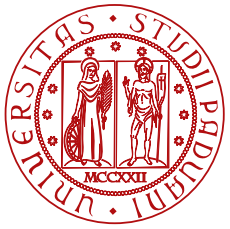

Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Department of Information Engineering

Master's Degree in Bioengineering

Model-informed methodologies to optimize insulin dosing for high-fat and high-protein meals in individuals with type 1 diabetes

Candidate

Sebastiano Bellese

Supervisor

Prof. Chiara Dalla Man

University of Padova

Co-supervisor

Dr. Edoardo Faggionato

University of Padova

Academic Year: 2024–2025

Graduation Date: 12th September 2025

Abstract

For individuals diagnosed with Type 1 Diabetes (T1D), effective insulin therapy is crucial for maintaining Blood Glucose (BG) levels within the target range, thus reducing the risk of both acute and chronic complications associated with diabetes. Since their pancreas no longer produces insulin, they rely on external insulin intake to regulate BG. Daily insulin administration is usually done via Multiple Daily Injections (MDI), using insulin pens, or Continuous Subcutaneous Insulin Infusion (CSII) using Insulin Pumps (IP). The treatment is complemented with a BG monitoring technology, such as glucometers to sample capillary glucose and Continuous Glucose Monitoring (CGM) devices to measure subcutaneous glucose.

Despite the great technological progress, most of the patients affected by diabetes still fail to achieve their glycemic goals, mainly due to poor postprandial BG control. One of the reasons is that available insulin dosing strategies are based solely on carbohydrate counting to determine prandial insulin requirements. This approach is suboptimal, since it does not account for the influence of fat and protein on glycemic excursion. In fact, these two macronutrients are known to delay glucose absorption and increase insulin resistance, and thus, their presence can lead to early hypoglycemia and/or delayed and sustained postprandial hyperglycemia if not considered in the treatment. To date, some studies have been conducted to propose insulin adjustments based on meal composition. They demonstrated that the Dual-wave Bolus (DWB) insulin strategy, combining a square-wave and a standard bolus using CSII, proved to be a good strategy to address complex meals. However, the sample sizes were relatively small and the studies were heterogeneous, making it difficult to define clear guidelines for the DWB strategy.

In this optic, a methodology to infer meal macronutrient composition or to perform meal classification exploiting minimally invasive data (such as CGM and IP infusion profiles) acquired in real-life conditions would be useful to develop individualized and adaptive DWB strategies. This thesis addresses this gap through an explorative analysis based on model-driven machine learning approaches. Specifically, we exploited features derived from parameters of the Minimally-invasive Oral Minimal Model (MI-OMM), a physiologically-based

model describing glucose and insulin interactions from CGM and IP data in individuals with T1D. The model was identified in a dataset comprising individuals with T1D, where macronutrient content of each meal was recorded alongside minimally invasive data under free living conditions. Both supervised and unsupervised learning techniques were evaluated. Supervised learning showed limited performance caused by the variability of meal effects on individual postprandial BG excursion, making the macronutrient content label unreliable. On the contrary, unsupervised methods, especially *k*-means, that circumvent these issues, effectively separated meals into two groups: one characterized by higher macronutrient content (fat, protein, carbohydrates) and the other by lower content. The proposed model-based methodology was further validated *in silico* using the UVa/Padova T1D Simulator. Despite its strong limitations, simulations confirmed that the DWB strategy improved postprandial glycemic control in complex meals rich in protein and fat.

Overall, this model-based classification framework added predictive value in identifying meals with special insulin requirements. This methodology has the potential to drive the change in insulin therapies that go beyond carbohydrate counting, highlighting the importance of considering fat and protein when optimizing insulin dosing.

Sommario

Per gli individui con diagnosi di Diabete di Tipo 1 (T1D), una terapia insulinica efficace è cruciale per mantenere i livelli di glicemia (BG) entro il range target e ridurre il rischio di complicanze acute e croniche del diabete. Poiché il loro pancreas non produce insulina, dipendono dall'assunzione esogena per regolare la BG. La somministrazione spesso avviene tramite Iniezioni Multiple Giornaliere (MDI), utilizzando penne insuliniche, oppure attraverso Infusione Sottocutanea Continua di Insulina (CSII) con microinfusori (IP). Il trattamento è integrato da tecnologie di monitoraggio della BG, come glucometri per il campionamento della glicemia capillare e i sistemi di Monitoraggio Continuo del Glucosio (CGM) per la misura del glucosio sottocutaneo.

Nonostante i grandi progressi tecnologici, gran parte dei pazienti non raggiunge gli obiettivi glicemici, principalmente a causa del difficile controllo della glicemia postprandiale. Una delle ragioni è che le strategie disponibili per il calcolo della dose insulinica si basano solo sul conteggio dei carboidrati per stabilire il fabbisogno prandiale, trascurando l'influenza di grassi e proteine sull'escursione glicemica. Infatti, questi sono noti per ritardare l'assorbimento del glucosio e aumentare la resistenza insulinica e, quindi, possono causare ipoglicemia precoce e/o iperglicemia postprandiale tardiva e prolungata, se non considerati. Alcuni studi sono stati condotti per proporre aggiustamenti insulinici che si basano sulla composizione del pasto. Hanno dimostrato che la strategia insulinica del Bolo a Onda Doppia (DWB), che combina un bolo standard e un bolo a onda quadra mediante CSII, si è rivelata una buona strategia per trattare pasti complessi. Tuttavia, le ridotte popolazioni studiate e l'eterogeneità degli studi rendono difficile definire linee guida chiare per il DWB.

In quest'ottica, una metodologia per stimare la composizione in macronutrienti dei pasti o di classificarli usando dati minimamente invasivi (come CGM e IP) raccolti in condizioni di vita quotidiana potrebbe rivelarsi utile per sviluppare strategie DWB individualizzate e adattive. Questa tesi affronta tale lacuna con un'analisi esplorativa basata su approcci di machine learning basati su modelli. In particolare, sono state sfruttate feature derivate da parametri del Minimally-invasive Oral Minimal Model (MI-OMM), un modello basato sulla fisiologia che descrive interazioni del glucosio e dell'insulina da dati CGM e IP in soggetti con T1D. Il modello è stato identificato in un dataset con individui con T1D, in

cui il contenuto in macronutrienti di ogni pasto è stato registrato insieme a dati minimamente invasivi raccolti in condizioni di vita quotidiana. Sono state valutate sia tecniche di apprendimento supervisionato che non. L'apprendimento supervisionato ha mostrato prestazioni limitate a causa della variabilità degli effetti dei pasti sull'escursione postprandiale individuale, rendendo inaffidabile l'etichettatura sul contenuto dei macronutrienti. D'altra parte, i metodi non supervisionati, in particolare il *k*-means, che aggirano tali problemi, hanno permesso di separare i pasti in due gruppi: uno con un più alto contenuto di macronutrienti (grassi, proteine, carboidrati) e l'altro con un contenuto più basso. La metodologia proposta, basata sul modello, è stata validata *in silico* utilizzando il simulatore UVa/Padova. Nonostante le forti limitazioni, le simulazioni confermano che il DWB migliora il controllo glicemico postprandiale per pasti complessi ricchi di proteine e grassi.

Nel complesso, questo framework di classificazione supportato da un modello ha aggiunto un valore predittivo nell'identificazione di pasti con particolari esigenze insuliniche. Tale metodologia può spingere il cambiamento delle terapie insuliniche andando oltre il conteggio dei carboidrati, sottolineando l'importanza di considerare grassi e proteine nella strategia insulinica.

Contents

List of Figures

List of Tables

List of Acronyms

AIC Akaike Information Criterion

AP Artificial Pancreas

AUC Area Under the Curve

BG Blood Glucose

BMI Body Mass Index

BW Body Weight

COV Coefficient of Variation

CV Cross Validation

CGM Continous Glucose Monitoring

CR Insulin-to-carbohydrate Ratio

CSII Continuous Subcutaneous Insulin Infusion

DSS Decision Support Systems

DWB Dual-wave Bolus

FPR False Positive Rate

GMM Gaussian Mixture Model

GR Gastric Retention

HC High-carbohydrate

HE Hypoglycemic Events

HF High-fat

HP High-protein

IP Insulin Pump

IQR Interquartile Ranges

LC Low-carbohydrate

LF Low-fat

LL Log-likelihood

LME Linear Mixed Effects

LP Low-protein

MAP Maximum a Posteriori

MI-OMM Minimally-invasive Oral Minimal Model

MDI Multiple Daily Injections

MPC Model Predictive Control

OOP Optimal Operating Point

PID Proportional-integral-derivate

Ra Glucose Rate of Appearance

ROC Receiver Operating Characteristic

SAP Sensor-augmented Pump

SMBG Self-monitoring of Blood Glucose

SSE Sum of Squares Error

T1D Type 1 Diabetes

T2D Type 2 Diabetes

TAR Time Above Range

TBR Time Below Range

TIR Time in Range

TPR True Positive Rate

1 Introduction

In this chapter, the fundamental concepts and background information necessary to understand this thesis are presented. The chapter introduces diabetes, with a particular focus on Type 1 Diabetes (T1D), insulin therapies, and Blood Glucose (BG) monitoring technologies. It also highlights the main challenges in the complex treatment of T1D. Finally, the goal and objectives of this thesis, aimed at helping to address some of these challenges, are stated.

1.1 Diabetes Mellitus and Type 1 Diabetes

Diabetes Mellitus is a group of metabolic disorders characterized by defects in insulin secretion, insulin action, or both [1]. The majority of diabetes cases fall into two main categories. The first is T1D, which accounts for 5–10% of all cases. It is caused by complete insulin deficiency due to the body's own immune system, which destroys the pancreatic β-cells. The second, more prevalent form is Type 2 Diabetes (T2D), representing 90–95% of all cases. In T2D, insulin secretion or insulin action, or both, are impaired, leading to an inadequate compensatory response, often in the context of insulin resistance [2].

This thesis focuses on T1D, which results from the complete lack of endogenous insulin secretion and therefore requires lifelong exogenous insulin replacement.

1.1.1 Overview and Management

T1D is caused by immune-mediated destruction of the pancreatic β-cells that produce insulin, leading to absolute insulin deficiency. The aim of treatment is to replace endogenous insulin with frequent administration of analogs of this hormone to maintain glucose levels in the physiological ranges [3]. In this way, patients can reduce T1D complications and live longer and healthier.

Management strategies for T1D include: (i) maintaining BG levels within the target range (70-180 mg/dL) to prevent complications, and minimizing episodes of hypoglycemia (<70 mg/dl) and hyperglycemia (>180 mg/dl); (ii) managing cardiovascular risk factors; (iii) providing approaches, treatments, and devices that reduce the burden on individuals with T1D and their caregivers, thereby mitigating diabetes-related stress and improving psychological well-being [4].

Maintaining BG levels within this tight range is particularly challenging because they are influenced by external factors such as meals, physical activity [5], and both physiological and psychological stress [6]. Nonetheless, achieving stable glucose control is crucial, as uncontrolled levels can cause both short- and long-term adverse effects [7].

Short-term effects of hypoglycemia and hyperglycemia include headaches, sweating, rapid heartbeat, confusion, visual disturbances, nausea, fatigue, and weakness [8]. Prolonged hyperglycemia damages the vascular system, increasing the risk of cardiovascular disease, retinopathy, neuropathy, and nephropathy [9]. If left untreated, dangerously high BG levels may ultimately result in diabetic coma. It is important to note that hypoglycemia is associated with more dangerous short-term symptoms compared to hyperglycemia, since it causes a disruption in the supply of glucose to the brain and neurons. However, the long-term complications of hyperglycemia are equally severe.

1.1.2 Insulin Therapies

Effective T1D management requires appropriate exogenous insulin replacement. Insulin therapy is therefore a cornerstone of treatment, as it represents one of the most critical aspects of T1D care. Individuals with T1D should follow insulin regimens that mimic physiological insulin secretion as closely as possible in order to reduce the risk of adverse events. Administering more insulin than required (insulin overdosing) can cause prolonged hypoglycemic

events. Conversely, insufficient insulin dosing may lead to hyperglycemia. It is therefore evident that selecting the most suitable insulin therapy is essential for optimizing BG levels control.

Currently, the two most employed insulin therapies are either the Multiple Daily Injections (MDI) of insulin or the Continuous Subcutaneous Insulin Infusion (CSII), with the latter allowing for more precise and flexible insulin dosing [10].

The mode of insulin delivery is different for the two systems, illustrated in Fig. 1.1. In case of MDI, the individuals with T1D must undergo subcutaneous insulin injection via insulin pens (Fig. 1.1, Panel A). An MDI insulin therapy is composed of the following: (i) long-acting insulin analogs injected once a day to cover basal insulin requirements; (ii) administration of pre-meal boluses of fast-acting insulin analogues to mitigate the postprandial variation of BG levels; (iii) corrective boluses to treat underdosing or other hyperglycemic events.

Figure 1.1: The two most widespread modes for insulin injection. Panel A: Insulin pen (NovoPen 6, Novo Nordisk A/S, Bagsvaerd, Denmark. Source: [11]). Panel B: Insulin pump (Omnipod DASH Insulin Management System, Insulet Corp., Acton, MA, USA. Source: [12]).

On the other hand, with CSII therapy, instead of insulin pens, the patient is provided with an Insulin Pump (IP) (Fig. 1.1, Panel B), where, instead of a daily injection of a long-acting insulin analog, a continuous basal insulin infusion composed of micro boluses of rapid-acting insulin analogs is used to cover the individual insulin basal requirements. Corrective and meal boluses are still present [4].

Other insulin delivery methods also exist, such as oral and inhaled insulin

analogs. However, their clinical use remains limited due to higher costs compared to MDI and CSII, potential contraindications, and uncertain efficacy [13].

1.1.3 Blood Glucose Monitoring

Alongside effective insulin therapy, the successful treatment of T1D also relies on the ability to reliably measure and collect BG samples in order to monitor glucose levels. Employing appropriate monitoring techniques reduces the risk of administering an incorrect insulin dose, thereby mitigating the occurrence of hypo- and hyperglycemic events. The pursuit of adequate glycemic control has therefore driven the development of BG monitoring technologies. In this thesis, we will focus on widespread minimally-invasive BG monitoring technologies. Of note, other popular BG sampling methods are available, but they require invasive plasma sampling and are only employed in hospitalized settings.

Currently, two main minimally-invasive and portable technologies are available, enabling glucose monitoring in free-living conditions: Self-monitoring of Blood Glucose (SMBG) and Continous Glucose Monitoring (CGM). These two BG monitoring technologies are illustrated in Fig. 1.2.

