $\beta_j$ coefficients that minimizes RSS. To do so, let's rewrite Equation 5.6 as:

$$RSS = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5.7}$$

$$= \sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2 \tag{5.8}$$

$$\tag{5.9}$$

that can be expressed in matrix form as a function of $\beta$, i.e. the unknown vector of model coefficients, as:

$$RSS(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) \tag{5.10}$$

where $\mathbf{y}$ is the $N$ x 1 vector composed of the target variable values, and $\mathbf{X}$ is a $N$ x $(p+1)$ matrix containing in each row $i$ the $p$ features of the i-th row of the dataset.
In order to obtain $\beta$ such that it minimizes RSS, Equation 5.10 has to be differentiated with respect to $\beta$:

$$\frac{\delta RSS(\beta)}{\delta \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) \tag{5.11}$$

$$\tag{5.12}$$

Assuming that $\mathbf{X}$ has full column rank, the $\beta$ value which minimizes RSS can be obtained by setting the derivative to zero, thus obtaining $\hat{\beta}$ as:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{5.13}$$

### 5.4.2 LASSO model

In order to deal with the multicollinearity problem, here, we propose the use of shrinkage methods, and in particular, the LASSO algorithm. Indeed, some regression techniques may be more sensitive to collinearity than others. Past developments in model selection methods have introduced methods for balancing model complexity and fit. Although not necessarily designed to be tolerant to collinearity, they offer approaches that may be less sensitive to this aspect [85]. In particular, to reduce the risk of overfitting and dealing with multicollinearities that could be present between dataset features, regularization based techniques can be used. Specifically, coefficient estimation through regularization strategies consists of minimizing not only RSS but also a further term $F(\beta)$, i.e. a function that puts a price on $\beta$ in order to discourage coefficients to become, in absolute value, too large, as it may happen in MLR. Hence, the function to minimize becomes:

$$L(\beta, \lambda) = RSS(\beta) + F(\beta) \tag{5.14}$$

where, in LASSO, it becomes:

$$L(\beta, \lambda) = RSS(\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \tag{5.15}$$

One of the most popular models that leverage regularization are the so called least absolute shrinkage and selection operator (LASSO) models. In LASSO $\hat{\beta}$ can be obtained by solving:

$$\hat{\beta} = \arg\min_{\beta} RSS(\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \tag{5.16}$$

where the coefficients of the multivariate model are penalized by considering the sum of their absolute values ($\lambda \geq 0$). By using Lagrangian multipliers [61], it can be shown that an equivalent way to write the minimization problem is as follows:

$$\hat{\beta} = \arg\min_{\beta} RSS(\beta) \tag{5.17}$$

$$\text{subject to} \sum_{j=1}^{p} |\beta_j| \leq t \tag{5.18}$$

where $t$ is proportional to $\lambda$. Because of the nature of the constraints, making $t$ sufficiently small will cause some of the coefficients to be exactly zero, leading to a sparse solution.

**Numerical methods for computing LASSO estimates**

Unfortunately, Equation 5.16 is not differentiable when $\beta$ contains zero values. Hence, a solution of 5.16 in closed form is not available and iterative methods are necessary to compute an approximate solution. As a consequence, for computing the LASSO solution, a wide variety of approaches have been proposed in the literature. Here, an iterative technique based on Newton's method is presented [86]. These methodologies update the vector $\beta$ at each iteration according to:

$$\beta_{k+1} = \beta_k - \alpha \nabla L(\beta_k) / \nabla^2 L(\beta_k) \tag{5.19}$$

where subscript $k$ represents the $k-th$ iteration.
Since the gradient $\nabla L(\beta_k)$ does not exist it some $\beta_j$ are zero, sub-gradients based algorithms can be used [86]. Here, for the sake of brevity, we do not present these methodologies. However, for an exhaustive explanation of sub-gradients algorithms we refer the reader to [87][88][89].

### 5.4.3 Multiple linear regression model with non-symmetric cost function (MLRNS)

In general, overestimation of $IMB$ is something that should be avoided. Indeed, people with diabetes are concerned more about the probability that a hypoglycemia event occurs (due to insulin overdosing) rather than hyperglycemia (due to insulin underdosing), being the short-term consequences of hypoglycemia more severe. Therefore, we decided to developed another multiple linear regression model, labeled as MLRNS, whose cost function, used during its training procedure, is non-symmetric and specifically designed to penalize heavily the overestimation of $IMB$ rather than its underestimation. To do so, we resort to the gradient descent algorithm, which allowed us to define arbitrary differentiable cost function to be used during the training process [60].

Here, we set such a cost function as:

$$\begin{cases} RSS & \text{if} \quad IMB_{OPT} \geq \hat{IMB} \\ 3 \cdot RSS & \text{if } IMB_{OPT} < \hat{IMB} \end{cases} \tag{5.20}$$

where $\hat{IMB}$ is the $IMB$ estimated using MLRNS.

## 5.5 Identification of the models

### 5.5.1 MLR

MLR model has been identified on the training set using the least squares algorithm. The resulting MLR is:

$$\hat{y} = 4.35 - 0.17 \cdot CR - 0.19 \cdot CF - 0.19 \cdot I_b - 0.36 \cdot BW + \tag{5.21}$$

$$+ 0.43 \cdot G_b + 0.53 \cdot COB - 0.19 \cdot IOB - 0.14 \cdot G_C + \tag{5.22}$$

$$+ 0.30 \cdot ROC + 0.25 \cdot CHO + 2.82 \cdot SF \tag{5.23}$$

Few comments can be done about MLR coefficient signs. It can be seen that negative or positive contributions given by each term are consistent. In particular, it is reasonable that $CHO$, $COB$ and $G_C$ have a positive sign within the model, since the greater their value, the higher should the injected insulin be. The ROC gives a positive contribution to the model, indeed if its value is falling, the $IMB$ will be lower, while if the glucose trend is rising the $IMB$ will be higher. A negative contribution is given by $CR$, since the lower is $CR$ the more resistant is the subject in terms of insulin response to $CHO$, thus the $IMB$ should be higher. A similar reasoning can be applied to $CF$, which has a negative sign, indeed a high $CF$ means that the patient needs a lower $IMB$. In addition, it can be seen that, as expected, $IOB$ has a negative contribution, since if any insulin amount is still active in the subject, then the $IMB$ should be lower. In conclusion, the signs of the different terms are reasonable in terms of interpretability of the model.

### 5.5.2 LASSO

The resulting LASSO model, identified on the training set, is the following:

$$\hat{y} = 4.35 - 0.19 \cdot CR - 0.22 \cdot BW + 0.39 \cdot G_b + 0.46 \cdot COB \tag{5.24}$$
$$- 0.09 \cdot IOB - 0.09 \cdot G_C + 0.36 \cdot ROC + 0.19 \cdot CHO \tag{5.25}$$
$$+ 2.86 \cdot SF \tag{5.26}$$

where the parameter $\lambda$ has been set equal to 0.03 through a grid search procedure. In this model, most of the coefficients have been shrunk towards zero, in particular the $CF$ coefficient and $I_b$ coefficient have been set exactly equal to zero. Again, as MLR, model coefficients are consistent. Moreover, it can be observed that the ROC and SF coefficients are the only two values which show an increase.

### 5.5.3 MLRNS

The resulting MLRNS model, identified on the training set, is the following:

$$\hat{y} = 3.95 - 0.31 \cdot CR - 0.05 \cdot CF - 0.09 \cdot I_b - 0.29 \cdot BW + \tag{5.27}$$
$$+ 0.31 \cdot G_b + 0.40 \cdot COB - 0.26 \cdot IOB - 0.07 \cdot G_C + \tag{5.28}$$
$$+ 0.37 \cdot ROC + 0.40 \cdot CHO + 2.50 \cdot SF \tag{5.29}$$

Looking at the absolute values of the coefficients, we can notice that many of them show lower magnitude compared to MLR, except for $IOB$, ROC, and $CHO$. Interestingly, also the coefficient associated with SF has a smaller magnitude, while the coefficients signs are consistent as the ones of the previous models.

## 5.6 Results in terms of capabilities of the the models in targeting optimal insulin bolus

Performance of each of the candidate models tested in the present chapter in predicting $IMB_{OPT}$ has been evaluated by quantifying model accuracy and goodness of fit. Specifically, we computed the mean squared error ($MSE$) and

**Table 5.1**: Comparison between prediction performance obtained using SF, BU, SC, and PE vs. MLR, LASSO, and MLRNS on the testing set.

| Metric | SF | BU | SC | PE | MLR | LASSO | MLRNS |
|---|---|---|---|---|---|---|---|
| $MSE$ | 2.00 | 1.98 | 1.73 | 1.96 | 0.82 | 0.79 | 0.93 |
| $R^2$ | 0.74 | 0.78 | 0.81 | 0.78 | 0.91 | 0.91 | 0.90 |

the coefficient of determination ($R^2$) that are defined as:

$$MSE = \frac{RSS}{N} \tag{5.30}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{5.31}$$

where $TSS$ indicates the total sum of squares defined as:

$$TSS = \sum_{i=1}^{N} (y_i - \bar{y}_i)^2 \tag{5.32}$$

where $\bar{y}$ is the mean value of the target variable $y$.

### 5.6.1 MLR

In Table 5.1 we compared the $MSE$ and $R^2$ distribution obtained using SF, BU, SC, and PE vs. MLR, the aim being assessing and quantifying whether the developed model is able to approximate more accurately the optimal target $IMB$, i.e. $IMB_{OPT}$, with respect to state-of-the-art methodologies. Compared to all the other methods, it can observed that MLR outperforms the other approaches since it is able to estimate $IMB$ values closer to the $IMB_{OPT}$. Specifically, obtained $MSE$ is 0.82 U using MLR and 2.00 U, 1.98 U, 1.73 U, and 1.96 U with SF, BU, SC, and PE respectively, while obtained $R^2$ are 0.91 using MLR and 0.74, 0.78, 0.81, 0.78 using SF, BU, SC, and PE, respectively.

### 5.6.2 LASSO

In Table 5.1 $MSE$ and $R^2$ obtained using SF, BU, SC, and PE are compared against LASSO. Again, the developed model allowed to achieve better prediction accuracy in terms of both $MSE$ and $R^2$ outperforming the other considered methods and obtaining slightly better performance than MLR.