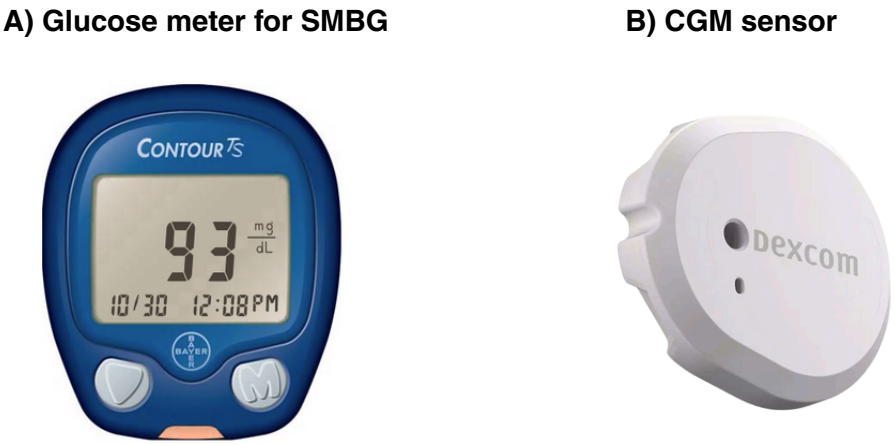

Figure 1.2: The two most used minimally-invasive BG monitoring devices. Panel A: Glucose meter for SMBG (Contour TS, Ascensia Diabetes Care AG (formerly Bayer), Basel, Switzerland. Source: [14]). Panel B: CGM sensor (Dexcom G7, Dexcom Inc., San Diego, CA, USA. Source: [15]).

The first BG monitoring technology, SMBG (Fig 1.2, Panel A), is the most widely used and accessible method. It is based on a glucometer, where BG is

measured by placing a small drop of blood, typically from a fingertip, onto a test strip that is then inserted into the device, which promptly displays its concentration. Despite its simplicity and broad use, SMBG presents several limitations. Measurement accuracy depends heavily on the calibration and reliability of the glucometer. Concerns have also been raised about possible contamination and the transmission of blood-borne pathogens. Moreover, its effectiveness is strongly dependent on patient adherence and compliance. Clinical guidelines recommend that individuals with T1D use SMBG before meals, before engaging in critical activities, and whenever symptoms of hypo- or hyperglycemia are suspected [16].

The second and more advanced technology is CGM (Fig. 1.2, Panel B). CGM systems were developed to address some of the limitations of SMBG. A CGM device consists of a subcutaneous sensor, typically inserted in the abdomen or on the back of the upper arm, that can remain in place for several days to weeks, depending on the system [17]. The sensor, connected via a small catheter, measures glucose levels in the interstitial fluid at frequent intervals, usually every 5 minutes, thereby providing a frequently sampled profile of glycemic excursions. CGM sensors convert interstitial glucose concentrations into electrical signals, which are then displayed on a monitor or transmitted to a smartphone. By providing real-time glucose data, CGM systems enable individuals with T1D to better manage their insulin therapy and maintain blood glucose levels within the target range, thereby improving overall glycemic control.

Despite its advantages, CGM also has limitations. First, SMBG measurements remain necessary, since fingertip samples are still required for calibration of most systems, and before taking any major intervention, to confirm the measured concentration of the CGM sensor, especially in case of suspected hypo- or hyperglycemia [18]. Second, CGM sensors, other than calibration issues, are also affected by drift and offset. Third, interstitial glucose measurements are subject to a physiological lag time relative to BG, which can lead to inaccuracies, particularly during rapid glucose fluctuations. Other limitations are present, such as the accuracy degradation of the sensor within the first 24 hours after insertion due to local tissue inflammation [16].

1.1.4 Advancements in Treatment

The combined use of CGM sensor and IP gave rise to the so-called Sensor-augmented Pump (SAP) therapy. Specifically, SAP combines CSII delivered through an IP with a CGM sensor in a single integrated system. The IP delivers fast-acting insulin through a subcutaneous catheter, programmed with basal rates and bolus doses, while the CGM sensor measures interstitial glucose levels in real time. The peculiarity of SAP is that data from the CGM sensor are transmitted to the IP, which displays glucose trends and alerts the user to take actions to avoid hyper- or hypoglycemia. This integration enables more responsive and individualized insulin delivery compared with MDI or the stand-alone use of IP, as users can make informed adjustments based on continuously available glucose data. The SAP system has proved to achieve improved glycemic control [19].

The new frontier of insulin therapy consists of closing the loop between the CGM sensor and the IP, thereby minimizing human involvement in treatment. This almost fully automated approach is commonly referred to as a hybrid closed-loop system or hybrid artificial pancreas Artificial Pancreas (AP) [20].

The primary goal of an AP system is to increase the time spent within the glucose target range while reducing human interventions. An AP system integrates three core components: an IP that serves as an actuator, a CGM sensor, and a control algorithm (the actual AP). The algorithm dynamically adjusts insulin delivery in real time based on CGM values and glucose trends (e.g., insulin infusion is increased or decreased in response to rising or falling glucose levels) [10].

The most employed control algorithm is the Model Predictive Control (MPC). MPC uses a mathematical model of glucose–insulin dynamics to predict future glucose trajectories over a prediction horizon. Based on these forecasts, the MPC optimizes insulin infusion rates by minimizing deviations from a target glucose range while accounting for constraints such as insulin action delays or glycemic safety ranges. Unlike classic Proportional-integral-derivate (PID) control, which reacts primarily to current and past glucose values [21], MPC anticipates future changes and adjusts insulin delivery proactively. AP systems with an MPC algorithm have been shown to significantly improve glucose control while increasing time maintained in the target range [22].

While these hybrid AP systems automate several aspects of insulin therapy,

user involvement remains necessary to achieve satisfactory treatment outcomes. Different AP devices differ in the degree of automation and user input required. In some systems, individuals must manually enter carbohydrate estimates for meals. This open-loop delivery of insulin analog, which requires human intervention, is also referred to as meal announcement. In others, the meal announcement process is simplified for the user by categorizing meals (e.g., small, medium, or large breakfast, lunch, or dinner) and adjusting the insulin therapy accordingly [20]. Nevertheless, carbohydrate counting or meal categorization remains a burden and error-prone task for most patients and affects their quality of life. Therefore, a fully automated closed-loop system that eliminates the need for meal announcement is highly desirable for T1D. Systems of this kind have been tested; however, without completely reaching the performance of the hybrid systems [23].

Thanks to the advent and large availability of CGM and IP technologies, diabetes care now benefits from a continuous and high-volume stream of data. This has paved the way for the development of Decision Support Systems (DSS) in T1D management. By leveraging large datasets, DSS can be trained using machine learning techniques to recognize complex patterns and improve insulin therapy decisions. These systems have been shown to assist in meal detection, reducing the need for manual meal announcements, as well as in the recognition of exercise events, faults in insulin delivery, and in providing more accurate insulin bolus calculators and carbohydrate estimation tools [24].

1.2 Meal Composition and Postprandial Glycemic Excursions

Despite the increasing adoption of advanced technologies, managing postprandial BG excursions, which significantly contribute to glucose variability, remains a major challenge for individuals with T1D [25]. Several factors contribute to this difficulty, including suboptimal therapy parameters due to physiological variability, imprecise carbohydrate counting, and the influence of meals with High-fat (HF) and/or High-protein (HP) content.

Among macronutrients, carbohydrates have the greatest impact on BG levels. However, recent research has highlighted the significant role of fat and protein in shaping postprandial glycemic responses [26]. These findings underscore the

importance of considering meal composition, including the relative contribution of different macronutrients such as the share of fat, protein, and carbohydrates, when addressing postprandial BG control.

1.2.1 Meal Composition Effects on Postprandial Blood Glucose

Carbohydrates are considered the main contributor in a meal in influencing the postprandial BG, and therefore most insulin strategies rely solely on carbohydrate counting [7]. The influence of other macronutrients, i.e., fat and protein, is known to affect postprandial BG excursion, especially on Glucose Rate of Appearance (Ra), Gastric Retention (GR), and Insulin Sensitivity (S_I) [27].

Fat influences postprandial BG by reducing the early glucose response within the first 2–3 hours after a meal and by delaying the time to glucose peak. This effect is largely attributed to the slowing of gastric emptying, which delays glucose absorption. HF meals can therefore result in late postprandial hyperglycemia and/or an increased risk of early postprandial hypoglycemia when treated with the standard carbohydrate-based insulin dosing.

Protein produces similar effects, with notable differences in glycemic levels reported. High-protein (HP) meals, like HF meals, delay the glucose response, although their influence typically appears within the first 2 hours after meal ingestion [28].

1.2.2 Current Insulin Therapy corrections

There is a clear need to adapt insulin therapy to account for HF/HP meals. Insulin dosing strategies based solely on carbohydrate counting are suboptimal in these cases, as they fail to capture the effects of fat and protein on postprandial glycemia. Current closed-loop insulin delivery systems also do not adequately consider these macronutrient effects.

To date, few studies have been conducted to propose empirical [29, 30] or model-based [31] insulin adjustments based on meal composition, but they were performed under strictly controlled hospitalized settings. In those studies, they demonstrated that the Dual-wave Bolus (DWB), combining a square-wave and a standard bolus using CSII, proved to be effective in addressing the needs of

complex meals.

In the DWB strategy, the standard bolus targets the rapid glycemic rise caused by carbohydrates, while the square-wave bolus addresses the delayed rise due to fat and protein effect on gastric-emptying, thereby reducing both early hypoglycemia and late hyperglycemia risks. Clinical evidence suggests that the DWB strategy improves postprandial glycemic control after HF/HP meals compared with standard bolus delivery in individuals with T1D on IP therapy [32]. However, in these studies, the sample sizes were relatively small, the tested meals were predefined, and the studies were heterogeneous within each other, making it difficult to define clear guidelines for the DWB strategy.

1.3 Mechanistic Modeling of Postprandial Blood Glucose

As current AP systems do not rely on macronutrient share of a meal, but only the carbohydrate amount, there is a need for control algorithms that exploit information from the postprandial glycemic excursion to estimate the type of meal and accordingly suggest an improved insulin strategy, such as the DWB. The estimation of the macronutrient share in the meal could be possible through a physiological model quantitatively describing the postprandial BG dynamics operating only with minimally-invasive data. Among the physiologically based mechanistic models of the postprandial BG excursion, two tools employed in this thesis are the Minimally-invasive Oral Minimal Model (MI-OMM) and the UVa/Padova T1D simulator, which will be briefly introduced in this section.

1.3.1 The Minimally-invasive Oral Glucose Minimal Model

The MI-OMM is a physiologically-based compartmental model that has been shown to accurately estimate parameters related to glucose absorption and to reproduce the postprandial BG profile. The model requires only minimally invasive data, which includes CGM data, SMBG measurements, and IP infusion profiles, in order to infer physiologically meaningful parameters linked to glucose absorption and insulin kinetics. As illustrated in Fig. 1.3, the MI-OMM is composed of several interconnected sub-models, each one describing a particular biological process happening in an individual with T1D using CGM in

combination with an IP. The complete description of the MI-OMM is reported in [33].

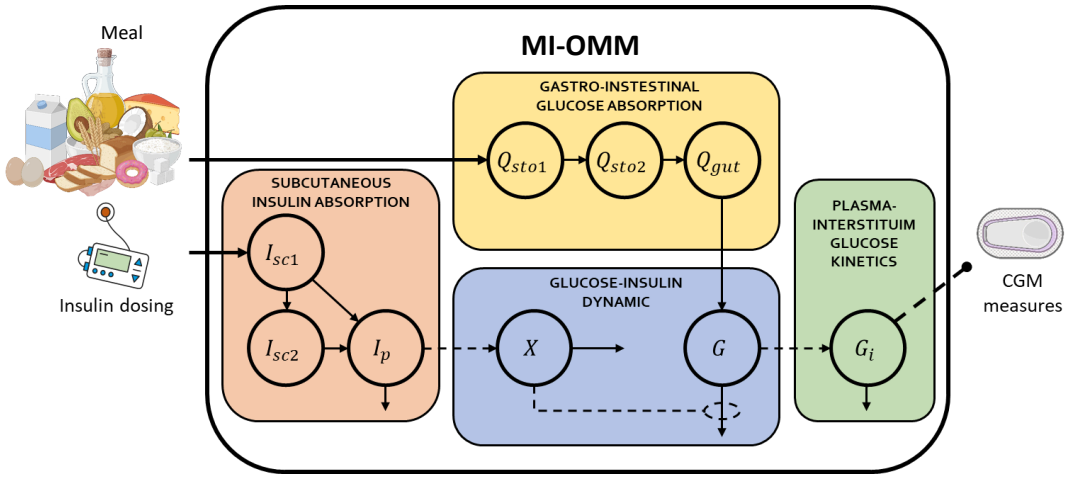

Figure 1.3: Schematic representation of the MI-OMM. The model consists of four blocks, each representing a sub-model. The gastro-intestinal glucose absorption model describes the postprandial glucose rate of appearance (yellow block). The subcutaneous insulin absorption model describes the plasma insulin concentration (orange block). The oral glucose minimal model describes the glucose-insulin interaction (blue block). The plasma-to-interstitium glucose kinetics model describes the diffusion process from blood to interstitium (green block).

The strength of employing a physiologically-based model, such as the MI-OMM, in the task of integrating meal composition effects into T1D insulin therapy lies in its ability to estimate meaningful physiological parameters that can help distinguish meals with varying macronutrient compositions. By quantitatively capturing the differential impact of carbohydrates, fat, and protein on postprandial glucose dynamics, the MI-OMM can support the task of tailoring insulin delivery strategies that go beyond carbohydrate counting alone.

1.3.2 The UVa/Padova T1D Simulator

The UVa/Padova T1D simulator is an FDA-approved simulation platform designed for the *in silico* testing and evaluation of insulin treatment strategies on virtually generated subjects with T1D [34]. The simulator is based on a physiologically based maximal model that integrates all the key processes contributing

to glucose homeostasis, along with necessary parameters to simulate individual BG dynamics over time. A schematic overview of the simulator is provided in Fig. 1.4. This tool enables the investigation of diverse scenarios, including the simulation of postprandial glycemic responses following meal intake, as well as the assessment of various insulin therapy strategies, such as the DWB.

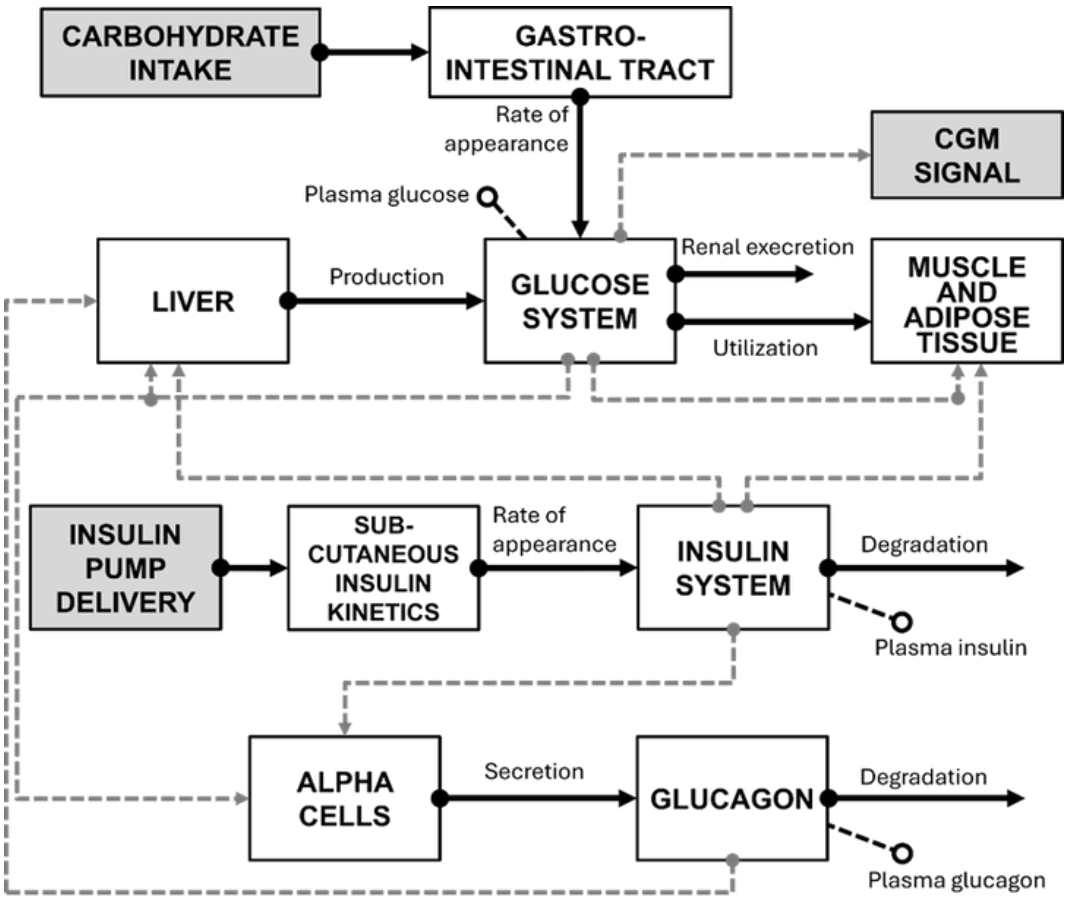

Figure 1.4: The model scheme of the UVa/Padova T1D simulator. White blocks represent sub-modules of the simulator describing the biological processes of gastro-intestinal glucose absorption, subcutaneous insulin absorption, insulin kinetics, glucose-insulin interaction, hepatic insulin extraction and glucose production, muscle and adipose tissue glucose utilization, glucagon secretion, and glucagon kinetics. Gray blocks represent inputs that are insulin pump delivery and carbohydrate intake, and output that is the CGM signal.

The capability of the UVa/Padova T1D simulator to generate virtual subjects and reproduce meal intake, particularly the postprandial BG response, makes it a promising pre-clinical tool. In particular, it may enable the simulation of BG

dynamics after complex meals, such as HF/HP meals, and provide a framework for the *in silico* evaluation of improved insulin strategies aimed at mitigating their impact on postprandial glycemia and assessing treatment efficacy.

1.4 Scope of the Thesis

As it was abovementioned highlighted, individuals with T1D would benefit from information regarding the macronutrient composition of their meals, particularly to better manage the increased insulin requirements associated with HF/HP meals. This calls attention to the need for a robust methodology capable of recognizing such meal types under real-life conditions.

This thesis aimed to address this gap by investigating supervised and unsupervised machine learning methods for inferring meal macronutrient composition or, more broadly, meal classes. This was done by employing features with predictive potential for this task coming both from the postprandial CGM trends and the identification of a physiologically-based model of the glucose-insulin system. For the latter, we selected the MI-OMM, described in the previous Section, which relies solely on minimally invasive data such as CGM profiles, SMBG values (when available), and IP data acquired under free-living conditions.

Furthermore, the thesis explored the usability of the developed methodology to identify complex meals by examining the efficacy of the DWB in managing meals classified as such. The specific objective is to assess whether this approach can improve postprandial glycemic control, reducing the risks of early hypo- and late hyperglycemia when encountering HF/HP meals.

Finally, the thesis aimed to demonstrate the added predictive value of the model-based approach carried out in this work, which leveraged features derived from parameters identified through a mechanistic model.

1.5 Contents of the Thesis

This thesis is organized as follows:

Chapter 2 describes the datasets employed in this work. Specifically, Dataset 1, extracted from [35], is used for the development of the meal classification methodology, while Dataset 2, derived from the UVa/Padova T1D Simulator [34] and composed of four sub-datasets extracted from [36], is employed for

validation.

Chapter 3 details the procedures carried out for the development of the methods for inferring meal macronutrient composition. First, the extraction of relevant features is presented, followed by the description of the preliminary analysis conducted regarding the relationship between these features and macronutrients. Subsequently, both supervised and unsupervised learning approaches are outlined in detail. The chapter concludes with the validation procedure using Dataset 2 and an assessment of the added predictive value of the proposed model-based approach.

Chapter 4 reports the results obtained during the development and validation stages, as well as the evaluation of the predictive performance of the model-based methodology against its non-model-based counterpart.

Chapter 5 provides a critical interpretation of the results, compares them with existing literature, and highlights both the encouraging findings and the limitations of this work.

Finally, Chapter 6 summarizes the main contributions of the thesis, outlines its limitations, and discusses future research directions to address these challenges. In particular, a detailed outline of future work based on the application of the developed methodology is provided.

2

Datasets

Two different datasets are employed in this thesis work. The first dataset (Dataset 1) was employed for the development and testing of the methodologies proposed for meal classification. The second dataset (Dataset 2) was employed to assess the usability of the developed methodologies.

The first dataset was acquired in a multicenter supervised clinical trial involving a population of adults, adolescents, and children with T1D to assess the safety and performance of a MPC algorithm in a real-life scenario [35]. During the study, participants were monitored with a CGM sensor, and an IP device was used for insulin administration. The peculiarity of this dataset is that it presents data from a numerous and heterogeneous population of subjects with T1D, and, for each meal, the complete macronutrient content (i.e., amount of carbohydrates, fat, and protein) was recorded with the assistance of healthcare personnel. This makes the dataset an optimal support for the scope of the thesis, where we aim to infer meal composition from minimally invasive wearable devices.

The second dataset (Dataset 2) consists of data generated in silico using the UVa/Padova T1D simulator platform [34], properly modified to account for the presence of fat and protein in the meal. Specifically, 100 virtual subjects ingested two different meals with the same carbohydrate amount but different fat and protein content. Leveraging the simulator platform not only allowed us to test the developed algorithm for meal classification on new data, but, most importantly, to validate the methodology by applying a hypothetical therapy correction to the meals labelled as HF/HP.

2.1 Dataset 1

2.1.1 Data Acquisition Protocol

The dataset was collected in a multicenter outpatient supervised clinical trial (registration identifier: NCT03216460). During the trial, 120 participants with T1D (age=15.5±11.5 years, Body Weight (BW)=51.3±28.0 kg, Body Mass Index (BMI)=21.7±5.8 kg/m^2) were enrolled from three different centers: (i) Stanford School of Medicine, Stanford, Palo Alto, CA, USA (56 participants); (ii) Barbara Davis Center of Childhood Diabetes, University of Colorado Denver, Aurora, CO, USA (34 subjects); (iii) Yale University School of Medicine, New Haven, CT, USA (30 subjects).

Participants were monitored for 3 to 5 consecutive days, in a supervised hotel or rental home setting under free-living conditions. Participants were under a SAP therapy, wearing a CGM sensor (Dexcom G4 505 Share Artificial Pancreas System, Dexcom Inc., San Diego, CA, USA) and an IP (Ominpod Insulin Management System, Insulet Corp., Acton, MA, USA) with a personalized MPC algorithm. The CGM sensor measured glucose every 5 minutes, and with the same frequency, the IP device administered a micro-bolus of insulin. SMBG measures were also acquired as per standard guidelines to confirm hypo- and hyperglycemic episodes.

Participants were studied under free-living conditions, as they were free to engage in any activity, such as physical exercise, and to choose meals of their choice. Meals and snacks were recorded with the assistance of medical personnel. Each meal log has precise information about timing, type (breakfast, lunch, or dinner), and macronutrient composition (i.e., carbohydrate, fat, and protein content). Prandial insulin boluses were calculated based on carbohydrate counting rules by each participant or by the assisting clinical personnel. Hypoglycemic events were treated with the consumption of fast-acting carbohydrates until normal glucose levels were restored. Physical activity of the participants was encouraged during the study, registering the timing and intensity level of the performed activity. For more information about the enrolled participants and the study protocol, we refer to the original publication [35].

2.1.2 Data Extraction

From the complete dataset, postprandial CGM and IP profiles were extracted from 1 hour before to 4 hours after each meal log. This time window was selected to get enough data for model identification without discarding too many meals from the analysis. Meals were merged if they were recorded within 30 minutes from each other. In total, 1211 postprandial profiles were extracted from 110 subjects. Data coming from 10 subjects were discarded due to the lack of at least one of the following: demographics information, IP data, or meal information. Among the 1211 profiles, a subset was selected based on the following requirements:

i $\geq$ 85% of CGM sensor samples available in the selected time window;

ii Meal information was complete and carbohydrate amount was > 0 g;

iii No meals were registered within 5 hours before and 4 hours after the ongoing meal;

iv Neither moderate nor intense physical exercise was performed within 2 hours before and 4 hours after the ongoing meal, and no mild physical exercise was performed by the subject within 1 hour before and 4 hours after the ongoing meal;

v The CGM profile was physiologically plausible (e.g., correct meal timing, no evident unreported bolus, no evident unreported calibration is present).

According to these requirements, a total of 353 meals coming from 105 subjects were extracted and used for the analysis. Median and interquartile range of CGM measures and administered insulin in the extracted meals are shown in Panels A and B of Fig. 2.1, respectively. Median and interquartile range of macronutrient content of the extracted meals are shown in Fig. 2.2. The age and body weight distribution of the 105 subjects are shown in Panels A and B of Fig. 2.3, respectively.

2.2 Dataset 2

The dataset was generated using the 100 in silico subjects available in the UVa/Padova T1D simulator (age=33.8±9.6 years, BW=75.5±12.1 kg, information about BMI was not available) [34]. For our purposes, we used a recently

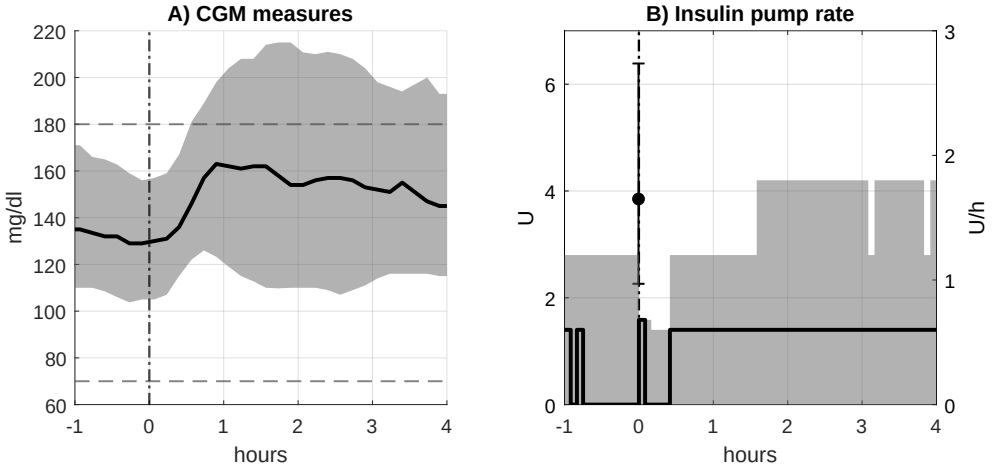

Figure 2.1: CGM measures, IP rates, and insulin boluses in Dataset 1. Vertical dashed-dotted line represents mealtime. Panel A: Median (solid black line) and 25th-75th percentiles (gray colored area) of CGM sensor measures. Gray horizontal dashed lines represent hypo- and hyperglycemic thresholds. Panel B, right y-axis: median (solid black line) and 25th-75th percentiles (gray colored area) of IP rates. Panel B, left y-axis: Median (black dot) and 25th-75th percentiles (vertical black bars) of prandial insulin boluses.

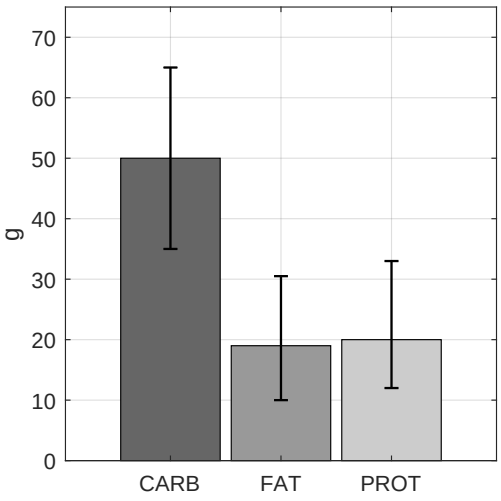

Figure 2.2: Median (gray colored bars) and 25th-75th percentiles (vertical black bars) of meal macronutrient composition in Dataset 1. CARB = carbohydrates. FAT = fat. PROT = protein.