### 5.6.3 MLRNS

Observing Table 5.1, it can be stated that the metrics slightly deviate from the values computed in the preceding models, even though the difference is not quantitatively significant. Anyway, as MLR and LASSO, MLRNS outperforms all the other considered literature methodologies by halving their error in estimating $IMB_{OPT}$.

### 5.6.4 Recap of the results

In this subsection, a brief recap and analysis of the results obtained with MLR, LASSO, and MLRNS is reported.
When considering the $MSE$ and $R^2$ as evaluation metrics, LASSO provides the best performance, with a $MSE$ of 0.79 and a $R^2$ equal to 0.91. However, the metrics of the other proposed models differ just slightly from LASSO. In addition, all the models improve considerably not only the metrics resulting from SF but also from all the literature methodologies considered in Chapter 2, i.e. BU, SC, and PE.
In conclusion, satisfactory performance has been achieved in terms of $MSE$ and $R^2$. However, an in silico evaluation in term of glycemic outcomes is needed. To do so, in the next section we will run in silico simulations to compare the glycemic control achieved by the adoption of the proposed models against SF, BU, SC, and PE.

## 5.7 Performance of the models in terms of glycemic control indexes

In this Section, for each subject of the testing set, for each scenario, performance of MLR, LASSO, and MLRNS were compared against the considered literature methods the aim being assessing whether or not, beside being more accurate in estimating the optimal $IMB$, they achieve better glucose control within the population. Specifically, performance of each methods have been compared against the same performance achieved through the use of OPT in terms of ΔBGRI, percentage time in hypoglycemia (<70 mg/dL) ($\Delta\%T_{HYPO}$); percentage time in hyperglycemia (>180 mg/dL) ($\Delta\%T_{HYPER}$); and percentage time in glucose target range ([70-180] mg/dL) ($\Delta\%T_{TARGET}$). By mean of this comparison we are able to assess which of the considered methods better ap-

**Table 5.2:** Median [interquartile range] ΔBGRI, $\Delta\%T_{HYPO}$, $\Delta\%T_{HYPER}$ $\Delta\%T_{TARGET}$, obtained by comparing SF, BU, SC, PE, MLR, LASSO, and MLRNS against OPT on the test set.

| | **ΔBGRI** | **$\Delta\%T_{HYPO}$** | **$\Delta\%T_{HYPER}$** | **$\Delta\%T_{TARGET}$** |
|---|---|---|---|---|
| SF-OPT | 0.59 [0.06, 2.50] | 0.00 [0.00, 19.11] | 0.00 [-3.32, 1.66] | -2.77 [-15.79, 0.00] |
| BU-OPT | 0.51 [0.05, 2.37] | 0.00 [0.00, 19.94] | 0.00 [-3.60, 1.11] | -1.94 [-15.51, 0.00] |
| SC-OPT | 0.53 [0.06, 2.31] | 0.00 [0.00, 28.81] | 0.00 [-3.32, 1.39] | -2.22 [-15.24, 0.00] |
| PE-OPT | 0.52 [0.05, 2.70] | 0.00 [0.00, 21.88] | 0.00 [-4.16, 0.83] | -2.22 [-17.17, 0.00] |
| MLR-OPT | 0.31 [0.05, 1.24] | 0.00 [0.00, 1.11] | 0.00 [-1.39, 1.94] | -1.39 [-6.65, 0.00] |
| LASSO-OPT | 0.29 [0.05, 1.15] | 0.00 [0.00, 1.11] | 0.00 [-1.39, 1.94] | -1.39 [-6.09, 0.00] |
| MLRNS-OPT | 0.27 [0.04, 0.85] | 0.00 [0.00, 0.00] | 0.28 [0.00, 3.60] | -1.66 [-4.99, 0.00] |

proximate the "optimal" bolus calculator.
As we did in Chapter 4, these metrics are reported as mean (± standard deviation) for Gaussian distributed metrics or median [interquartile range]. For this purpose, non-Gaussianity of each distribution was checked by mean of the Lilliefors test with a 5% confidence level.

### 5.7.1 Assessment against the "optimal" bolus calculator

Figure 5.7 shows the distributions of ΔBGRI, $\Delta\%T_{HYPO}$, $\Delta\%T_{HYPER}$ $\Delta\%T_{TARGET}$, obtained by comparing SF, BU, SC, PE, MLR, LASSO, and MLRNS against OPT on the test set. It is possible to observe that, consistently with what we obtained in the Section 5.6, MLR, LASSO, MLRNS outperform the other considered methods. In particular, MLR, LASSO, and MLRNS distributions show that they are all better approximation of OPT compared to SF, BU, SC, and PE. Indeed, Figure 5.7 shows that MLR-OPT, LASSO-OPT, and MLRNS-OPT distributions are squished around zero. Analyzing the numerical results reported in Table 5.2 this aspect can be furtherly appreciated. For example, focusing on the third quartile of $\Delta\%T_{HYPO}$, obtained results are 19.11 for SF-OPT, 19.94 for BU-OPT, 28.81 for PE-OPT, 1.11 for MLR-OPT, 1.11 for LASSO-OPT, and

**Table 5.3**: Median [interquartile range] ΔBGRI, $\Delta\%T_{HYPO}$, $\Delta\%T_{HYPER}$ $\Delta\%T_{TARGET}$, obtained by comparing MLR vs. LASSO, MLR vs. MLRNS, and LASSO vs. MLRNS on the test set.

| | **ΔBGRI** | **$\Delta\%T_{HYPO}$** | **$\Delta\%T_{HYPER}$** | **$\Delta\%T_{TARGET}$** |
|---|---|---|---|---|
| MLR-LASSO | 0.00 [-0.04, 0.08] | 0.00 [0.00, 0.00] | 0.00 [-0.28, 0.28] | 0.00 [-0.28, -0.28] |
| MLR-MLRNS | 0.00 [-0.24, 0.40] | 0.00 [0.00, 1.66] | -1.11 [-2.22, 0.00] | 0.55 [-0.83, 1.66] |
| LASSO-MLRNS | 0.01 [-0.26, 0.39] | 0.00 [0.00, 1.94] | -1.39 [-1.94, 0.00] | 0.83 [-0.83, 1.66] |

0.00 for MLRNS-OPT, demonstrating that the adoption of the developed models for the computation of $IMB$ is able to significantly reduce hypoglycemia within the virtual population. In particular, results obtained for MLRNS-OPT are expected since the cost function used during the training of MLRNS penalizes $IMB$ overestimation. Considering the median results (see Table 5.2), the best outcomes were achieved using MLRNS in terms of $\Delta BGRI$ and $\Delta\%T_{HYPO}$, LASSO in terms of $\Delta\%T_{TARGET}$, while both MLR and LASSO perform the same in terms of $\Delta\%T_{HYPER}$. Again, this is expected since higher prediction accuracy (MLR and LASSO) should translate in better $\Delta\%T_{TARGET}$, while better $\Delta\%T_{HYPO}$ should be obtained using MLRNS due to the nature of the cost function we used during its training.

### 5.7.2 Comparison among MLR, LASSO, and MLRNS

Figure 5.8 shows the distributions of ΔBGRI, $\Delta\%T_{HYPO}$, $\Delta\%T_{HYPER}$ and $\Delta\%T_{TARGET}$ obtained by comparing MLR vs. LASSO, MLR vs. MLRNS and LASSO vs. MLRNS on the test set. Numerical results are reported in Table 5.3. It can be noticed that, MLR and LASSO obtain basically the same results being the ΔBGRI, $\Delta\%T_{HYPER}$ and $\Delta\%T_{TARGET}$ distributions symmetrical around 0 and all median values equal to 0.00. Moreover, there are no differences in terms of $\Delta\%T_{HYPO}$. Numerical results are reported in Table 5.3. Different results are obtained comparing MLR vs. MLRNS. Of course, this is due to the asymmetric cost function used to train MLRNS. Indeed, using MLRNS we get $\%T_{HYPO}$ but higher $\%T_{HYPER}$. However, obtained $\Delta\%T_{HYPER}$ is acceptable (on average -1.11 %). Same results are obtained comparing LASSO vs. MLRNS. Indeed, since MLR and LASSO obtained the same results, this is expected.

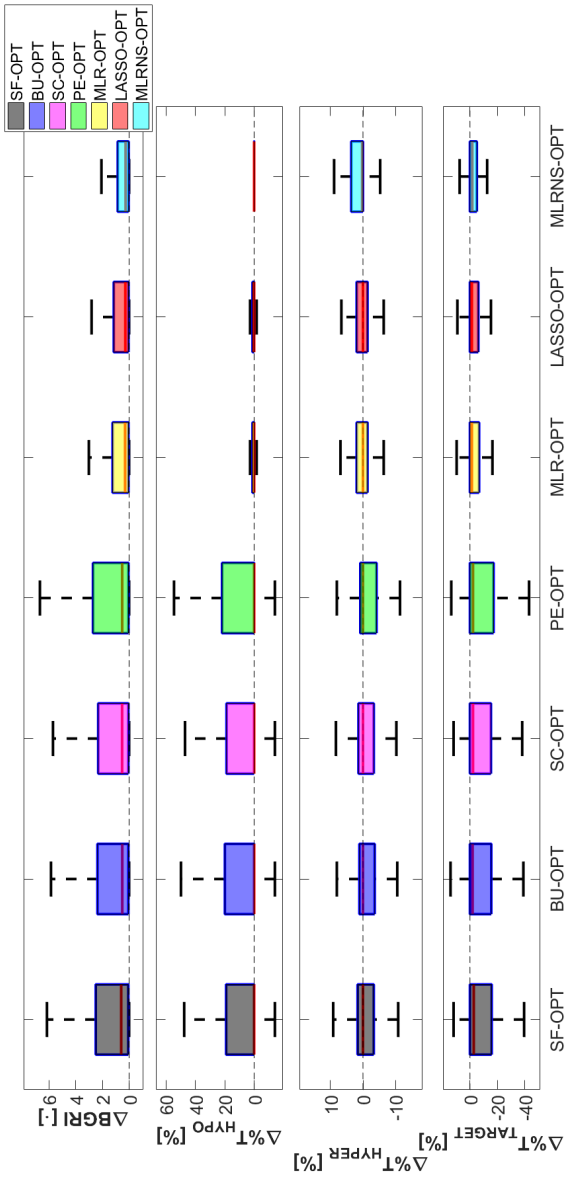



**Figure 5.7:** Boxplot distributions of ΔBGRI, $\Delta\%T_{HYPO}$, $\Delta\%T_{HYPER}$ $\Delta\%T_{TARGET}$, obtained by comparing SF, BU, SC, PE, MLR, LASSO, and MLRNS against OPT on the test set.

## 5.8 Summary of the obtained results and ideas for future developments

In this chapter we developed new models for the calculation of insulin boluses, exploiting machine learning approaches, with the aim of improving the standard formula used for bolus insulin dosage. Results show that linear re-

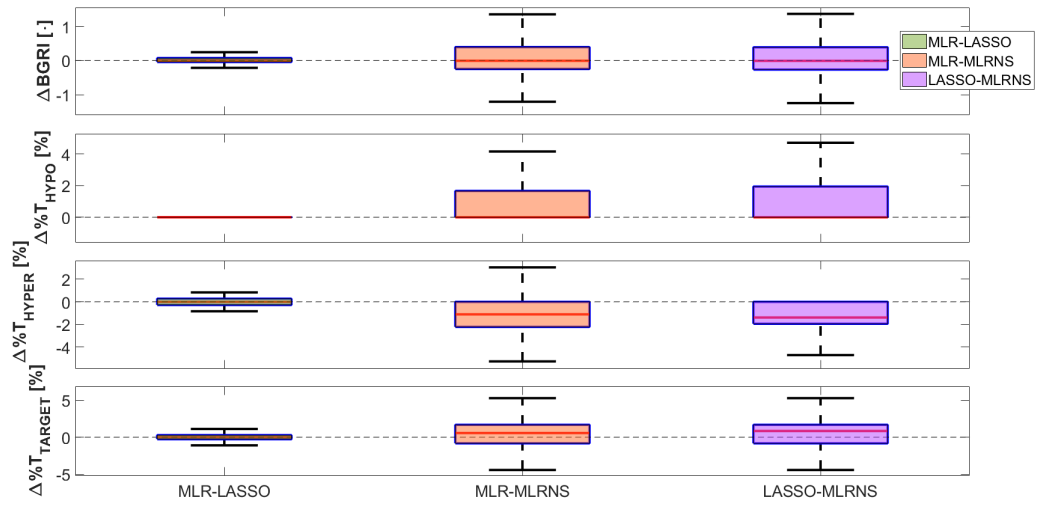



**Figure 5.8:** Boxplot distributions of ΔBGRI, $\Delta\%T_{HYPO}$, $\Delta\%T_{HYPER}$ $\Delta\%T_{TARGET}$, obtained by comparing MRL vs. LASSO, MLR vs. MRLNS, LASSO vs. MLRNS on the test set.

gression models are viable techniques for the scope. In particular, they outperform SF, BU, SC, and PE in approximating OPT by maintaining interpretability, simplicity, and, from a practical point of view, being easily implementable in currently available insulin pumps as a substitute of SF. Moreover, a comparison between the glycemic outcomes obtained with the proposed models and literature methodologies for $IMB$ calculation is performed. As a result, we can state that each of the models makes a significant contribution to this improvement. In particular, use of LASSO, allowed to obtain results closer to what can be achievable with the "optimal" BC, i.e. OPT, suggesting that it could represent a promising method for the diabetes treatment management. Furthermore, we would like to highlight the significant improvement brought by MLRNS. When tested in silico the model reaches the minimum average time spent in the hypoglycemic range among all the proposed models at the cost of a slightly increase of the time spent in hyperglycemia. This result will lead us to further investigate this type of approaches method during future developments of the this line of research. In particular, MLRNS represents the ideal $IMB$ calculation methodology for those patients that, are highly insulin sensitive and, more in general, they have to necessarily avoid $IMB$ overestimation.

# Chapter 6

# Conclusion and future development

## 6.1 Summary of the main achievements

Use of CGM devices opened new scenarios in T1D management. In particular, FDA approval of non-adjunctive use of CGM increased the interest toward methods to "correct" insulin therapy in order to account for CGM-derived information such as the glucose ROC. The primary aim of this thesis was to develop novel insulin dosing techniques, usable for T1D treatment, to leverage this information and improve glycemic control through the population.
For this purpose, in Chapter 2 we first reviewed three popular methods, namely BU, SC, and PE, that use ROC magnitude and direction to correct SF. The aim was to evaluate whether these algorithms help in improving the performance of SF and to identify their criticalities and limitations as well as margins for improvement by new approaches. In particular, by an ad-hoc in silico clinical trial based on the 2013 version of the UVa/Padova T1D Simulator, we showed that, overall, none of the considered approaches clearly prevails on the others, since the best modulation of the insulin bolus results strongly related to preprandial conditions and patient characteristics.
The above results suggested us to explore new algorithms for the adjustment of SF which leverage not only CGM data, but also accessible patient parameters including, for example, preprandial BG concentration, body weight, and meal carbohydrate intake. Therefore, given their ability of capturing complex non linear input-output relationships, such as glucose-insulin dynamics, we decided to adopt a NN. Specifically, we assessed whether or not it can be applicable to the considered case of study, i.e. the insulin dosing personalization. Therefore, in Chapter 3 we developed a new algorithm to correct SF, based on

NN, trained to learn the optimal insulin dose using the SF parameters, ROC, body weight, insulin pump basal infusion rate, and insulin as features. Results on simulated data show that this approach outperforms the other considered algorithms by bringing to a small but statistically improvement of glucose control. Moreover, this suggested us to investigate other strategies based on machine learning models to furtherly improve glycemic control across the population.

Following these results, in Chapter 4 we proposed a different personalized approach to SF adjustment. Specifically, a new model based on XGB was used firstly to classify, at meal time, the postprandial glycemic status using CGM data together with carbohydrate intakes data and insulin infusion recordings, then its outcome was used to modulate, in real-time, the insulin bolus amount obtained through SF. Results obtained on simulated data show that, when used to adjust SF, the proposed approach improves glycemic control. In particular, percentage time in target [70, 180] mg/dL improved without increasing hypoglycemia.

The results we obtained in Chapter 3 and Chapter 4 showed that our strategy, i.e. leveraging machine-learning power to learn the underneath relationships between optimal insulin dose and patient status, is promising in improving the glycemic outcomes achievable through the use of SF. Indeed, we showed that using SF alone one obtains bad glucose control due to its definition. However, this aspect is critical. In fact, since both methodologies were designed to correct something sub optimal, this can represent a possible limitation. Therefore, this drove us to develop new models for insulin dosing that get rid of SF and provide patients with new rules that naturally take into account for the information on their physiological status. For this scope, in Chapter 5 we investigated the use of three different LR models labelled in the thesis MLR, LASSO, and MLRNS. Results on simulated data demonstrate the superiority of MLR, LASSO, and MLRNS when compared to SF, BU, SC, and PE.

In conclusion, the developed algorithms are promising tools to improve glycemic control in patients affected with T1D, since they properly use the available information on patient, available at meal-time, to provide effective insulin dosage.

## 6.2 Limitations of the study

Regarding the review of state-of-the-art techniques to correct SF according to CGM data carried out in Chapter 2, we acknowledge that it is not exhaustive since it does not include the latest methodologies published after the beginning of that part of the work, e.g. the one proposed by Aleppo et al. [90], Ziegler et al. [91], and Klonoff et al. [92]. For this reason, work currently under way involve their inclusion in a new study aimed at comparing their performance with BU, SC, and PE [1].

As far as the NN-based methodology of Chapter 3 is concerned, it is important to stress that, in the present investigation of tools to predict the optimal correction of SF, we limited ourselves to evaluating neural networks only. Of course, other non linear machine learning techniques (e.g., kernel support vector machines or regression trees) could be considered for the scope as well. Implementation of alternative methods will be matter of future investigations, together with a comprehensive analysis of the relative performance of the so-obtained SF "corrector". In addition, another interesting aspect would be furtherly exploring both the NNC structure by reducing the number of nodes in the middle layer and introducing fuzzification as a way to accommodate person-to-person variation and expanding the feature set we used to train NNC. For example, it will be worth adding as inputs also a preprandial window of CGM values to exploit fully the information on BG dynamic provided by CGM devices. Finally, it would be also interesting to investigate how the NNC performance is influenced by the variability of CGM sensor accuracy, for example, observed for different days of sensor wear or different number of sensor's calibrations per day.

Regarding the development of the methodology based on XGB, described in Chapter 4, the presented study has some limitations that need to be addressed in future work. An important aspect that needs to be further studied is the feature selection process. Given the inherent capability of XGB at automatically discriminating important features, no preliminary feature selection was performed during the data preparation phase. However, future work is needed to investigate if it is possible to reduce its dimensionality and improve the interpretability of the results. Additional work can be done to develop a more

[1] G. Noaro, G. Cappon, S. Del Favero, G. Sparacino, and A. Facchinetti,"In silico assessment of insulin dosing algorithms accounting for glucose rate of change provided by continuous glucose monitoring data," in the 13th International Conference on Advanced Technology Treatment for Diabetes, ATTD, 2020. (submitted)

specific and optimized policy to adjust the meal insulin bolus according to XGB classifications. Furthermore, future work will study the impact of each of the input features in characterizing the target classes. Indeed, such a study is fundamental to fully interpret XGB and furtherly improve its performance. Finally, future work will also compare XGB with neural networks and deep learning techniques to assess whether, by sacrificing model interpretability, it is possible to achieve better classification accuracy.

Some additional comments can be done about the methodologies we presented in Chapter 5. Describing the $IMB$ dose with a model which is linear in the parameters is challenging, although using a linear model has many advantages, in terms of interpretability but also computational burden. Indeed, with high probability, a non linear model could further improve the performances, since glucose and insulin dynamics are non linear. A first attempt exploiting gradient boosted decision tree was carried out. This type of non linear model has been chosen for its comprehensibility compared to other types of non linear algorithms such as NNs or support vector machines. Current works undergoing in our laboratory show that the application of non linear models can further improve the performance, but the results should be tested in-silico and the method further investigated.