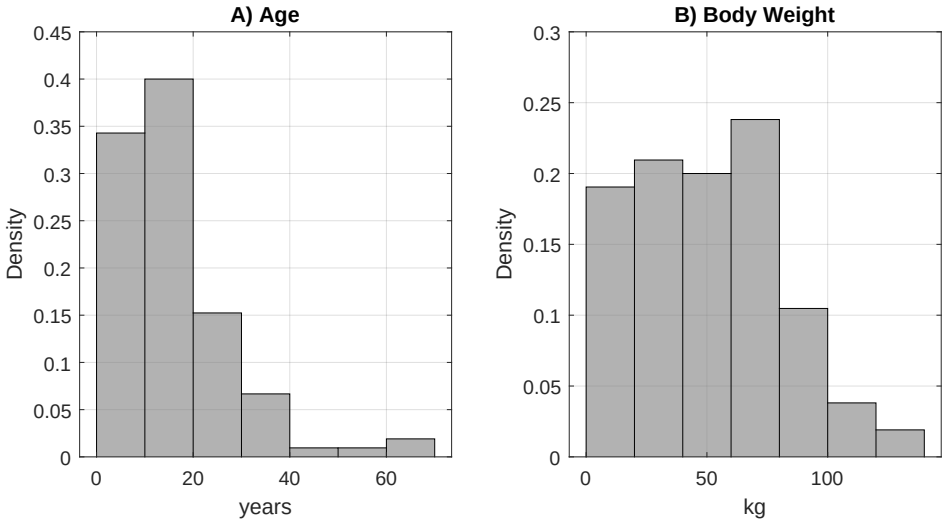

Figure 2.3: Demographics of the subjects in Dataset 1. Panel A: Distribution of age. Panel B: Distribution of body weight.

updated version of the platform that allows the simulation of meals with different macronutrient composition [36].

For each virtual subject, two different lunches were simulated. A nominal meal, consisting of 75 g of carbohydrates, 20 g of fat, and 24 g of protein, and a HF/HP, with the same composition as the nominal meal with the addition of 20 g of fat and 20 g of protein (i.e., with 75 g of carbohydrates, 40 g of fat, and 44 g of protein). Macronutrient composition of nominal and HF/HP meals is shown in Fig. 2.6. Specifically, this dataset is composed of four sub-datasets with each a set of 100 HF/HP meals with progressively more delayed dynamics with respect to nominal meals, forming Datasets 2a, 2b, 2c, and 2d, respectively. In particular, the parameters related to glucose absorption of HF/HP meals were progressively accentuated with the aim of slowing their dynamics, as shown by their CGM profiles (Fig. 2.5). CGM tracks of Datasets 2a-d, and IP profiles used for the analysis were extracted from 1 hour before the meal and 4 hours after (Fig. 2.5 and 2.4, respectively).

Basal insulin was provided through an IP delivering with microboluses every 5 minutes, whereas prandial insulin requirements were covered with boluses administered at mealtime (Fig. 2.4). Prandial insulin dosage was based on carbohydrate counting rules using the optimal Insulin-to-carbohydrate Ratio (CR) of each subject. Carbohydrate-counting error was also included in the simulation [37]. Hypo-glycemic treatments and corrective insulin boluses were

not allowed in the simulations. CGM sensor measures were collected every 5 minutes (Fig. 2.5), and synthetic SMBG measurements were generated at -15 minutes and at 120 minutes with mean equal to the current plasma glucose level and Coefficient of Variation (COV) set at 2%. The distribution of age and BW of the virtual subjects is shown in Fig. 2.6. Further details about the generation of the *in silico* subjects, the simulation platform, and the implementation of macronutrient effects in the simulator are available in [36].

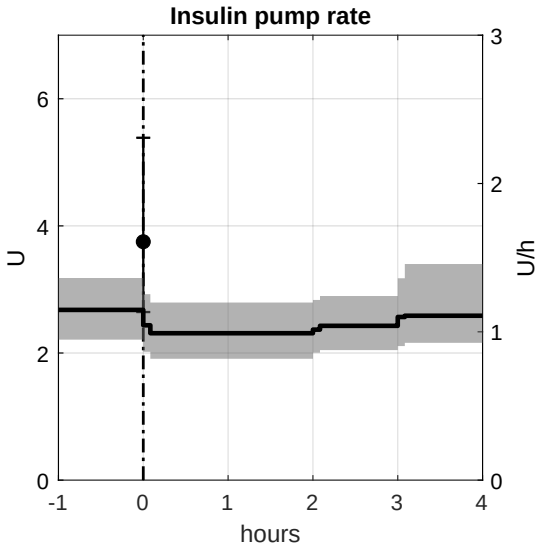

Figure 2.4: IP rates and insulin boluses in Datasets 2a-d. Vertical dashed-dotted line represents mealtime. Right y-axis: median (solid black line) and 25th-75th percentiles (gray colored area) of IP rates. Left y-axis: Median (black dot) and 25th-75th percentiles (vertical black bars) of prandial insulin boluses.

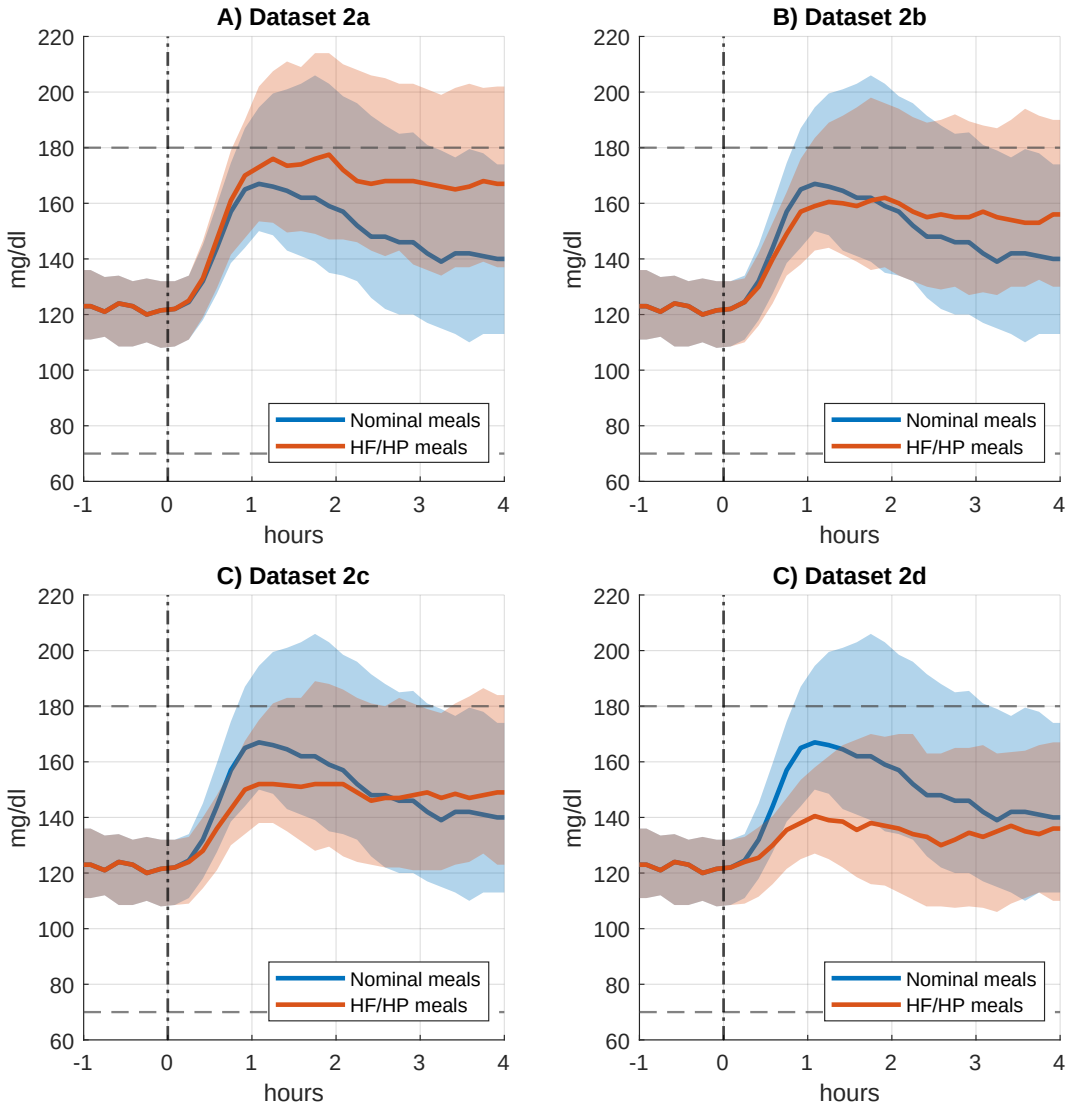

Figure 2.5: CGM measures in Dataset 2a-d. Vertical dashed-dotted line represents mealtime. Median (solid lines) and 25th-75th percentiles (colored areas) of CGM sensor measures of nominal (in blue) and HF/HP (in orange) meals. Gray horizontal dashed lines represent hypo- and hyperglycemic thresholds. Panel A: CGM measures of Dataset 2a. Panel B: CGM measures of Dataset 2b. Panel C: CGM measures of Dataset 2c. Panel D: CGM measures of Dataset 2d.

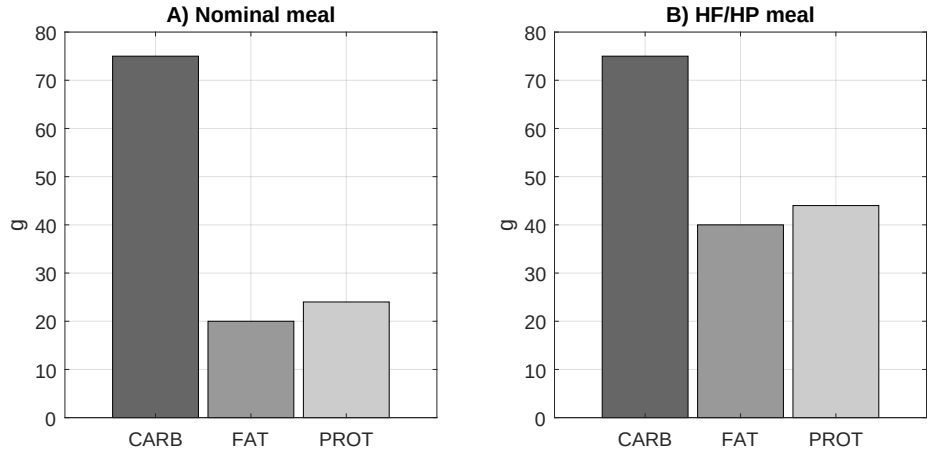

Figure 2.6: Meal macronutrient composition (gray colored bars) of Dataset 2. Panel A: nominal meal. Panel B: HF/HP meal. CARB = carbohydrates. FAT = fat. PROT = protein.

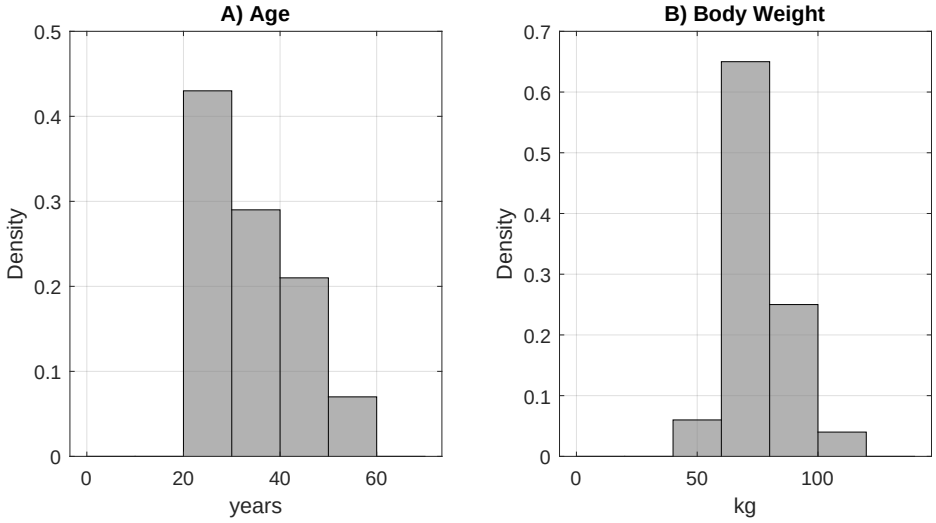

Figure 2.7: Demographics of the subjects in Dataset 2. Panel A: Distribution of age. Panel B: Distribution of body weight.

3

Methodologies

In this chapter, the methodology for developing and implementing the techniques for predicting meal class using features coming from a physiological model is presented.

First, the physiological model employed is described. This model is the MI-OMM, developed to estimate insulin sensitivity and Ra from CGM and IP data in patients with T1D.

Then, the various stages to develop a methodology for meal classification are described. Firstly, the relevant features to achieve this objective were extracted either from the demographics of the subjects, directly from the CGM profiles, or from parameters and internal states obtained by numerical identification of the MI-OMM. Following the extraction of the features, a preliminary data analysis was performed. This included a correlation analysis, useful to detect potentially important features and exclude features that are too correlated. Then, different machine learning methods were explored. In this stage, we evaluated both supervised learning methods, which utilize the known amount of macronutrients in a meal or the meal classification labels, and unsupervised learning techniques, which, on the contrary, do not require such information or labels. Last, to evaluate the impact of the model-based features in the meal classification task, the best-performing method was assessed while removing those features.

Once a good method was identified, it was then validated using a simulator platform for T1D patients (Dataset 2), where we can classify meals based on their composition and, most importantly, simulate changes in the therapy for meals classified as HF/HP.

3.1 The Minimally-Invasive Oral Minimal Model

3.1.1 Model Equations

The MI-OMM is composed of several sub-models, each describing a biological process happening in the glucose-insulin regulation system of individuals with T1D [33].

Gastro-intestinal tract model

This model allows the description of the postprandial Glucose Rate of Appearance (Ra). The model equations are the following:

$$\begin{cases} \dot{Q}_{sto1}(t) = -k_{max}Q_{sto1}(t) + D\delta(t) & Q_{sto1}(0) = 0 \\ \dot{Q}_{sto2}(t) = -k_{empt}(t)Q_{sto2}(t) + k_{max}Q_{sto1}(t) & Q_{sto2}(0) = 0 \\ \dot{Q}_{gut}(t) = -k_{abs}Q_{gut}(t) + k_{empt}(t)Q_{sto2}(t) & Q_{gut}(0) = 0 \\ Ra(t) = \frac{f}{BW}k_{abs}Q_{gut}(t) \end{cases} \tag{3.1}$$

where $Q_{sto1}(t)$ and $Q_{sto2}(t)$ [mg] represent the amount of glucose in the stomach in solid and liquid phase and $Q_{gut}(t)$ [mg] is the amount of glucose in the intestine. D [mg] is the ingested glucose, k_{max} and k_{abs} [min^{-1}] are the maximal constant rate of gastric emptying and the constant rate of intestinal absorption. f [dimensionless] represents the glucose fractional bioavailability, BW is the body weight [kg]. k_{empt} [min^{-1}] is the rate of gastric emptying defined as:

$$k_{empt}(t) = k_{min} + \frac{k_{max} - k_{min}}{2}\{\tanh[\alpha(Q_{sto}(t) - b \cdot D)] - \tanh[\beta(Q_{sto}(t) - c \cdot D)] + 2\} \tag{3.2}$$

where $Q_{sto}(t) = Q_{sto1}(t)+Q_{sto2}(t)$, i.e., the amount of glucose in the stomach, $\alpha = 5/[2 \cdot D(1-c)]$ and $\beta = 5/(2 \cdot D \cdot d)$. d and c [dimensionless] represent, respectively, the first and second flex points describing the behavior of k_{empt} as function of Q_{sto}. k_{min} and k_{max} [min^{-1}] represent the maximal and minimal constant rates of gastric emptying. More details about the model and its parameters are reported in [38].