Finally, of course the results obtained using a state-of-the-art in silico environment are very encouraging. However, in order to fully validate each methodology, the next logical step will certainly involve their assessment using retrospective clinical data and its subsequent evaluation in a prospective clinical trial. With respect to this aspect, the developed methodologies, especially NNC and the new formula for insulin dosing developed in Chapter 5, share the same limitation: their training will not be possible in a clinical setting since no such a thing as the optimal target bolus will be available. To overcome this issue, a possible solution is represented by, first training models on in silico data generated using the UVa/Padova T1D Simulator, then calibrating these model on real patient data in order to adjust model parameters, thus obtaining a model representation of the processes of interest on the real population. Future work will investigate extensively this approach.

## 6.3 Possible further developments

### 6.3.1 Methodological aspects

The results achieved in this thesis encourage further investigations on machine learning-based methodologies to provide patients with new tools to ease their daily insulin therapy routine. In this context, big data analysis can be exploited to discover daily or weekly patterns, whether patients perform better than others and even cluster each subject into different categories according to their outcomes. For example, a possible advantage of having "clusters" of patients instead of a single one is the possibility of using not only that subject information for the insulin bolus calculation, but also the whole cluster information, that, containing much more data, makes it possible to deploy powerful deep learning techniques. To give an example, it is known that, in general, the amount of injected insulin strongly depends on the $CR$ of a subject: a low $CR$ leads to higher amounts of insulin needed to compensate a meal, while subjects having a high $CR$ require less insulin compared to the previous case. Hence, the idea of exploiting this information applying clustering methodologies to the subjects, obtaining two (or more) clusters of patients which shows similar characteristics in terms of $CR$ parameter will be explored.

### 6.3.2 A possible application: development of a decision support system

The increased amount of available information brought by wearable devices, such as CGM systems and physical activity monitoring bands, has led to the development of decision making tools and applications, such as the methodologies developed in this thesis, that can enhance the management of the disease [93]. Decision support systems (DSSs) give the possibility to support users with proactive and personalized decisions in any scenario of their daily living and allows to react at shorter time scales. Over the past few years, DSSs for diabetes have been an emerging concept in health care. By means of this new technology, data can be automatically collected, transmitted, aggregated with other physiological data, analyzed, stored, and presented to the patient. By integrating e-health and tele-monitoring systems, DSSs for T1D have the potential to improve glycemic outcomes thanks to prevention of hypo- or hyperglycemic events, reducing uncertainty when making critical self-management decisions

[94]. A DSS for diabetes treatment provides an alternative to the closed-loop system, the so-called artificial pancreas (AP). Indeed, a wide range of users do not feel confident with the use of AP systems, being concerned about errors occurring in the insulin pump, and they prefer an open-loop therapy, which can be assisted by DSSs. Most of DSSs already available in the literature are composed by a predictive glucose module (which alerts the user whenever its BG is predicted to fall outside the safe range in the next future), an insulin suspension module (which temporarily suspends basal insulin delivery to avoid hypoglycemia when BG is critically low in patient using insulin pumps), and a BC module to compute $IMB$ at meal-time. Therefore, a straightforward application of the work presented in this thesis, consist in designing a new BC module that implements the proposed methodologies in order to develop a new, more sophisticated, DSS that is able to achieve better glycemic control when compared to other state-of-the-art algorithms.

# Bibliography


[1] K. Alberti and P. Zimmet, "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation," *Diabet Med*, vol. 15, pp. 539–553, 7 1998.

[2] D. Daneman, "Type 1 diabetes," *Lancet*, vol. 367, no. 9513, pp. 847–858, 2006.

[3] World Health Organization (WHO), "Diabetes facts sheet," *Available at: http://www.who.int/ mediacentre/factsheets/fs312/en/*.

[4] American Diabetes Association, "Classification and diagnosis of diabetes," *Diabetes Care*, vol. 40, no. S1, pp. S11–S24, 2017.

[5] C. Hayes and A. Kriska, "Role of physical activity in diabetes management and prevention," *J Am Diet Assoc*, vol. 108, no. 4, pp. S19–S23, 2017.

[6] M. Montagnana, M. Caputo, D. Giavarina, and G. Lippi, "Overview on self-monitoring of blood glucose," *Clin Chim Acta*, vol. 402, no. 1, pp. 7–13, 2009.

[7] L. Clores, and R. Man, "12 Tips for injecting insulin," *nursingcrib.com. Available at: https://nursingcrib.com/nursing-notes-reviewer/fundamentals-of-nursing/12-tips-for-injecting-insulin/*, 2015.

[8] Medtronic Inc., "What is insulin therapy?," *medtronic-diabetes.ie. Available at: https://www.medtronic-diabetes.ie/what-insulin-pump-therapy*, 2015.

[9] WorldChronicle24, "Self-Monitoring Blood Glucose Devices Market Size, Trends, Share and Forecast by 2026: Terumo Corporation, AgaMatrix Inc, Abbott," *WorldChronicle24.com. Available at:*

*https://worldchronicle24.com/2019/25455/self-monitoring-blood-glucose-devices-market-size-trends-share-and-forecast-by-2026-terumo-corporation-agamatrix-inc-abbott/*, 2019.

[10] S. Schmidt and K. Norgaard, "Bolus calculators," *J Diabetes Sci Technol*, vol. 8, no. 5, pp. 1035–1041, 2014.

[11] R. R. John Walsh, "Test and adjust your carb factor," in *Pumping Insulin. Everything you need for success on an insulin pump.*, ch. 12, pp. 141–158, San Diego, CA: Torrey Pines Press, 5.1 ed., 2012.

[12] R. R. John Walsh, "Test and adjust your correction factor," in *Pumping insulin. Everything you need for success on an insulin pump.*, ch. 13, pp. 159–172, San Diego, CA: Torrey Pines Press, 5.1 ed., 2012.

[13] P. C. Davidson, H. R. Hebblewhite, R. D. Steed, and B. W. Bode, "Analysis of guidelines for basal-bolus insulin dosing: Basal insulin, correction factor, and carbohydrate-to-insulin ratio," *Endocr Prat*, vol. 14, no. 9, pp. 1095–1101, 2008.

[14] T. Gross, D. Kayne, A. King, C. Rother, and S. Juth, "A bolus calculator is an effective means of controlling postprandial glycemia in patients on insulin pump therapy," *Diabetes Technol Ther*, vol. 5, no. 3, pp. 365–369, 2003.

[15] S. Garg, T. Bookout, K. McFann, W. Kelly, C. Beatson, S. Ellis, R. Gutin, and P. Gottlieb, "Improved glycemic control in intensively treated adult subjects with type 1 diabetes using insulin guidance software," *Diabetes Technol Ther*, vol. 10, no. 5, pp. 369–375, 2008.

[16] S. Marden, P. Thomas, and S. Z.A, "Poor numeracy skills are associated with glycaemic control in type 1 diabetes," *Diabet Med*, vol. 29, no. 5, pp. 662–669, 2012.

[17] Roche, "Accu-Check Connect app," *Available at: https://www.accu-chek.com/apps-and-software/connect-app*.

[18] K. M. Miller, R. W. Beck, R. M. Bergenstal, R. S. Goland, M. J. Haller, J. B. McGill, H. Rodriguez, J. H. Simmons, I. B. Hirsch, and T1D Exchange Clinic Network, "Evidence of a strong association between frequency of

self-monitoring of blood glucose and hemoglobin A1c levels in T1D exchange clinic registry participants," *Diabetes Care*, vol. 36, no. 7, pp. 2009–2014, 2013.

[19] A. Pfützner, J. Weissmann, S. Mougiakakou, E. Daskalaki, N. Weis, and R. Ziegler, "Glycemic variability is associated with frequency of blood glucose testing and bolus: Post hoc analysis results from the ProAct study," *Diabetes Technol Ther*, vol. 17, no. 6, pp. 392–7, 2015.

[20] A. J. Graveling and B. M. Frier, "Impaired awareness of hypoglycaemia: A review," *Diabetes Metab*, vol. 36 Suppl 3, pp. 64–74, 2010.

[21] A. E. Gold, K. M. Macleod, and B. M. Frier, "Frequency of severe hypoglycemia in patients with type I diabetes with impaired awareness of hypoglycemia," *Diabetes Care*, vol. 17, no. 7, pp. 697–703, 1994.

[22] J. Walsh, R. Roberts, and T. Bailey, "Guidelines for optimal bolus calculator settings in adults," *J Diabetes Sci Technol*, vol. 5, no. 1, pp. 129–135, 2011.

[23] A. B. King, A. Kuroda, M. Matsuhisa, and T. Hobbs, "A review of insulin-dosing formulas for continuous subcutaneous insulin infusion (CSII) for adults with type 1 diabetes," *Curr Diab Rep*, vol. 16, no. 83, 2016.

[24] C.E. Smart, B.R. King, P. McElduff and C.E. Collins, "In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia," *Diabetic Med*, vol. 29, no. 7, pp. e21–e24, 2012.

[25] J. E. Lane, J. P. Shivers, and H. Zisser, "Continuous glucose monitors: Current status and future developments," *Curr Opin Endocrinol Diabetes Obes*, vol. 20, no. 2, pp. 106–111, 2013.

[26] G. Cappon, G. Acciaroli, M. Vettoretti, A. Facchinetti and G. Sparacino, "Wearable continuous glucose monitoring sensors: A revolution in diabetes treatment," *Electronics*, vol. 6, no. 3, p. 65, 2017.

[27] G. Cappon, M. Vettoretti, G. Sparacino and Andrea Facchinetti, "Continuous glucose monitoring sensors for diabetes management: A review of technologies and applications," *Diabetes Metab J*, vol. 43, no. 4, pp. 383–397, 2019.

[28] S.E. Clarke and J.R. Foster, "A history of blood glucose meters and their role in self-monitoring of diabetes mellitus," *J Biomed Sci*, vol. 69, no. 4, pp. 83–93, 2012.

[29] M. Christiansen, T. Bailey, E. Watkins, D. Liljenquist, D. Price, K. Nakamura, R. Boock, and T. Peyser, "A new-generation continuous glucose monitoring system: Improved accuracy and reliability compared with a previous-generation system," *Diabetes Technol Ther*, vol. 15, no. 10, pp. 881–888, 2013.

[30] T. S. Bailey, A. Chang, and M. Christiansen, "Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm," *J Diabetes Sci Technol*, vol. 9, no. 2, pp. 209–214, 2015.