Subcutaneous insulin absorption and plasma insulin kinetic model

This model describes the plasma insulin concentration, indicated as $I_p(t)$ [mU/l], exploiting the infused insulin by the IP as input.

$$\begin{cases} \dot{Q}_{sc1}(t) = -(k_{a1} + k_d)Q_{sc1}(t) + U(t - \tau_1) & Q_{sc1}(0) = Q_{sc1,0} \\ \dot{Q}_{sc2}(t) = -k_{a2}Q_{sc2}(t) + k_dQ_{sc1}(t), & Q_{sc2}(0) = Q_{sc2,0} \\ \dot{Q}_p(t) = -k_eI_p(t) + k_{a1}Q_{sc1}(t) + k_{a2}Q_{sc2}(t) & Q_p(0) = Q_{p,0} \\ I_p(t) = \frac{Q_p(t)}{V_I} & I_p(0) = I_{p,0} \end{cases} \tag{3.3}$$

where $Q_{sc1}(t)$ and $Q_{sc2}(t)$ [mU/kg] represent the amount of insulin in the subcutis in non-monomeric and monomeric state, $Q_p(t)$ [mU/kg] is the amount of plasma insulin. $U(t)$ [mU/kg/min] represent the IP rate infusion with τ_1 temporal delay in absorption on insulin. k_{a1} and k_{a2} [min^{-1}] represent the constant rate of insulin absorption in plasma from Q_{sc1} and Q_{sc2} compartments, k_d and k_e [min^{-1}] are the constant rate of insulin dissociation in monomers and the constant fractional clearance in plasma. V_I [l/kg] represents the distribution volume of insulin in plasma.

Oral glucose minimal model

This model, proposed by Dalla Man and coworkers [39], allows to describe the insulin sensitivity, S_I. The model equations are:

$$\begin{cases} \dot{G}(t) = -(p_1 + X(t))G(t) + p_1G_b + \frac{Ra(t)}{V_G} & G(0) = G_b \\ \dot{X}(t) = -p_2X(t) + p_2S_I(I_p(t) - I_b) & X(0) = 0 \end{cases} \tag{3.4}$$

where $G(t)$ [mg/dl] represents the glucose in plasma, $X(t)$ [min^{-1}] is the insulin action in a remote compartment, with I_p [μU/ml] plasma insulin concentration. p_1 and p_2 [min^{-1}] represent, respectively, the fractional glucose effectiveness and the rate constant describing the dynamics of insulin action. S_I [$min^{-1}/(\mu U/ml)$] represents the insulin sensitivity, V_G [dl/kg] is the volume of glucose distribution. G_b [mg/dl] is the steady state level of plasma glucose and I_b [μU/ml] the steady state level of plasma insulin.

Plasma-to-interstitium glucose kinetics model

This model describes the diffusion process from the blood to the interstitium using a single compartmental model.

$$\dot{G}_i(t) = -\frac{1}{\tau_G} G_i(t) + \frac{1}{\tau_G} G(t) \quad G_i(0) = G_{i,0} \tag{3.5}$$

where $G_i(t)$ [mg/dl] represents the interstitial glucose concentration and τ_G [min] is the equilibration time constant between plasma and interstitium.

3.1.2 Model Identification

The MI-OMM was identified on the extracted meals employing minimally-invasive data, using the subcutaneously injected exogenous insulin profiles and the carbohydrate amount consumed in the meal as model inputs, and CGM sensor measurements and SMBG samples, if available, as output.

Model parameters were estimated using a Maximum a Posteriori (MAP) estimator [40] assuming a log-normal *a priori* distribution, with known mean μ_p and covariance matrix Σ_p of the log-transformed model parameters coming from the literature [41, 42, 43].

In other words, the minimization problem can be formalized as follows:

$$\hat{p} = \underset{p}{\operatorname{argmin}}\, J(p) \tag{3.6}$$

where $J(p)$ is the objective function to minimize, p is the set of unknown model parameters, and $\hat{p}$ represents their estimate using the MAP estimator. Therefore, the objective function is:

$$J(\hat{p}) = \hat{v}^T \Sigma_v^{-1} \hat{v} + \hat{w}^T \Sigma_w^{-1} \hat{w} + (\hat{p} - \mu_p)^T \Sigma_p^{-1} (\hat{p} - \mu_p) \tag{3.7}$$

where $\hat{v}$ is defined as the difference between measurements of the CGM sensor and model prediction, Σ_v is the covariance matrix of the autocorrelated CGM error noise, modelled with a second order autoregressive model [43], $\hat{w}$ is the difference between the measured SMBG value and the predicted plasma glucose, and Σ_w is the covariance matrix of the measurement error noise of the SMBG samples [44].

The numerical identification was carried out using the computing software

MATLAB (The MathWorks Inc., Version: R2024b Update 4 [45]) using `ode45.m` to solve the ordinary differential equations of the MI-OMM and `lsqnonlin.m` to solve the MAP minimization problem (eq. 3.6) and estimate model unknown parameters. More details about the model are outlined in [33].

3.2 Features Extraction

In this section, the complete list of extracted features used for the subsequent analysis is presented. The features were derived from: (i) demographic data; (ii) CGM sensor profiles; (iii) numerical identification of the MI-OMM.

3.2.1 Demographic Data

For each subject, the following demographic information was extracted:

- Age [years]
- BW [kg]
- Sex, as a categorical variable (male or female)
- BMI [kg/m^2]

In Dataset 2, no information on Sex and BMI is generated from the simulator. Regarding the BMI, needed for the numerical identification of the MI-OMM, the value was calculated from BW by fixing the height of each *in silico* subject to 1.8 m.

3.2.2 Glucose Rate of Change

Following the work of Ekhlaspour and coworkers [46], which suggested that the glucose rate of change is significantly lower in HF meals with respect to Low-fat (LF) meals between 32-62 min after mealtime and higher between 104-150 min, the interstitial glucose rate of change (i.e., the derivative of the interstitial glucose), within those time windows, was extracted.

Specifically, we exploited the model of plasma-to-interstitium glucose kinetics (Eq. 3.5), to assess the derivatives of $G_i(t)$ [mg/dl/min] at 45 min and at 120 min were extracted, i.e., $\dot{G}_i(t=45min)$ and $\dot{G}_i(t=120min)$. A visual representation is provided in Fig. 3.1.

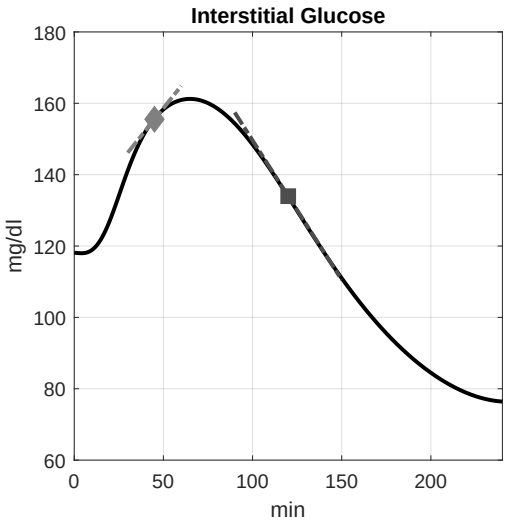

Figure 3.1: Interstitial glucose curve (solid black line) with its derivative (dashed gray lines) evaluated at 45 minutes (light-gray diamond) and 120 minutes (dark-gray square) after mealtime for a representative subject.

3.2.3 Physiologically-based Features

Features obtained from numerical identification of the MI-OMM were also used in the analysis. We evaluated the use of estimated model parameters together with features calculated from inner model states. This set of features was chosen as it is known to be causally linked with meal composition; thus, we expected them to be relevant for meal classification.

Insulin Sensitivity

The first feature extracted from the MI-OMM was insulin sensitivity (parameter S_I [$\text{min}^{-1}/(\mu\text{U/ml})$] in Eq. 3.4). This parameter summarizes the ability of insulin to lower glucose production from the liver and to promote glucose uptake.

Glucose Absorption Parameters

The following features, related to glucose absorption, were extracted directly from the estimated model parameters: k_{min} [min^{-1}], k_{max} [min^{-1}], c and d [dimensionless] from Eq. 3.2 as indicated from Fig. 3.2 and k_{abs} [min^{-1}] from Eq. 3.1.

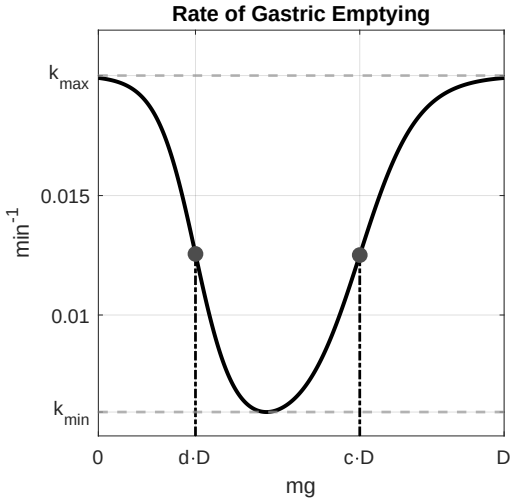

Figure 3.2: Rate of gastric emptying (solid black line) as function of the total glucose in the stomach, Q_{sto}, for a representative subject. Minimal and maximal constant rates are indicated with gray dashed lines, and flex points with black dash-and-dot lines.

Basal Plasma Glucose Concentration

From the identification of the MI-OMM, we also extracted the basal plasma glucose cocnetration (parameter G_b [mg/dl] in Eq. 3.4), which defines the steady-state level of the plasma glucose compartment at a given level of insulin (I_b [μU/ml] in Eq. 3.4).

Glycemia at Mealtime

Exploiting the plasma-to-interstitium glucose kinetics model (Eq. 3.5), the linear interpolation of the interstitial glucose concentration G_i [mg/dl] in the sampling grid of the CGM sensor measurements, that is the prediction of the CGM profile, indicated as G_s [mg/dl], was calculated. The feature extracted is the value of G_s extrapolated on the first CGM sensor sample after mealtime, i.e., $G_s(t_{CGM}=0min)$, as shown in Fig. 3.3.

Times of Gastric Retention

From the gastro-intestinal tract model (Eq. 3.1), the times for the stomach to digest 75%, 50% and 25% of the glucose present in the meal were calculated from GR, GR(t) [%]. This is defined as the ratio between the total amount of

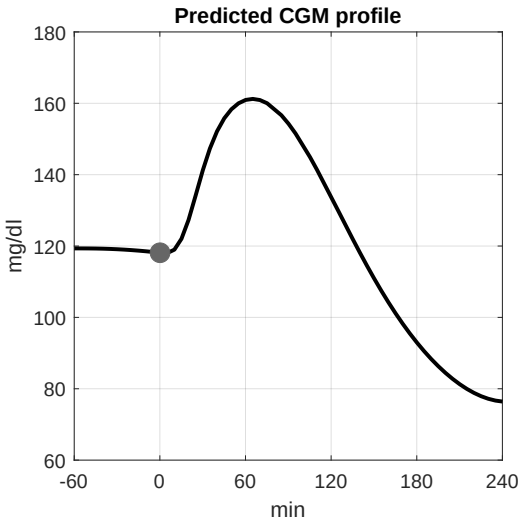

Figure 3.3: Predicted CGM profile curve (solid black line) and its value at mealtime (gray dot) for a representative subject.

glucose in the stomach over time, $Q_{sto}(t)$ [mg], and the total ingested dose of glucose, D [mg]:

$$GR(t) = 100 \cdot \frac{Q_{sto}(t)}{D} \tag{3.8}$$

From GR(t), the time to reach a certain GR(t) was extracted as a feature for the analysis:

$$TGR(P) = \{t : GR(t) = P\} \quad [min] \tag{3.9}$$

with P = 25%, 50%, and 75%. In Fig. 3.4, the calculation of the different TGR(P) is illustrated.

Areas Under the Curve of Glucose Rate of Appearance

The rate of appearance of glucose in plasma, Ra(t) [min^{-1} mg/kg] is described by the model with the gastro-intestinal tract equations (Eq. 3.1). A normalized version of this curve, NRa(t) [min^{-1}], was used to avoid meal and body size effects in the analysis. NRa(t) was calculated by dividing Ra(t) by the consumed carbohydrate amount in the meal, D [mg], normalized by subject's BW [kg]:

$$NRa(t) = Ra(t) \cdot \frac{1}{D/BW} \tag{3.10}$$

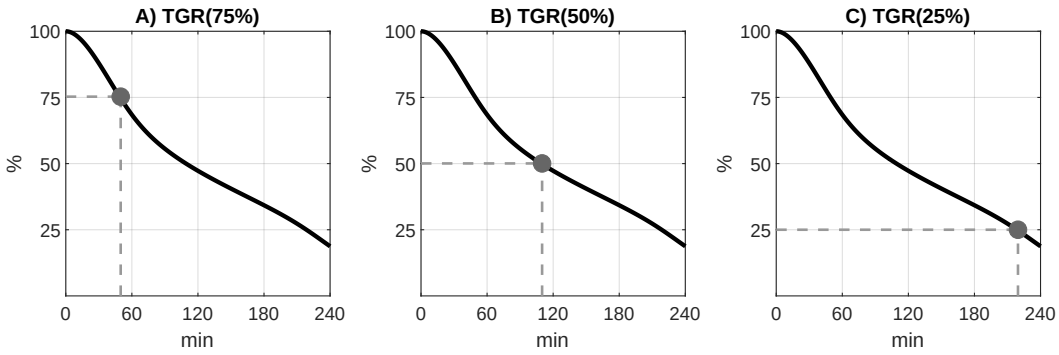

Figure 3.4: Gastric Retention curve (black solid line) and time required to reach the digestion of a percentage of glucose equal to P (gray dashed lines and gray dot) for a representative subject. Panel A: P = 75%. Panel B: P = 50%. Panel C: P = 25%.

From this, the Area Under the Curve (AUC) of NRa(t), was computed at different time intervals, T [min], and used as features for the subsequent meal analysis:

$$AUC_{Ra}(T) = \int_0^T NRa(t)dt \tag{3.11}$$

with T=30, 60, 90, 120 min. In Fig. 3.5, a visual representation of this index is reported.

3.2.4 Macronutrient Transformations

Other than the features that were used to predict the meal class, the following macronutrient transformations were calculated and used for the development and evaluation of the methodologies for meal classification:

- Amount [g]: The absolute amounts [g] consumed of fat, protein, and carbohydrates, hereinafter indicated as FAT_g, $PROT_g$, and $CARB_g$, respectively;
- Amount contribution [%]: The percentage of each macronutrient to the total amount of macronutrients, defined as follows:

$$FAT_{g\%} = \frac{FAT_g}{FAT_g + PROT_g + CARB_g} \tag{3.12}$$

$$PROT_{g\%} = \frac{PROT_g}{FAT_g + PROT_g + CARB_g} \tag{3.13}$$

$$CARB_{g\%} = \frac{CARB_g}{FAT_g + PROT_g + CARB_g} \tag{3.14}$$

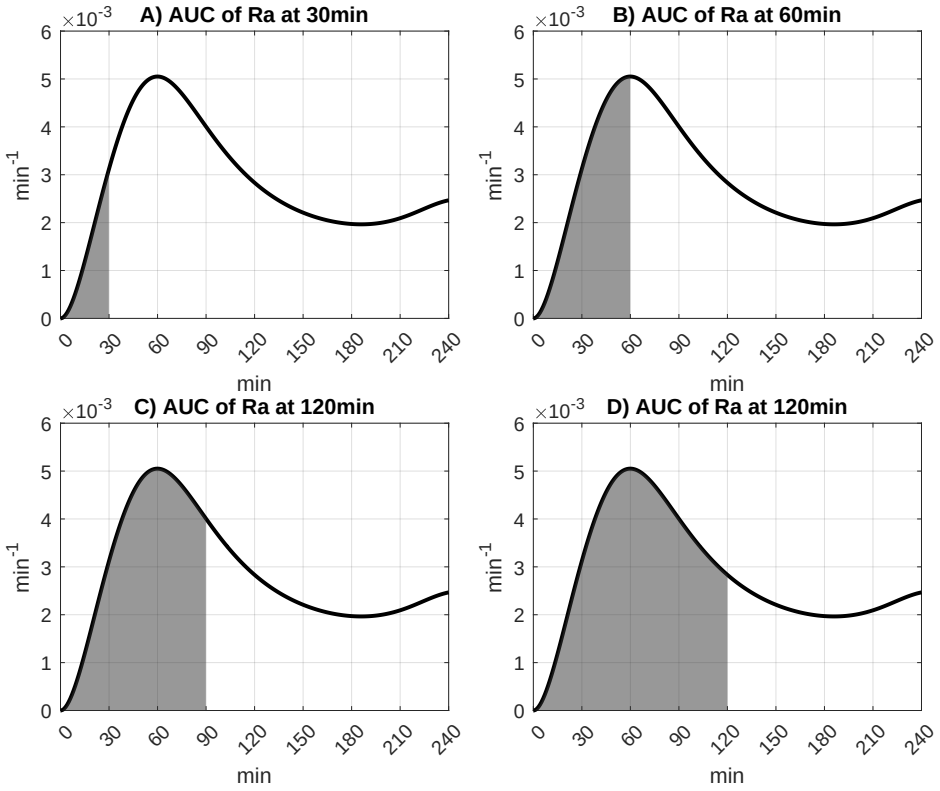

Figure 3.5: Normalized Glucose Rate of Appearance (black line) and its AUC (gray area) calculated for a representative subject from mealtime up until T. Panel A: T = 30 minutes. Panel B: T = 60 minutes. T = Panel C. T = 90 minutes. Panel D: T = 120 minutes.

- Caloric contribution [%]: The percentage of each caloric intake to the total amount of calories, defined as follows:

$$\mathrm{FAT_{kcal\%}} = \frac{\mathrm{FAT_{kcal}}}{\mathrm{FAT_{kcal}} + \mathrm{PROT_{kcal}} + \mathrm{CARB_{kcal}}} \tag{3.15}$$

$$\mathrm{PROT_{kcal\%}} = \frac{\mathrm{PROT_{kcal}}}{\mathrm{FAT_{kcal}} + \mathrm{PROT_{kcal}} + \mathrm{CARB_{kcal}}} \tag{3.16}$$

$$\mathrm{CARB_{kcal\%}} = \frac{\mathrm{CARB_{kcal}}}{\mathrm{FAT_{kcal}} + \mathrm{PROT_{kcal}} + \mathrm{CARB_{kcal}}} \tag{3.17}$$

  where $\mathrm{FAT_{kcal}}$, $\mathrm{PROT_{kcal}}$, and $\mathrm{CARB_{kcal}}$ are the caloric amounts [kcal] of fat, protein, and carbohydrates, respectively, calculated from their amounts [g] as $\mathrm{FAT_{kcal}}$=9·$\mathrm{FAT_g}$, $\mathrm{PROT_{kcal}}$=4·$\mathrm{PROT_g}$, $\mathrm{CARB_{kcal}}$=4·$\mathrm{CARB_g}$;

- Amount of fat and protein [g]: The sum of fat and protein amounts ($\mathrm{FAT_g}$+$\mathrm{PROT_g}$);

- Amount contribution of fat and protein [%]: The sum of fat and protein amount contributions ($\mathrm{FAT_{g\%}}$+$\mathrm{PROT_{g\%}}$);

- Caloric contribution of fat and protein [%]: The sum of fat and protein caloric contribution ($FAT_{kcal\%}$+$PROT_{kcal\%}$).

Of note, other transformations were not attempted since they would have led to one of the already tested transformations after feature normalization (see Section 3.2.5).

3.2.5 Features Normalization

All extracted features, including macronutrient transformations, were normalized to ensure that variables with differing scales or distributions were given equal weight in the subsequent analyses. This is particularly important, especially during model training.

In Dataset 1, z-score normalization was applied. Each feature (or macronutrient) was rescaled by subtracting its mean and dividing by its standard deviation, so that the normalized feature belongs to a distribution with zero mean and unitary standard deviation.

For Dataset 2, normalization was performed using the mean and standard deviation computed from Dataset 1, to ensure a fair validation of the developed methodology on unseen scenarios.

3.3 Preliminary Analysis

A preliminary correlation analysis of the extracted features and macronutrient transformations was performed in Dataset 1. The goal was to identify possible meaningful relationships among features and assess their strength. This initial investigation serves as the starting point of the methodologies that were tested in the subsequent steps, as it helps to highlight features that are linked with macronutrients in the meal, and thus, may be useful in the development of a technique for meal classification. On the other hand, this analysis detects features that are too correlated, and thus, non-informative, which are therefore removed from the set of features to avoid redundancy.

First, Spearman's Rank Correlation was computed to evaluate the correlation among features, as well as between features and macronutrients. Spearman's correlation coefficient, ρ, quantifies how strong the relationship between two variables can be described using a monotonic function.

Second, a corrected version of Pearson's correlation for Repeated Measures [47] was computed to evaluate inter-individual associations in measures acquired in two or more occasions. By accounting for variability across multiple observations, this statistical technique allows for reducing the bias caused by simple correlation in aggregated data. However, unlike Spearman's, this type of correlation detects only linear relationships.

3.3.1 Spearman's Rank Correlation

The first analysis was performed to find the degree of association among the extracted features. Spearman's rank correlation coefficient, ρ, was computed for each pair of features. The rationale lies in identifying potentially redundant features. A high absolute Spearman's ρ between two features suggests a strong monotonic relation, indicating that the features may carry overlapping information.

Then, the degree of association between each feature and the macronutrients was analyzed. The Spearman's rank correlation coefficient was used to quantify the strength and direction of these associations to detect the most descriptive features. A high correlation, in absolute value, between a given feature and a macronutrient may indicate a strong relation, suggesting that the feature could be informative and could help in performing meal classification.

3.3.2 Repeated Measures Correlation

Repeated measures correlation is a statistical technique used to assess the association between two variables measured on multiple occasions from the same individuals. Unlike standard correlation methods, it accounts for intra-individual variability and is able to correct for the presence of measures coming from the same subject. Similarly to other correlation coefficients, the repeated measures correlation coefficient varies between -1 and 1 and represents the strength and direction of the relation. In this case, the correlation coefficient is obtained by computing the best regression line for each individual, with a common slope shared by all individuals, but varying intercepts [47].

In this thesis work, the repeated measures correlation was computed for each feature-macronutrient pair, and provides more reliable correlation values as it accounts for variability between individuals. The analysis was performed

using R Statistical Programming Software (version 4.3.2) [48] and employing the `rmcorr` R package [49] to compute repeated measures correlation.

3.4 Supervised Learning

The first machine learning techniques tested for meal classification are two supervised learning approaches. In supervised learning, other than the features, the desired outcome is also employed during model training.

First, a battery of Linear Mixed Effects (LME) models was tested. This class of models is particularly suited to settings with repeated measures and can work with continuous responses (i.e., the various macronutrient transformations). Second, binary logistic regression models were built, where the classes to be predicted are binary categorical outcomes representing meal classes (i.e., the meal type).

3.4.1 Linear Mixed Effects Modeling

LME modeling is a supervised statistical technique that is characterized by extending linear regression modeling for data that are collected in groups. A LME model tries to explain the relation between a continuous response variable (i.e., macronutrient transformations, such as the fat amount, in our case), and independent variables (i.e., the extracted set of features). The peculiarity of LME models is that they describe this relationship with a hierarchy of models: the linear model describing the individual response, and a stochastic model describing the variability of linear model parameters. The residual unexplained variability is usually assumed to be normally distributed. Parameters of the population linear model are the so-called 'fixed effects', which are shared among all the subjects. On the other hand, the specific individual realization is defined by the so-called 'random effects' and can be considered as the deviation of the individual linear parameter from the respective fixed effect. [50]

The general form of a LME model can be written as:

$$Y = X\beta + Zb + \varepsilon \tag{3.18}$$

where Y is the response vector, the term β represents the fixed effects, shared among all individuals, with X the corresponding design matrix, the term b refers

to the random effects, different for each individual, with Z its design matrix. b is assumed to be a vector of random variables normally distributed with zero mean and variance Ψ. The error representing what remains unexplained by the model is represented by ε assumed to be Gaussian random noise with variance $\sigma\mathbb{I}$. Therefore, the measured outcome assumes a distribution of the form $Y \sim \mathcal{N}(X\beta, Z\Psi Z^T + \sigma\mathbb{I})$.

As an illustrative example, let's consider an LME model formulated to predict the fat amount. In this model, the fixed effects include S_I, TGR(25%), and k_{min}, along with the intercept term. No interaction terms are specified among the fixed effect predictors. Therefore, the total number of model parameters is $p = 4$. The random effects are included for the intercept and for S_I, defining the deviation of the individual parameter from the respective population estimate, i.e., the fixed effect. This specification allows both the intercept and the effect of S_I to vary across individuals. Meanwhile, TGR(25%) and k_{min} are not allowed to vary between individuals, as in the LME model, and are present only as fixed effects. No interaction terms are specified among the random effects. In this case, the total number of random parameters is $q = 2$.

Consider Dataset 1, composed of $n = 353$ meals recorded for $m = 105$ individuals. Assuming that subject 1 has $n_1 = 2$ meals recorded and subject 105 has $n_{105} = 4$ meals, the design matrices of the LME model are constructed as follows:

- The fixed effects design matrix, X, is a $n \times p$ matrix defined as follows:

$$X = \begin{bmatrix} 1 & S_{I\,1,1} & TGR(50\%)_{1,1} & k_{min\,1,1} \\ 1 & S_{I\,1,2} & TGR(50\%)_{1,2} & k_{min\,1,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & S_{I\,i,j} & TGR(50\%)_{i,j} & k_{min\,i,j} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & S_{I\,150,1} & TGR(50\%)_{150,1} & k_{min\,150,1} \\ 1 & S_{I\,150,2} & TGR(50\%)_{150,2} & k_{min\,150,2} \\ 1 & S_{I\,150,3} & TGR(50\%)_{150,3} & k_{min\,150,3} \\ 1 & S_{I\,150,4} & TGR(50\%)_{150,4} & k_{min\,150,4} \end{bmatrix} \in \mathbb{R}^{353\times 4} \tag{3.19}$$

  where the first column, composed of 1s, represents the intercept term, the following columns represent, respectively, S_I, TGR(25%), and k_{min} values of the j-th meal underwent by the i-th subject. Therefore, the fixed effects, are contained in $\beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3]^T \in \mathbb{R}^{p\times 1}$, that multiplies X.

- The random effects design matrix, Z, is a $n \times (q \cdot m)$ block-diagonal matrix

defined as follows:

$$Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Z_{150} \end{bmatrix} \in \mathbb{R}^{353\times210} \tag{3.20}$$

where Z_i represents the random effect matrix for the i-th subject. In this case, Z_1 is a $q \times n_1$ matrix, and Z_{150} is a $q \times n_{105}$ matrix:

$$Z_1 = \begin{bmatrix} 1 & \mathrm{S_{I\,1,1}} \\ 1 & \mathrm{S_{I\,1,2}} \end{bmatrix}; \qquad Z_{150} = \begin{bmatrix} 1 & \mathrm{S_{I\,1,1}} \\ 1 & \mathrm{S_{I\,1,2}} \\ 1 & \mathrm{S_{I\,1,3}} \\ 1 & \mathrm{S_{I\,1,4}} \end{bmatrix} \tag{3.21}$$

with, analogously to the X matrix, the first column composed of 1s, representing the intercept term, and the second column, composed of $\mathrm{S_I}$ values. Therefore, the random effects, which can be considered as deviations from the fixed effects, are contained in $b \in \mathbb{R}^{(q\cdot m)\times1}$, that multiplies Z. Specifically, the structure of the column vector b is the following:

$$b = \begin{bmatrix} b_{0,1} \\ b_{1,1} \\ \vdots \\ b_{0,150} \\ b_{1,150} \end{bmatrix} \in \mathbb{R}^{210\times1} \tag{3.22}$$

with $b_{0,\mathrm{i}}$ and $b_{1,\mathrm{i}}$ representing, respectively, the i-th subject-specific deviation of the intercept and $\mathrm{S_I}$ from the fixed effects contained in β.