[31] J. R. Castle and P. G. Jacobs, "Nonadjunctive use of continuous glucose monitoring for diabetes treatment decisions," *J Diabetes Sci Technol*, vol. 10, no. 5, 2016.

[32] T. Battelino, I. Conget, B. Olsen, I. Schütz-Fuhrmann, E. Hommel, R. Hoogma, U. Schierloh, N. Sulli, J. Bolinder, and SWITCH Study Group, "The use and efficacy of continuous glucose monitoring in type 1 diabetes treated with insulin pump therapy: A randomised controlled trial," *Diabetologia*, vol. 55, no. 12, pp. 3155–3162, 2012.

[33] S. Garg, H. Zisser, S. Schwartz, T. Bailey, R. Kaplan, S. Ellis, and L. Jovanovic, "Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor: A randomized controlled trial," *Diabetes Care*, vol. 29, no. 1, pp. 44–50, 2006.

[34] T. Battelino, M. Phillip, N. Bratina, R. Nimri, P. Oskarsson, and J. Bolinder, "Effect of continuous glucose monitoring on hypoglycemia in type 1 diabetes," *Diabetes Care*, vol. 34, no. 4, pp. 795–800, 2011.

[35] S. V. Edelman and T. S. Bailey, "Continuous glucose monitoring health outcomes," *Diabetes Technol Ther*, vol. 11 Suppl 1, pp. 68–74, 6 2009.

[36] R. M. Bergenstal, W. V. Tamborlane, A. Ahmann, J. B. Buse, G. Dailey, S. N. Davis, C. Joyce, T. Peoples, B. A. Perkins, J. B. Welsh, S. M. Willi, M. A. Wood, and STAR 3 Study Group, "Effectiveness of sensor-augmented insulin-pump therapy in type 1 diabetes," *N Engl J Med*, vol. 363, no. 4, pp. 311–320, 2010.

[37] The Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group, "Continuous glucose monitoring and intensive treatment of type 1 diabetes," *N Engl J Med*, vol. 359, no. 14, pp. 1464–1476, 2008.

[38] R. Gifford, "Continuous glucose monitoring: 40 years, what we've learned and what's next," *Chemphyschem*, vol. 14, no. 10, pp. 2032–2044, 2013.

[39] M. Vettoretti, G. Cappon, G. Acciaroli, A. Facchinetti and G. Sparacino, "Continuous glucose monitoring: current use in diabetes management and possible future applications," *J Diabetes Sci Technol*, vol. 12, no. 5, pp. 1064–1071, 2018.

[40] P. Herrero, P. Pesl, J. Bondia, M. Reddy, N. Oliver, P. Georgiou, and C. Toumazou, "Method for automatic adjustment of an insulin bolus calculator: In silico robustness evaluation under intra-day variability," *Comput Methods Programs Biomed*, vol. 119, no. 1, pp. 1–8, 2015.

[41] P. Herrero, P. Pesl, M. Reddy, N. Oliver, and P. Georgiou, "Advanced insulin bolus advisor based on run-to-run control and case-based reasoning," *IEEE J BIomed Health Inform*, vol. 15, no. 3, pp. 1087–1096, 2015.

[42] Dexcom Inc., "Dexcom G5 system," *Available at: https://www.dexcom.com/g5-cgm-system*.

[43] Senseonics Inc., "Eversense system," *Available at: https://www.eversensediabetes.com/eversense-cgm-system*.

[44] B. Buckingham, D. Xing, S. Weinzimer, R. Fiallo-Scharer, C. Kollman, N. Mauras, E. Tsalikian, W. Tamborlane, T. Wysocki, K. Ruedy, and R. Beck, "Use of the DirecNet Applied Treatment Algorithm (DATA) for diabetes management with a real-time continuous glucose monitor (the FreeStyle Navigator)," *Pediatr Diabetes*, vol. 9, no. 2, pp. 142–147, 2008.

[45] G. Scheiner, "Maximizing the benefits in real time," in *Practical CGM: Improving patient outcomes through continuous glucose monitoring*, ch. 2, pp. 23–25, Alexandria: American Diabetes Association, 1st ed., 2015.

[46] J. Pettus and S. V. Edelman, "Recommendations for using real-time continuous glucose monitoring (rtCGM) data for insulin adjustments in type 1 diabetes," *J Diabetes Sci Technol*, 2016.

[47] U.S. Food and Drug Administration, "FDA executive summary âĂŞ Dexcom G5 Mobile continuous glucose monitoring system," *Available at: http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/ClinicalChemistryandClinicalToxicologyDevicesPanel/UCM511810.pdf*, 2016.

[48] J. A. Dimasi, "Cost of developing new drugs," *Available at: http://csdd.tufts.edu/files/uploads/Tufts_CSDD_briefing_on_RD_cost_study_-_Nov_18,_2014..pdf*, 2014.

[49] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. Orsini Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering, and M. E. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol Meas*, vol. 25, no. 4, pp. 905–920, 2004.

[50] M. E. Wilinska, L. J. Chassin, C. L. Acerini, J. M. Allen, D. B. Dunger, and R. Hovorka, "Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes," *J Diabetes Sci Technol*, vol. 4, no. 1, pp. 132–144, 2010.

[51] S. S. Kanderian, S. Weinzimer, G. Voskanyan, and G. M. Steil, "Identification of intraday metabolic profiles during closed-loop glucose control in individuals with type 1 diabetes," *J Diabetes Sci Technol*, vol. 3, no. 5, pp. 1047–1057, 2009.

[52] A. Haidar, M. E. Wilinska, J. A. Graveston, and R. Hovorka, "Stochastic virtual population of subjects with type 1 diabetes for the assessment of closed-loop glucose controllers," *IEEE Trans Biomed Eng*, vol. 60, no. 12, pp. 3524–3533, 2013.

[53] C. Dalla Man, R. a. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Trans Biomed Eng*, vol. 54, no. 10, pp. 1740–1749, 2007.

[54] R. Basu, C. Dalla Man, M. Campioni, A. Basu, G. Klee, G. Jenkins, G. Toffolo, C. Cobelli, and R. A. Rizza, "Mechanisms of postprandial hyperglycemia in elderly men and women: Gender specific differences in insulin secretion and action," *Diabetes*, vol. 55, pp. 2001–2014, 2006.

[55] R. Visentin, C. Dalla Man, Y. C. Kudva, A. Basu, and C. Cobelli, "Circadian variability of insulin sensitivity: Physiological input for in silico artificial pancreas," *Diabetes Technol Ther*, vol. 17, no. 1, pp. 1–7, 2015.

[56] L. Hinshaw, C. Dalla Man, D. K. Nandy, A. Saad, A. E. Bharucha, J. A. Levine, R. A. Rizza, R. Basu, R. E. Carter, C. Cobelli, Y. C. Kudva, and A. Basu, "Diurnal pattern of insulin action in type 1 diabetes: Implications for a closed-loop system," *Diabetes*, vol. 62, no. 7, pp. 2223–9, 2013.

[57] R. Visentin, C. Dalla Man, and C. Cobelli, "One-day Bayesian cloning of type 1 diabetes subjects: Towards a single-day UVA/Padova type 1 diabetes simulator," *IEEE Trans Biomed Eng*, vol. 63, no. 11, pp. 2416–2424, 2016.

[58] R. Visentin, E. Campos-Nanez, M. Schiavon, D. Lv, M. Vettoretti, M. Breton, B.P. Kovatchev, C. Dalla Man and C. Cobelli, "The UVA/Padova Type 1 Diabetes Simulator Goes From Single Meal to Single Day," *J Diabetes Sci Technol*, vol. 12, no. 2, pp. 273–281, 2018.

[59] C. Marling and R. Bunescu, "The OhioT1DM dataset for blood glucose level prediction," *The 3rd International Workshop on Knowledge Discovery in Healthcare Data*, 2018.

[60] T. Hastie, R. Tibshirani, and J. H. Friedman, "The elements of statistical learning: data mining, inference, and prediction," *Springer, 2nd Edition, Springer Series in Statistics*, 2009.

[61] C. M. Bishop, "Pattern recognition and machine learning," *Springer New York*, vol. 4.

[62] I. Contreras, and J. Vehi, "Artificial intelligence for diabetes management and decision support: literature review," *J Med Internet Res*, vol. 20, no. 5, p. e10775, 2018.

[63] G. Cappon, F. Marturano, M. Vettoretti, A. Facchinetti, and G. Sparacino, "In silico assessment of literature insulin bolus calculation methods accounting for glucose rate of change," *J Diabetes Sci Technol*, vol. 13, no. 1, pp. 103–110, 2019.

[64] J. Pettus, D. A. Price, and S. V. Edelman, "How patients with type 1 diabetes translate continuous glucose monitoring data into diabetes management decisions," *Endocr Pract*, vol. 21, no. 6, pp. 613–620, 2015.

[65] J. Pettus and S.V. Edelman, "Use of glucose rate of change arrows to adjust insulin therapy among individuals with type 1 diabetes who use continuous glucose monitoring," *Diabetes Technol Ther*, vol. 18, no. S2, pp. S2–S34, 2016.

[66] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli, "The UVA/PADOVA type 1 diabetes simulator: New features," *J Diabetes Sci Technol*, vol. 8, no. 1, pp. 26–34, 2014.

[67] B. P. Kovatchev, E. Otto, D. Cox, L. Gonder-Frederick, and W. Clarke, "Evaluation of a new measure of blood glucose variability in diabetes," *Diabetes Care*, vol. 29, no. 11, pp. 2433–2438, 2006.

[68] C. Fabris, S. D. Patek, and M. D. Breton, "Are risk indices derived from CGM interchangeable with SMBG-based indices?," *J Diabetes Sci Technol*, vol. 10, no. 1, pp. 50–59, 2016.

[69] B. P. Kovatchev, D. J. Cox, L. Gonder-Fredrick, and W. Clarke, "Symmetrization of the blood glucose measurement scale and its applications," *Diabetes Care*, vol. 10, no. 1, pp. 50–59, 1997.

[70] G. Cappon, M. Vettoretti, F. Marturano, A. Facchinetti, and G. Sparacino, "A neural network based approach to personalize insulin bolus calculation using continuous glucose monitoring," *J Diabetes Sci Technol*, vol. 12, no. 2, pp. 265–272, 2018.