To summarize, the LME model of fat amount, represented by the response vector Y$\in \mathbb{R}^{353\times1}$, for the j-th meal of the i-th individual, indicated as $y_{\mathrm{i,j}}$, can be expressed as:

$$y_{\mathrm{i,j}} = \beta_0 + \beta_1\,\mathrm{S_{I\,i,j}} + \beta_2\,\mathrm{TGR(25\%)_{i,j}} + \beta_3\,\mathrm{k_{min\,i,j}} + \mathrm{b_{0,i}} + \mathrm{b_{1,i}}\,\mathrm{S_{I\,i,j}} + \varepsilon_{\mathrm{i,j}} \tag{3.23}$$

LME models were applied to the dataset in order to find a relation between a macronutrient transformation and the extracted features. The primary aim was to find the best macronutrient transformation that is explained by a subset of the extracted features and to assess the predictive capability of the model.

To determine the best model for each macronutrient transformation, an exhaustive search was conducted across all possible feature combinations. For each combination, the model was evaluated using the Akaike Information Criterion (AIC), and the model with the lowest AIC was selected as the optimal one. All the extracted features were employed as predictors, excluding demographics and the features with an absolute Spearman's $\rho \geq 0.9$.

Then, to evaluate which macronutrient transformation was best described by the available features, two metrics were designed, one based on the Sum of

Squares Error (SSE) and the second one based on the Log-likelihood (LL), which also accounts for the uncertainty of the measures.

SSE is simply defined, in matrix form, as

$$\mathrm{SSE} = (\mathrm{y} - \mathrm{X}\beta)^{\mathrm{T}}(\mathrm{y} - \mathrm{X}\beta) \tag{3.24}$$

Meanwhile, the LL of the LME model is defined as

$$\mathrm{LL} = -\frac{\mathrm{n}}{2}\log(2\pi) - \frac{1}{2}\log[\det(\mathrm{V})] - \frac{1}{2}(\mathrm{y} - \mathrm{X}\beta)^{\mathrm{T}}\mathrm{V}^{-1}(\mathrm{y} - \mathrm{X}\beta) \tag{3.25}$$

where V, the covariance matrix of Y, is defined as $\mathrm{V} = \mathrm{Z}\Psi\mathrm{Z}^{\mathrm{T}} + \sigma\mathbb{I}$ [51].

A null model, defined as the model that includes no features (i.e., without any fixed effect), was identified for each tested macronutrient transformation. From this null model, the baselines of SSE and LL (denoted with SSE_0 and LL_0, respectively) were calculated. Then, SSE and LL were computed also for the best model suggested by the AIC (denoted with $\mathrm{SSE}_{\mathrm{best}}$ and $\mathrm{LL}_{\mathrm{best}}$, respectively). The two performance metrics were then defined as

$$\Delta\mathrm{LL} = \mathrm{LL}_0 - \mathrm{LL}_{\mathrm{best}} \tag{3.26}$$

$$\Delta\mathrm{SSE} = \mathrm{SSE}_0 - \mathrm{SSE}_{\mathrm{best}} \tag{3.27}$$

The macronutrient transformation that was best predicted by an LME model was identified as the one that achieved the highest improvement over its null model, or, in other words, the one achieving the highest absolute value of either ΔLL or ΔSSE.

In the LME models implemented, random effects were designed by allowing only the intercept to vary across subjects. Specifically, each model included a random intercept for each individual, as subject-specific corrections from the overall population trend. The fixed effects represent a common slope shared among all individuals, while the random effects account for individual-level variability through the intercept. Thus, each model assumes that all subjects respond similarly to the predictors (same slope) but start from different baseline levels (varying intercepts).

To compare the outcomes of LME models with varying intercept, also models with random slopes added to the random intercept as random effects, as inter-individual corrections for each fixed effect, were tested.

First, the performances of the developed models were evaluated in terms of SSE and their improvement from the null models. Second, the residual distributions, as well as the observation-prediction plots, were inspected for the best models related to fat and protein, comparing them with their respective null model and checking whether LME models are suitable for meal classification purposes both at the individual and at the population level, by turning off random effects.

3.4.2 Logistic Regression

Logistic regression is a statistical method for modeling the relationship between one or more predictor variables and a categorical dependent variable. Unlike LME models, which predict continuous outcomes, logistic regression models estimate the probability that a given observation belongs to a particular category. It is particularly suited for binary classification problems, where the response variable has two possible outcomes.

In this case, logistic regression models are used to study the effects of predictor variables, i.e., the extracted features, on categorical outcomes, i.e., meal classes obtained from macronutrient transformations. Since the macronutrients are on a continuous scale, to implement binary classification, the categorical outcome was created by assigning a positive label (H-class) if the macronutrient was above its mean value, or negative (L-class) otherwise.

First, to avoid instability of the model, pairs of highly correlated features (with an absolute Spearman's $\rho \geq 0.9$) were identified, and one of the two was removed. Second, the observations composing the dataset were split into a training set and a test set, retaining respectively 70% and 30% of the total observations. This was done while ensuring the preservation of the relative proportion of each class in the train and test sets; therefore, maintaining the original class distribution. To confirm the validity of the split, the feature distributions in the two sets were visually inspected and compared.

To identify the optimal model for each binarized macronutrient transformation, i.e., to select the most informative set of features, a backwards stepwise selection procedure was employed. Starting from the full linear model that included all available features (excluding those exhibiting high correlation and demographics), features were iteratively removed. At each step, the feature whose exclusion resulted in the greatest reduction in the AIC was discarded.

This process continued until no further reduction in AIC could be achieved, obtaining a final set of predictors.

After identifying the best model for each macronutrient transformation using the training set, the predicted probability of each observation in the test set belonging to the positive class was computed. To determine the classification threshold, denoted as α, such that an observation is assigned to the positive class if its predicted probability exceeds α, three different threshold selection strategies were evaluated:

1. α at maximum allowable False Positive Rate (FPR):
   The threshold corresponding to a maximum FPR of 5% on the Receiver Operating Characteristic (ROC) curve [52].
   The rationale for setting the threshold at FPR= 5% was to strictly limit the proportion of L-class meals incorrectly classified as H-class, as such misclassification could lead to inappropriate insulin therapy. In particular, L-class meals misclassified as H-class might receive a higher insulin dose than required, causing potential safety concerns, whereas misclassifying an H-class meal would not produce any intervention.

   Formally, FPR can be parametrized as function of the threshold α, and it's defined as:
   $$\mathrm{FPR}(\alpha) = \frac{\mathrm{FP}(\alpha)}{\mathrm{FP}(\alpha) + \mathrm{TN}(\alpha)} \tag{3.28}$$
   where FP are the false positives and TN the true negatives counts when employing a probability threshold α.
   The threshold at maximum allowable FPR (5%), $\alpha_{\mathrm{FPR}\,5\%}$, is found by solving the following minimization problem:
   $$\alpha_{\mathrm{FPR}\,5\%} = \underset{\alpha \in [0,1]}{\operatorname{argmin}} \, |\mathrm{FPR}(\alpha) - 0.05| \tag{3.29}$$

2. α at the optimal operating point on the ROC curve:
   The point closest to the top-left corner of the ROC curve that maximizes the trade-off between sensitivity and specificity.
   In other words, the aim was to find the Optimal Operating Point (OOP) of the ROC curve that minimizes the FPR and maximizes the True Positive Rate (TPR). Therefore probability threshold related to the OOP, α_{OOP}, can be found by solving the following:
   $$\alpha_{\mathrm{OOP}} = \underset{\alpha \in [0,1]}{\operatorname{argmin}} \sqrt{(\mathrm{FPR}(\alpha))^2 + (1 - \mathrm{TPR}(\alpha))^2} \tag{3.30}$$

3. α maximizing the F1-score:
   The threshold that yields the highest F1-score, also referred to as F-measure [53], on the precision-recall curve [54].
   The F1-score, parametrized as function of α, is defined as:
   $$\mathrm{F1}(\alpha) = \frac{2}{\mathrm{precision}^{-1}(\alpha) + \mathrm{recall}^{-1}(\alpha)} = \frac{2\mathrm{TP}(\alpha)}{2\mathrm{TP}(\alpha) + \mathrm{FP}(\alpha) + \mathrm{FN}(\alpha)} \tag{3.31}$$

where TP, FP, and FN are, respectively, the true positives, false positives, and false negatives counts obtained employing the threshold α. The threshold that maximizes the F1-score, α_{maxF1}, is found by solving the following:

$$\alpha_{\text{maxF1}} = \underset{\alpha \in [0,1]}{\text{argmax}}\, \text{F1}(\alpha) \tag{3.32}$$

The model performance was evaluated by analyzing the ROC and the prediction-recall curves. The predictive accuracy was calculated from the confusion matrices.

3.5 Unsupervised Learning

After exploring supervised learning techniques for meal classification, unsupervised learning methods were tested. Unsupervised approaches do not rely on a predefined label. In our case, this can be a powerful advantage since the definition of HF/HP meal is highly subjective (the same amount of macronutrient differently affects glucose metabolism in different people, e.g., due to dietary habits, age, and weight) and is based on the measured macronutrient (which is noisy, since it is measured from a person). In addition, no standard consensus on the definition of HF/HP meal across different populations is available in the literature. Therefore, we deemed it useful to attempt approaches that are free from the use of labels.

The objective is to assess whether unsupervised learning techniques were able to find hidden patterns in Dataset 1 and split it into different groups representing a particular meal class, utilizing only the extracted features.

The unsupervised learning methods implemented and analyzed were k-means clustering, hierarchical clustering, and Gaussian Mixture Model (GMM) clustering. A summary of the main characteristics of each technique is provided in Table 3.1.

An unsupervised Cross Validation (CV) algorithm was implemented on each of the clustering techniques to identify the best one and to find the set of features that maximizes the CV prediction accuracy.

To visualize and interpret the clustering structure, observations from both Dataset 1 and Dataset 2 were projected into the feature space defined by the two most relevant features, as determined by a feature importance metric based on the Silhouette coefficient. Specifically, the importance of a feature p was

Table 3.1: Summary of the characteristics of unsupervised learning methods tested.

	Optimization function	Input	Clusters Type	Assignment
k-means clustering	Distance between objects and centroids	Continuous observations	Spheroidal clusters with equal diagonal variance	Hard
Hierarchical clustering	Distance between objects	Pairwise distance between observations	Arbitrarily shaped clusters depending on linkage method	Hard
GMM clustering	Mixture of Gaussian distributions	Continuous observations	Spheroidal clusters with different covariance structures	Soft

quantified using the following measure:

$$\Delta\bar{S}_p = \bar{S} - \bar{S}_{-p} \tag{3.33}$$

where $\bar{S}$ denotes the mean Silhouette score (see Eq. 3.34) of the clustering when using all features employed to train the clustering, and $\bar{S}_{-p}$ represents the mean silhouette score when feature p is excluded. A larger $\Delta\bar{S}_p$ indicates a greater contribution of feature p to the clustering. This analysis was conducted for each feature used in the model, and the two most influential features were selected for visualization. The resulting plots were used to assess the physiological relevance of the identified clusters and to evaluate their consistency with findings reported in the literature.

For each of the clustering techniques, the clustering efficacy was analyzed by checking the distribution of the macronutrients of each cluster as external cluster labels. The Wilcoxon Rank Sum test was utilized to check whether the medians of the macronutrient content of the two clusters were different. Furthermore, the distributions of individual features within each cluster were examined to better understand the physiological characteristics associated with each group. This analysis served to assess the alignment between the clustering outcomes and established findings in the literature.

3.5.1 *K*-means Clustering

K-means clustering was implemented and applied to Dataset 1 to group observations into k mutually exclusive partitions. In this unsupervised learning

technique, each cluster is characterized by a cluster center, called centroid. Each observation is assigned to the cluster that minimizes the distance to its centroid, based on a predefined distance metric. With an iterative procedure, k-means clustering minimizes the sum of the distances between the cluster's centroid and the observation belonging to the cluster, until convergence. In our case, the squared Euclidean distance was selected as the distance metric for evaluating proximity between observations and centroids. In addition, since in k-means each observation has a location in space, the algorithm is sensitive to feature scaling and requires normalization of the data.

To improve initialization of clustering centroids coordinates the k-means++ algorithm was employed, allowing for avoiding falling into a local minima when computing clustering, improving computing times and clustering accuracy [55].

To further minimize the risk of converging to a local minimum, the clustering process was repeated 5 times with different randomly assigned initial centroids. Thereafter, the clustering with the lowest total sum of distances amongst all the replicates was chosen as the best.

To identify the optimal number of clusters for k-means, which performs hard assignment of observations (i.e., each item is assigned to a single cluster), the Silhouette method was employed. In our case, candidate values for the number of clusters, k, ranged from 2 to 6. For each k, clustering was performed and the mean silhouette score $\bar{S}_k$ was computed. The Silhouette score, S, described in Eq. 3.35, ranges from -1, indicating observations assigned to a wrong cluster, through 0, indicating observations that are at the border of clusters, to 1, indicating observations assigned to a cluster correctly and very distant from the neighboring ones [56]. In other words, S quantifies how well each observation lies within its assigned cluster compared to the others. Higher silhouette values indicate better clustering.

Formally, $\bar{S}_k$ is defined as:

$$\bar{S}_k = \frac{1}{n} \sum_{\mathrm{i}=1}^{n} S(\mathrm{i}) \tag{3.34}$$

where $S(\mathrm{i})$ is the Silhouette score for the i-th observation and n is the total amount of observations. The Silhouette score for the i-th observation, $S(\mathrm{i})$, assigned to a

cluster, indicated as C, can be written as follows:

$$S(\mathrm{i}) = \frac{b(\mathrm{i}) - a(\mathrm{i})}{\max\{a(\mathrm{i}), b(\mathrm{i})\}} \tag{3.35}$$

where $a(\mathrm{i})$ is the mean distance to the other members of C and $b(\mathrm{i})$ is the smallest mean distance to all observations of any other cluster that is not C.

The optimal number of clusters k^* was selected as the value that maximizes the average Silhouette score:

$$k^* = \underset{k}{\operatorname{argmax}}\, \bar{S}_k \tag{3.36}$$

In k-means clustering, an unseen observation was assigned to the cluster whose centroid was closer according to a distance metric. The two distance metrics employed and tested for this purpose were:

- Squared Euclidean distance between the observation, p, and the centroid, c, defined as: $d(p, c) = \sqrt{(p - c)^{\mathrm{T}}(p - c)}$.
- Mahalanobis distance calculated as: $d_{\Sigma_C}(p, c) = \sqrt{(p - c)^{\mathrm{T}}\Sigma_C^{-1}(p - c)}$, where the distance d is weighted by the covariance matrix of each cluster, Σ_C, using the observations already assigned to that cluster. Unlike Euclidean distance, Mahalanobis distance accounts for the variance and correlation of the features within each cluster, effectively weighting the distance based on the cluster's internal structure. This makes it a potentially more robust and better-suited metric for assigning observations to clusters, especially in cases where each cluster presents a different shape.