[71] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J Machine Learning Res*, vol. 15, pp. 1929–1958, 2014.

[72] G. Hinton, N. Srivastava, and K. Swersky, "Overview of mini-batch gradient descent," *Neural Network for Machine Learning, Coursera Course. Available at: http://www.cs.toronto.edu/tijmen/csc321/slides/lecture_slides_lec6.pdf*, 2017.

[73] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kegl, "Algorithms for hyper-parameter optimization," *Advances in Neural Information Processing Systems 24*, pp. 2546–2554, 2011.

[74] G. Cappon, A. Facchinetti, G. Sparacino, P. Georgiou, and P. Herrero, "Classification of postprandial glycemic status with application to in-

sulin dosing in type diabetes: An in silico proof of concept," *Sensors*, vol. 19, p. e3168, 2019.

[75] T. Chen, and C. Guestrin, "XGBoost: A scalable tree boosting system," *22nd ACM SIGDD Int Conf on Knowledge Discovery*, pp. 13–17, 2016.

[76] J. Walsh, R. Roberts, and L. Heinemann, "Confusion regarding duration of insulin action: A potential source for major insulin dose errors by bolus calculators.," *J Diabetes Sci Technol*, vol. 8, no. 1, pp. 170–178, 2014.

[77] S. Guerra, G. Sparacino, A. Facchinetti, M. Schiavon, C. Dalla Man, and C. Cobelli, "A dynamic risk measure from continuous glucose monitoring data," *Diabetes Technol Ther*, vol. 13, no. 8, pp. 843–852, 2011.

[78] M. E. Willinska, L. J. Chassin, H. C. Schaller, L. Schaupp, T. R. Pieber, and R. Hovorka, "Insulin kinetics in type 1 diabetes: Continuous and bolus delivery of rapid acting insulin," *IEEE Trans Biomed Eng*, vol. 52, no. 1, pp. 3–12, 2005.

[79] D. P. Howsmon, F. Cameron, N. Baysal, T. T. Ly, G. P. Forlenza, D. M. Maahs, and B. A. Buckingham, "Continuous glucose monitoring enables the detection of losses in infusion set actuation (LISAs)," *Sensors*, vol. 17, no. 1, p. e161, 2017.

[80] C. M. Florkowski, "Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: Communicating the performance of diagnostic tests," *Clin Biochem Rev*, vol. 29, no. S1, pp. S83–S87, 2017.

[81] M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, "Type 1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments," *IEEE Trans Biomed Eng*, vol. 65, no. 6, pp. 1281–1290, 2017.

[82] M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, "A model of self-monitoring blood glucose measurement error," *J Diabetes Sci Technol*, vol. 11, no. 4, pp. 724–735, 2017.

[83] A. Facchinetti, S. Del Favero, G. Sparacino, J. R. Castle, W. K. Ward, and C. Cobelli, "Modeling the glucose sensor error," *IEEE Trans Biomed Eng*, vol. 61, pp. 620–9, 3 2014.

[84] D. M. Maahs, B. A. Buckingham, J. R. Castle, A. Cinar, E. R. Damiano, E. Dassau, J. H. DeVries, F. J. Doyle, S. C. Griffen, A. Haidar, L. Heinemann, R. Hovorka, T. W. Jones, C. Kollman, B. Kovatchev, B. L. Levy, R. Nimri, D. N. O'Neal, M. Philip, E. Renard, S. J. Russell, S. A. Weinzimer, H. Zisser, and J. W. Lum, "Outcome measures for artificial pancreas clinical trials: A consensus report," *Diabetes Care*, vol. 39, no. 7, pp. 1175–1179, 2016.

[85] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, T. Diekotter, J. Garcia Marquez, B. Gruber, B. Lafourcade, P. Leitao, T. Munkemuller, C. Mcclean, P. Osborne, B. Reineking, B. Schroder, A. Skidmore, D. Zurell, and S. Lautenbach, "Collinearity: A review of methods to deal with it and a simulation study evaluating their performance," *Ecography*, vol. 36, pp. 27–46, 2013.

[86] M. Schmidt, G. Fung, and R. Rosales, "Optimization methods for l1-regularization," *University of British Columbia, Technical Report TR-2009-19*, 2009.

[87] W. J. Fu, "Penalized regression: the bridge versus the lasso," *J Comput Graph Stat*, vol. 7, no. 3, pp. 397–416, 1998.

[88] S. Perkins, K. Lacker, and J. Theiler, "Grafting: fast incremental feature selection by gradient descent in function space," *J Mach Learn Res*, vol. 3, pp. 1333–1356, 2003.

[89] A. Galen, and G. Jianfeng, "Scalable training of l1-regularized log-linear models," *Proceedings of the 24th Internation Conference on Machine Learning*, pp. 33–40, 2007.

[90] G. Aleppo, L. M. Laffel, A. J. Ahmann, I. B. Hirsch, D. F. Kruger, A. Peters, R. S. Weinstock, and D. R. Harris, "A practical approach to using trend arrows on the Dexcom G5 CGM system for the management of adults with diabetes," *J Endocr Soc*, vol. 1, no. 12, pp. 1445–1460, 2019.

[91] R. Ziegler, S. Von Sengbusch, J. Kroger, O. Schubert, P. Werkmeister, D. Deiss, T. Siegmund, "Therapy adjustments based on trend arrows using continuous glucose monitoring systems," *J Diabetes Sci Technol*, vol. 13, no. 4, pp. 763–773, 2019.

[92] D. Klonoff, and D. Kerr, "A simplified approach using rate of change arrows to adjust insulin with real-time continuous glucose monitoring," *J Endocr Soc*, vol. 11, no. 6, pp. 1063–1069, 2018.

[93] P. J. O'Connor, J. M. Sperl-Hillen, C. J. Fazio, B. M. Averbeck, B. H. Rank, and K. L. Margolis, "Outpatient diabetes clinical decision support: current status and future directions," *Diabet Med*, vol. 33, no. 6, pp. 731–741, 2016.

[94] P. Jia, P. Zhao, J. Chen, and M. Zhang, "Evaluation of clinical decision support systems for diabetes care: an overview of current evidence," *J Eval Clin Pract*, vol. 25, no. 1, pp. 66–77, 2019.

[95] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Phycol Rev*, vol. 62, no. 6, pp. 386–408, 1958.

[96] J. Bruck, and J. W. Goodman, "On the power of neural networks for solving hard problems," *J Compl*, vol. 6, no. 2, pp. 129–135, 1990.

[97] F. Hutter, H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," *LION-5, Extended version as UBC Tech report TR-2010-10*, 2011.

[98] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *J Global Opt*, vol. 21, pp. 345–383, 2001.

[99] P. Li, Q. Wu, and C. J. Burges, "Mcrank: Learning to rank using multiple classification and gradient boosting," *Proceedings of the 20th Conference on Advances in Neural Information Processing Systems*, pp. 897–904, 2008.

[100] S. Tyree, K. Weinberger, K. Agrawal, and J. Paykin, "Parallel boosted regression trees for web search ranking," *Proceedings of the 20th Conference on World Wide Web*, pp. 387–396, 2011.

# Appendix A

# Neural networks (NN)

## A.1 Introduction to neural networks

Neural networks (NNs) are complex and powerful machine learning based models originally devised to represent the mechanisms regulating information processing within biological systems [95]. So far, NN have been used to solve a variety of problems [96]. Moreover, NN base structure has been exploited to develop new, more sophisticated models to represent peculiar input-output relationships, e,g, recurrent NN, LSTM, and so and so forth. In this thesis, we will refer to the term NN as a specific class of NNs that have been proven to be of great practical value for both classification (i.e. when the target variable domain is a discrete finite set of values) and regression (i.e. when the target variable domain is a continuous set of values) tasks, namely the multilayer perceptron.

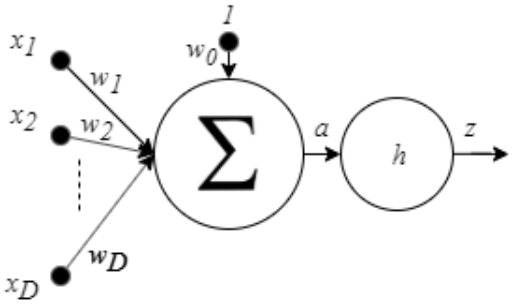



**Figure A.1:** Structure of a perceptron, i.e. a NN made of a single neuron.

**The most basic NN model: the perceptron**

Figure A.1 shows the perceptron, i.e. a NN composed of a single "neuron". Formally, it can be described as a functional transformations. First, it constructs a linear combination of the input variables $x_1, x_2, \ldots, x_D$ in the form:

$$a = \sum_{i=1}^{D} w_i x_i + w_0 \tag{A.1}$$

where $w_i$ and $w_0$ are commonly referred as weights and bias, respectively, and $a$ is the so-called activation. Then, the activation is transformed using a differentiable, non linear function $h(\cdot)$, namely the activation function, to obtain:

$$z = h(a) \tag{A.2}$$

where $z$ is the output of the perceptron.

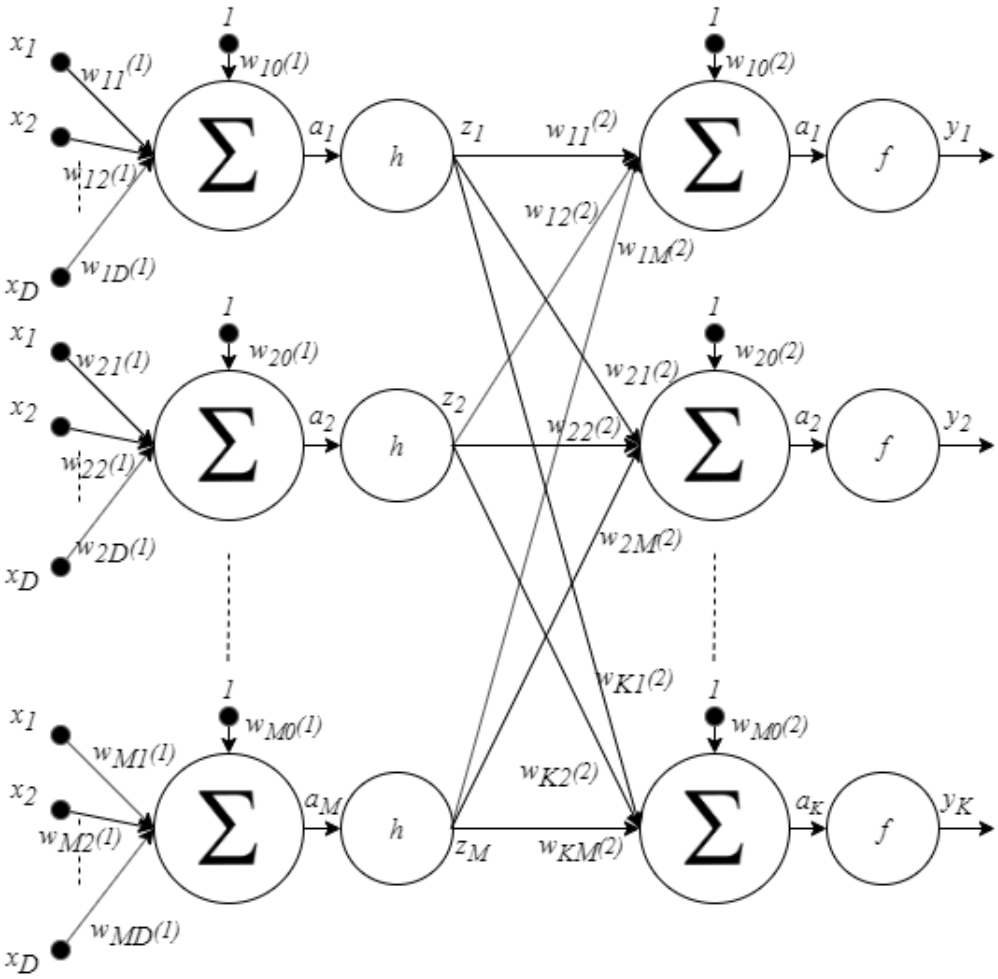



**Figure A.2:** A simple NN composed of 2 hidden layers, $D$ inputs and $K$ outputs.

**Putting perceptrons together** In general, NN models consist of an ensemble of perceptrons connected together to form the overall model. In particular, NN are commonly organized in "layers" where each layer is composed of multiple neurons each of them having as input all the outputs of the previous layer and a specific activation function.
To visualize this concept, let's describe how the simple NN model showed in Figure A.2 works. First, a set of inputs, i.e. $x_1, x_2, \ldots, x_D$, are used to fed the first layer of the NN (called the first "hidden" layer of the NN). Specifically, in our case, the first hidden layer consists of a set of $M$ perceptrons, which transform the inputs into $M$ linear combinations:

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \tag{A.3}$$

where superscript $(1)$ indicates the first hidden layer, and subscript $j = 1, \ldots, M$ denotes the j-th neuron of the first hidden layer. Then, Equation A.2 can be rewritten as:

$$z_j = h(a_j) \tag{A.4}$$

These quantities correspond to the outputs of the first hidden layer.
Then, these values are again linearly combined to give the output of the second hidden layer of the NN as:

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)} \tag{A.5}$$

where $k = 1, \ldots, K$, and $K$ is the total number of outputs.
Finally, the output unit activations are transformed using an appropriate activation function to give a set of NN outputs $y_k$ as:

$$y_k = f(a_k) \tag{A.6}$$

In particular, the choice of the activation function $f(\cdot)$ is determined according to the nature of the data and the assumed distribution of target variables.
In summary, the overall NN model takes the form:

$$y_k(\mathbf{x}, \mathbf{w}) = f(\sum_{j=1}^{M} w_{kj}^{(2)} h(\sum_{j=1}^{M} w_{ji}^{(1)} x_i + w_{j0}^{(1)}) + w_{k0}^{(2)}) \tag{A.7}$$

where the set of all weights and bias parameters and input variables have been grouped together in vector $\mathbf{w}$ and $\mathbf{x}$, respectively. Thus, NN is simply a non linear function from a set of input variables $\mathbf{x}$ to a set of output variables $\{y_k\}$ controlled by a vector $\mathbf{w}$ which represents the NN unknown parameters that need to be estimated.

## A.2 Neural network training through gradient descent: the backpropagation algorithm

Training, i.e. estimating $\mathbf{w}$, of NN is usually performed by resorting to the *backpropagation* algorithm, which consists in an iterative algorithm based on gradient descent that alternatively pass information forward and backward through the network. In the following, derivation of backpropagation is described for a general NN having arbitrary topology, arbitrary activation functions, and a broad class of error functions (i.e. functions used to quantify the error the NN makes in predicting the target $\{y_k\}$ given the current $\mathbf{w}$).
Commonly, the error function $E$ involve a sum of terms, one for each data point in the training set, so that:

$$E(\mathbf{w}) = \sum_{n=1}^{N} E_n(\mathbf{w}) \tag{A.8}$$

Here, the problem is evaluating $\nabla E_n(\mathbf{w})$ for one such term in the error function.
Considering a general NN, each unit computes a weighted sum of its inputs that is consequently non linearly transformed by an activation function to give activation $z_j$ of unit $j$ as follows:

$$z_j = h(a_j) = h(\sum_i w_{ji} z_i) \tag{A.9}$$

For each target variable of the training set, we know the corresponding input vector to NN and the respective activations of its units. Then, considering the evaluation of the derivative of $E_n$ with respect to a weight $w_{ji}$, it is possible to observe that $E_n$ depends on the weight $w_{ji}$ only by $a_j$ to unit $j$. As such, the

chain rule for partial derivatives can be applied to obtain:

$$\frac{\delta E_n}{\delta w_{ji}} = \frac{\delta E_n}{\delta a_j}\frac{\delta a_j}{\delta w_{ji}} = \delta_j z_i \tag{A.10}$$

where, $\delta_j = \frac{\delta E_n}{\delta a_j}$ and $z_i = \frac{\delta a_j}{\delta w_{ji}}$.
For the output unit, it can be derived that:

$$\delta_k = \hat{y}_k - y_k \tag{A.11}$$

where $y_k$ and $\hat{y}_k$ are the true and the predicted target value (obtained from the NN using the current $\mathbf{w}$), respectively. Then, using again the chain rule for partial derivatives on $\delta_j$ it is possible to obtain:

$$\delta_j = \sum_k \frac{\delta E_n}{\delta a_k}\frac{\delta a_k}{\delta a_j} \tag{A.12}$$

Finally, using the definitions of $\delta$, $a_j$, and $z_j$, it results the *backpropagation formula*:

$$\delta_j = h'(a_j)\sum_k w_{kj}\delta_k \tag{A.13}$$

which means that the value of $\delta$ for a particular unit can be obtained by propagating the $\delta$'s backward form units higher up in NN.

For the sake of clarity, the backpropagation algorithm has been summarized as follows:

1. Apply an input vector $\mathbf{x_n}$ to NN and find activations and outputs of all units;
2. Evaluate the $\delta_k$ for all the outputs units;
3. Backpropagate the $\delta$'s using the backpropagation formula for each uni of the NN;
4. Evaluate the derivatives of each unit.

**Algorithm 1:** The backpropagation algorithm

## A.3 Choosing NN architecture

While the design of the input and output layers of NNs is often straightforward, there can be a multitude of choices when choosing the structure of hid-

den layers. Indeed, it is not possible to sum up the design process for the hidden layers with few rules of thumb, so users developing NN have to fix the number of units per layer, the number of layers, which kind of activation function to use for each layer, and so on and so forth. These parameters are commonly referred to "hyperparameters" of the NN since they define, from the architectural point-of-view, the topology of the NN. The difficulty of tuning NN hyperparameters is basically due to the fact that, for given hyperparameter set, the computational cost associated with the NN training is very high. As such, this translates in the impossibility of searching for the optimal hyperparameter set by a simple exhaustive grid search, i.e. where the candidate optimal hyperparameters are defined upon a fine fixed grid and NN performance is evaluated for each of the possible grid points.
This problem calls for hyperparameter optimization techniques that are robust and allow to tune the NN structure with an acceptable computational cost. In the following, a state-of-the-art solution to this problem is discussed.

### A.3.1 Sequential model-based global optimization

Sequential model-based global optimization (SMBO) algorithms are a class of methodologies which have been widely used in the literature to tune models that are very computationally heavy to evaluate, e.g. NN or gradient boosted trees (described later in Appendix B) [97]. In particular, when the true fitness function $f : \mathcal{X} \longrightarrow \mathbb{R}$ is expensive to evaluate, SMBO algorithms approximate $f$ with another function $g$ cheaper to evaluate. Therefore, at each iteration, the point $x^*$ that maximizes $g$ becomes the proposal for where the true function $f$ should be evaluated.
SMBO algorithms available in the literature, have different criterion they use to obtain, at each iteration, $x^*$ given $g$, and different ways to model $f$ via observation history $\mathcal{H}$, i.e. the set of $(x^*, f(x^*))$ that have been evaluated upon the current iteration of the SMBO algorithm.
Here, we formally describe the Tree-structured Parzen Estimator approach (TPE), that is the SMBO algorithm we adopt in this thesis to optimize the hyperparameters of the developed machine learning models.
TPE belongs to a family of SMBO algorithms that optimize the criterion of Expected Improvement (EI) [98], i.e. the expectation under some model $M$ of $f$

that $f(x)$ will exceed (negatively) some threshold $y^*$ defined as:

$$EI_{y^*}(x) := \int_{-\infty}^{+\infty} \max(y^* - y, 0) p(y|x) dy \tag{A.14}$$

In TPE, $p(y|x)$ is not modeled directly, instead it models $p(x|y)$ and $p(y)$.
To do so, by assuming that the space in which we search for the optimal hyperparameter set $\mathcal{X}$ can be described by a graph-structured generative process, e.g. first choose the number of hidden layers of a neural network, then choose each layer's activation function, TPE models $p(x|y)$ by transforming that generative process, replacing the distributions of the configuration prior with non-parametric densities, i.e. uniform, log-uniform, quantized log-uniform, and categorical variables. Then, TPE substitute uniform densities with truncated Gaussian mixture densities, log-uniform densities with exponential truncated Gaussian mixture densities, and categorical with re-weighted categorical. In particular, using different observations $\{x_1, x_2, \ldots, x_k\}$ these substitutions represent the learning algorithm that produces a variety of densities over $\mathcal{X}$. TPE defines $p(x|y)$ using two such densities:

$$p(x|y) = \begin{cases} l(x) & \text{if} \quad y < y^* \\ g(x) & \text{if} \quad y \geq y^* \end{cases} \tag{A.15}$$

where $l(x)$ is the density formed by using the observations $\{x_i\}$ such that corresponding loss $f(x_i)$ was less than $y^*$ and $g(x)$ is the density formed by using the remaining observations. The TPE algorithm chooses $y^*$ to be some quantile $\gamma$ of the observed $y$ values, so that $p(y < y^*) = \gamma$, but no specific model for $p(y)$ is necessary. Indeed, by maintaining the observed variables in $\mathcal{H}$ sorted, the runtime of each iteration of TPE can scale linearly in $|\mathcal{H}|$ and linearly in the number of variables being optimized.

**How TPE optimize EI**

Starting from the definition of EI for a given $y^*$

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y) p(y|x) dy = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy$$

by construction, it can observed that $\gamma = p(y < y^*)$ and $p(x) = \int_{\mathbb{R}} p(x|y)p(y)dy = \gamma l(x) + (1-\gamma)g(x)$. It follows:

$$\int_{-\infty}^{y^*}(y^*-y)p(x|y)p(y)dy = l(x)\int_{-\infty}^{y^*}(y^*-y)p(y)dy = \gamma y^* l(x) + (1-\gamma)l(x)\int_{-\infty}^{y^*}p(y)dy$$

so that finally:

$$EI_{y^*}(x) = \frac{\gamma y^* l(x) - l(x)\int_{-\infty}^{y^*}p(y)dy}{\gamma l(x) + (1-\gamma)g(x)} \propto (\gamma + \frac{g(x)}{l(x)}(1-\gamma))^{-1}$$

This shows that, probabilistically speaking, to maximize improvement, it is necessary to have points with high probability under $l$ and low probability under $g$. From the practical point of view, given the tree form of $l$ and $g$, it is possible to easily draw candidates distributed according to $l$ and evaluate them according to $g(x)/l(x)$. As such, on each iterations, TPE, draws many candidates $x$ under $l$, evaluates their improvement, and finally returns the candidate $x^*$ with the greatest EI.

# Appendix B

# Gradient Boosted Trees (GBT)

Gradient boosted trees (GBT) [60], are powerful machine learning models that, in the last years have been proven to outperform other machine learning based techniques in many applications. Beside being accurate and efficient, GBT is popular also because it is highly interpretable. GBT is an ensemble technique that uses decision trees as base models where, at each step, a new decision tree is trained to fit the residual between ground truth and current prediction. Computationally speaking, many efficient implementations have been proposed in the literature, such as the one used in present work, i.e. eXtreme GBT (XGB) [75] that uses the second order gradient to guide the boosting process and improve the accuracy. In the following, the derivation of the objective function of GBT using the second order Taylor expansion, which is currently used by the XGB implementation, is presented, as well as the rationale behind the most used split finding methods, i.e. the methodologies decision trees implement in order to find the best topology.

## B.1 Derivation of the objective function

Let's define $\mathcal{D} = \{(\mathbf{x}_i, y_i) | i = 1, ;n\}$ as a dataset of $n$ samples $\mathbf{x}_i$, of $m$ features $\mathbf{x} \in \mathbb{R}^m$, each associated to its target $y_i$. In a decision tree, i.e. the base model of a GBT, prediction of $y_i$ are calculated as:

$$\hat{y}_i = f(\mathbf{x}_i) = w_{q(\mathbf{x})} \tag{B.1}$$

where $f : \mathbb{R}^m \longrightarrow \mathbb{R}$ is a function that maps a given $\mathbf{x}_i$ into a scalar, $q : \mathbb{R}^m \longrightarrow [1, L]$ is a function which selects a leaf of the decision tree given $\mathbf{x}$ and $w$, $L$ is the number of leaves of the decision tree, and $w_i$ is the weight of the $i$-th leaf.

I a GBT composed of $K$ decision trees, the prediction $\hat{y}_i$ is obtained as:

$$\hat{y}_i = \sum_{k=1}^{L} f_k(\mathbf{x}_i) \tag{B.2}$$

To train a GBT, i.e. learn the set of function $\{f_k | k = 1, \ldots, K\}$ associated to the $K$ decision trees, we want to minimize the following objective function:

$$\mathcal{L} = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \mathcal{R}(f_k) \tag{B.3}$$

where $l$ is a loss function which quantifies the prediction error of the model, and $\mathcal{R}$ is a regularization term used to penalize model complexity defined as:

$$\mathcal{R}(f) = \gamma L + \frac{1}{2}\lambda ||w_{q(\mathbf{x})}||^2 \tag{B.4}$$

where $\gamma$ is a parameter that weights the penalty contribution given by the number of leaves of the given decision tree, and $\lambda$ weights the penalty contribution given by the decision tree weights.
The objective function cannot be minimized using traditional methods. In fact, GBT is trained in an additive manner. To do so, let's define $\hat{y}_i^{(t)}$ as the prediction of $y_i$ at the $t$-th iteration of the training algorithm. The objective function becomes:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(\hat{y}_i^{(t)} + f_t(\mathbf{x}_i), y_i) + \sum_{k=1}^{K} \mathcal{R}(f_k) \tag{B.5}$$

hence, we add the $f_t$ function that minimizes the objective function greedily. Then, by using the second-order Taylor expansion, we can approximate $\mathcal{L}^{(t)}$ as:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} [l(\hat{y}_i^{(t)}) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \sum_{k=1}^{K} \mathcal{R}(f_k) \tag{B.6}$$

where $g_i$ and $h_i$ are the first and the second order derivatives of $l(\hat{y}_i, y_i)$ with respect to $\hat{y}_i$, respectively.
As such, by removing the constant terms, we obtain a simplified objective function at the $t$-th iteration of the training algorithm as:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} [g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \sum_{k=1}^{K} \mathcal{R}(f_k) \tag{B.7}$$

Then, by defining $Is_j = \{i|q(\mathbf{x}_j) = j\}$ as the instance set of leaf $j$, and expanding $\mathcal{R}$, we can obtain the following:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n}[g_i f_t(\mathbf{x}_i) + \frac{1}{2}h_i f_t^2(\mathbf{x}_i)] + \gamma L + \frac{1}{2}\lambda \sum_{j=1}^{L} w_j^2 \tag{B.8}$$

$$= \sum_{j=1}^{L}[(\sum_{i \in Is_j} g_i)w_j + \frac{1}{2}(\sum_{i \in Is_j} h_i + \lambda)w_j^2] + \gamma L \tag{B.9}$$

Then, zeroing $\frac{\delta \tilde{\mathcal{L}}}{\delta w}$, it results that, for a fixed $q$, we can obtain the optimal weight $w_j^*$ as:

$$w_j^* = -\frac{\sum_{i \in Is_j} g_i}{\sum_{i \in Is_j} h_i + \lambda} \tag{B.10}$$

and the respective optimal value of the simplified objective function at iteration $t$ as:

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{L}\frac{(\sum_{i \in Is_j} g_i)^2}{\sum_{i \in Is_j} h_i + \lambda} + \gamma L \tag{B.11}$$

## B.2 The split finding problem

One of the key problems of decision tree building is to find the best way to split a node of the tree given a set of samples. In practice, it means that we want to set a sort of optimal threshold $T$ such that a set of samples $\{\mathbf{x}_i \, | i = 1, \ldots, n\}$ are divided into:

$$\mathbf{x}_i \in \begin{cases} \text{left} & \text{if} \quad \mathbf{x}_{ij} \leq T \\ \text{right} & \text{if} \quad \mathbf{x}_{ij} > T \end{cases} \tag{B.12}$$

In this context, the optimal objective function after the split is reduced by:

$$\frac{1}{2}[\frac{(\sum_{i \in Is_L} g_i)^2}{\sum_{i \in Is_L} h_i + \lambda} + \frac{(\sum_{i \in Is_R} g_i)^2}{\sum_{i \in Is_R} h_i + \lambda} - \frac{(\sum_{i \in Is} g_i)^2}{\sum_{i \in Is} h_i + \lambda}] - \gamma \tag{B.13}$$

where $Is$ is the instance set associated to the node that has been split, $Is_L$ and $Is_R$ are the instance sets of the left and right nodes after the split.
As such, the split finding problem can be reformulated as the finding of the best threshold $T$ such that it maximizes Equation B.13. Several algorithms have been proposed in the literature [REF] to solve this problem. In the following, we will discuss the rationale behind two of the most popular algorithms for the scope: the exact split finding and the approximate split finding algorithms.

**Exact split finding algorithm** The exact split finding algorithm, is the most simple technique to find the best split. It consists of enumerating over all the possible splits on all features in order to solve the problem at hand. Of course, this is the most accurate split finding algorithm since it is possible to test all of the possible combinations of splits by exhaustive search and identify the better one. However, this is also computationally inefficient, especially in those cases where the number of samples is very high. To improve efficiency, practical implementations of the exact split search algorithm start sorting the samples according to their value and then they visit the samples in the sorted order to accumulate $g$ and $h$ which are used to compute Equation B.13.

**Approximate split finding algorithm** To ease the computational burden brought by the exact search algorithm, it is necessary to resort to approximate solutions. The rationale of these techniques is to divide samples into several subsets and enumerate over these subsets instead of each single samples. As a results, the complexity of the algorithm depends on the number of subsets instead of the number of samples. To do so, in the literature two methodologies are currently used, i.e. the quantile-based [99] and the histogram-based [100]. While the quantile-based strategy has the advantage of being distributable and recomputable, the histogram-based approach is significantly faster as such in the following we briefly describe the latter.
In the histogram-based approximate split finding algorithm, the sample space is divided into $b$ adjacent interval and each of the samples is associated to one of these intervals accordingly. Specifically, the $b$ intervals are built following the distribution of $j$th input column of the whole dataset. Then, keeping these intervals unchanged, a sample $\mathbf{x}_i$ is associated to the $k$th interval if $s_{k-1} < \mathbf{x}_{ij} \leq s_k$. In particular, the subset value of $\mathbf{x}_{ij}$ is obtained by binary-search to reduce complexity.