3.5.2 Hierarchical Clustering

Hierarchical clustering [57] was also applied to Dataset 1. In contrast to k-means, which partitions data into a fixed number of clusters, hierarchical clustering operates on dissimilarity between every pair of observations in the data, creating a multilevel hierarchy of clusters.

To implement hierarchical clustering, a bottom-up approach, usually referred to as agglomerative clustering, was employed. Agglomerative clustering begins with all the observations that represent each one a cluster. At each step, the iterative algorithm merges the two most similar clusters, with a similarity score based on the combination of a linkage method and a distance metric. The

procedure is repeated until all the clusters are merged into one. The resulting clustering tree is referred to as a dendrogram.

For this analysis, the linkage method chosen is Ward's method [58], which minimizes the inner squared distance between clusters, therefore minimizing the total within-cluster variance when performing merging. As distance metric, the Euclidean distance was employed to measure dissimilarity between observations.

To determine the optimal number of clusters, the same Silhouette approach used in k-means clustering (see Section 3.5.1) was utilized. The dendrogram was then cut at the height corresponding to the optimal number of clusters, k^*, as defined in Eq. 3.36.

In hierarchical clustering, a new observation was assigned to a cluster by manually computing the centroids for each cluster and finding the closest centroid according to a distance metric, following the same approach as in k-means.

3.5.3 Gaussian Mixture Model Clustering

A GMM, also referred to as Normal Mixtures Model [59], is a probabilistic model often employed to perform unsupervised partitioning. A GMM assumes all data points to be generated from a mixture of a finite number of Gaussian distributions with unknown parameters (i.e., mean and covariance).

In the previous two clustering techniques, k-means and hierarchical, each observation belongs to only one cluster (hard assignment). On the contrary, GMM clustering is a soft clustering method that assigns to each observation the probability of belonging to each cluster (soft-clustering). From this, it's possible to perform hard-clustering by maximizing the posterior probability.

The most important advantage of a GMM is that it incorporates the uncertainty quantifying the posterior probability of an observation belonging to a particular cluster, and tries to maximize it in the assignment procedure. Formally, a GMM represents the data distribution as a weighted sum of K Gaussian components:

$$p(x) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x \mid \mu_k, \Sigma_k) \tag{3.37}$$

where K is the number of clusters of the dataset. Each of the k Gaussian distributions is defined by its mean μ_k, and covariance matrix Σ_k, and π_k is the mixing probability that defines the weight of the distributions so that $\Sigma\pi_k = 1$,

and $p(x)$ maintains the properties of a probability distribution function.

The only assumption made in GMM is that the data comes from a mixture of normal distributions. If the features distribution is too skewed, the GMM may not achieve optimal cluster assignments. In our case, to meet this assumption, the original model parameters were used instead of the log-transformed parameters since they are assumed to be normally distributed.

The GMM was fitted to the features extracted from Dataset 1 using the Expectation-Maximization Algorithm [60] that iteratively computes the posterior probability for each cluster using mean, covariance matrices, and mixing proportions obtained via maximum likelihood estimation.

The estimation of mean and covariance matrices in GMM without appropriate regularization techniques and constraints can result in undesirable convergence of the algorithm due to the presence of numerous local maxima [61]. Specifically, the computed covariance matrices may be affected by numerical instabilities, such as preventing matrix inversion. Such issues typically arise when maximum likelihood estimation of GMM covariance matrices is performed without adequate regularization. Moreover, allowing unconstrained covariance matrix structures may lead to ill-conditioned matrices, inhibiting the capabilities of the algorithm to identify useful hidden patterns in the data [62].

In this work, to avoid undesired convergence of the algorithm, a small positive scalar was added to each diagonal in the covariance matrices to ensure the positiveness of all the eigenvalues, hence making the covariance matrix positive definite. Furthermore, to prevent ill-conditioning, the covariance matrices were constrained to be equal in each component (fuzzy k-means clustering [63]).

For GMM, being a probabilistic model that performs soft clustering, the presented Silhouette method was not applicable for determining the best number of clusters. Instead, the number of clusters was arbitrarily chosen based on the best number of clusters of k-means and hierarchical clustering according to Silhouette's method.

Unlike k-means and hierarchical clustering, in GMM clustering, an unseen observation not used during the training process was assigned to the cluster associated with the highest posterior probability, instead of using the nearest centroid approach. Therefore, for GMM clustering assignment, since it exploits the posterior probability, there was no need of the employment of a distance metric, such as squared Euclidean or Mahalanobis.

3.5.4 External Cluster Labelling

As all the clustering techniques employed in this study fall under unsupervised learning, labels or classes are not determined during the training process, in contrast to what happens in supervised learning methods. Nevertheless, cluster labelling is essential to evaluate the clustering results.

In this work, cluster labels were assigned *a posteriori* based on the median macronutrient distribution of the observations within each cluster. Assuming that clusters could be attributed a physiological interpretation, such as representing distinct meal classes, based on these median values, the labelling process followed a ranking approach. Specifically, clusters were ordered according to the medians of their macronutrient distributions, and labels were assigned accordingly.

For instance, in the case of two clusters, the labels "H" and "L" macronutrient classes were applied (e.g., supposing that the macronutrient analysed is protein, the labels reflect HP meals and Low-protein (LP) meals, respectively). The “L” label was assigned to the cluster with the lower median macronutrient value, while the “H” label was assigned to the cluster with the higher median value.

To assess whether the median macronutrient contents of the clusters differed significantly, the Wilcoxon Rank Sum test was applied. This statistical test also informed the selection of the most appropriate macronutrient transformation. By analyzing and comparing the macronutrient distributions of each cluster, the validity of the clustering process was evaluated, checking whether physiologically meaningful labels were correctly assigned to each group.

3.5.5 Unsupervised Cross Validation

Among the unsupervised learning techniques implemented in this study, the most robust method was identified through the development and evaluation of a custom unsupervised CV algorithm. Given the lack of true labelling in unsupervised learning, the classical CV procedure used in supervised learning was specifically adapted to the clustering context and the aim of this analysis.

In this algorithm, the clustering results obtained using all 353 observations from Dataset 1 were used as the reference baseline. The labels assigned to each cluster, and therefore to each of the 353 observations, were treated as the “true” class label, providing a baseline against which the clustering performance on un-

seen data could be evaluated. To obtain a reliable estimate of the performance, a 5-fold CV algorithm was designed. In each fold, 80% of the observations were randomly assigned as the training set for clustering computation, while the remaining 20% formed the test set. After the training on the train set, the observations in the test were assigned to a cluster. The assignment strategy is different for each clustering technique tested. Regarding *k*-means and hierarchical clustering, the squared Euclidean distance metric was used to determine the nearest centroid. Prediction accuracy in each fold was defined as the proportion of test observations that were assigned to the same cluster as in the reference clustering.

Specifically, an observation is considered correctly assigned if the predicted cluster shares the same label as the reference cluster for the corresponding macronutrient. In other words, the predicted cluster shares the same median-based order (rank) as the reference cluster. For example, an observation in the test set of a fold is assigned to the cluster with the highest median fat content, label "H". In this case, the assignment is considered correct, and therefore its prediction, if in the reference clustering, the same observation belongs to the cluster labelled "H", and thus likewise characterized by the highest median fat content.

3.6 Validation on Simulated Meals

The best classification method, developed using Dataset 1, was subsequently applied to Dataset 2 (i.e., to Datasets 2a, 2b, 2c, and 2d) to assess its performance to classify new meals and test in silico whether our methodology can be used to adjust insulin therapy.

3.6.1 Evaluation of Predictive Ability

Prior to classification, observations in Datasets 2a-d were normalized to account for a shift that prevented proper alignment with meals of Dataset 1. Specifically, each feature in Datasets 2a-d was standardized by subtracting the mean and dividing by the standard deviation of that feature, computed from Datasets 2a-d themself. This is due to the mismatch between real and simulated meals, with the latter happening too fast compared to a real scenario since they

are simulated using parameters obtained from a clinical set-up [34]. This is further discussed in the Discussion section of this thesis.

Each observation in Datasets 2a-d was classified with the developed methodology, and the resulting predicted meal class was compared to the true class labels using the various macronutrient transformations. The classification performance was then assessed using confusion matrices, evaluating the agreement between the predicted and actual meal classes of Datasets 2a-d.

3.6.2 Evaluation of Therapy Adjustment

For each observation classified as an HF/HP meal in Datasets 2a-d, DWB insulin strategy was applied. The total insulin bolus was increased by 30%, with 50% of the total insulin being delivered immediately at mealtime as a standard bolus, and the remaining 50% administered as a square wave bolus over the following two hours.

The meals were then re-simulated with the new therapy, and the predicted CGM profiles were compared with the original using standard glycemic outcomes. Comparison was done between the 200 original meals (100 nominal and 100 HF/HP) where standard insulin therapy was used in all meals (indicated as Datasets 2.1a, 2.1b, 2.1c, and 2.1d), and the re-simulated 200 meals where a DWB was administered to meals classified as HF/HP by the developed model, while a standard bolus was applied to meals classified as nominal (indicated as Datasets 2.2a, 2.2b, 2.2c, and 2.2d).

The following CGM metrics [64] were compared between Datasets 2.1a-d and Dataset 2.2a-d using CGM samples from mealtime to the following 8 hours:

- Time in Range (TIR) [%]: Percentage of CGM samples between 70 mg/dl and 180 mg/dl;
- Time Above Range (TAR) [%]: Percentage of CGM samples above 180 mg/dl;
- Time Below Range (TBR) [%]: Percentage of CGM samples below 70 mg/dl;
- Number of Hypoglycemic Events (HE): The number of CGM profiles going below 70 mg/dl at least once in the time window;
- Average glycemia ($\overline{\text{CGM}}$) [mg/dl]: defined as the mean of the measured CGM profile.

Specifically, at first, the CGM metrics of HF/HP meals in Dataset 2.2a-d were compared to those in Dataset 2.1a-d to determine whether the assigned dosing improved postprandial glucose profiles. Secondly, to evaluate whether the assigned therapy for HF/HP meals could restore glucose control to levels comparable to nominal meals, CGM metrics from HF/HP meals in Dataset 2.2a-d were compared to those from nominal meals in Datasets 2.1a-d. These comparisons aimed to determine both the potential efficacy and the improvement in the CGM metrics of the developed methodology for meal classification.

3.7 Evaluation of the Impact of the Model-based Features

To evaluate the contribution of model-informed physiologically-based features in the task of classifying meal types, a comparison was conducted by repeating the model training and validation of the developed methodology using features derived from CGM measurement only. This comparison was designed to quantify the added predictive value of utilizing physiologically-based features in the development of a machine learning technique.

The following set of features was utilized:

- CGM sensor measurement at mealtime, CGM(0min) [mg/dl];
- CGM sensor rate of change at 45 minutes after the meal, $\dot{\text{CGM}}$(45min) [mg/(dl min)];
- CGM sensor rate of change at 120 minutes after the meal, $\dot{\text{CGM}}$(45min) [mg/(dl min)].

These features correspond to the set of CGM-related features extracted without those obtained from the numerical identification of the MI-OMM. To completely remove the dependence on the physiological model, those features were obtained by filtering the CGM signal with a Butterworth low-pass filter with a cut-off frequency of 0.25 the Nyquist frequency, i.e., corresponding to approximately 417 μHz. A visual representation of the CGM filtering outcomes is shown in Fig. 3.6.

The best machine learning method was applied again with the abovementioned set of features, and results were compared with those obtained with the original set of features.

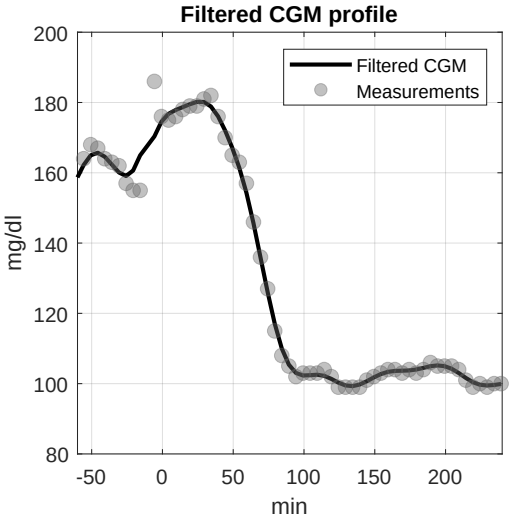

Figure 3.6: Filtered CGM sensor measures using a low-pass Butterworth filter (solid black line) with raw CGM measures (gray markers) overlaid of a representative subject.

4 Results

In this chapter, the results of the study are presented. First, the outcomes of the preliminary analysis are reported, and then, supervised and unsupervised learning techniques are assessed and compared. Second, the results of the validation of the best-performing method on Dataset 2 are described. Finally, the impact of the use of model-based features for the meal classification task is explored.

4.1 Results of the Preliminary Analysis

4.1.1 Spearman's Rank Correlation

Correlation plot obtained from the correlation analysis between features is shown in Fig. 4.1. For the sake of clarity, only interesting blocks are reported. It can be noticed that the AUC_{Ra}s calculated in adjacent periods and the TGRs calculated for adjacent percentages were characterized by a high Spearman's correlation coefficient, ρ (Fig. 4.1, Panels A and B, respectively). In contrast, the correlation obtained between model parameters related to glucose absorption and between CGM-based features was low (Fig. 4.1, Panels C and D, respectively). Furthermore, high negative correlations were detected between the TGRs and the AUC_{Ra}s (Fig. 4.1, Panel E), underlining their mutual influence. Outside of the reported blocks, the only correlations greater than 0.5, in absolute value, were between k_{abs} and AUC_{Ra}(30min), AUC_{Ra}(60min), and AUC_{Ra}(90min) (ρ = 0.79, 0.66, 0.55, respectively), between d and AUC_{Ra}(60min) ($\rho = -0.53$), be-

tween k_{min} and AUC_{Ra}(120min), and TGR(50%) (ρ = 0.53, −0.56, respectively), and between $\dot{G}_i$(120min) and AUC_{Ra}(60min), and AUC_{Ra}(90min) (ρ = −0.51 in both cases).

Regarding the associations between features and macronutrients, the highest Spearman's ρ coefficients were found with macronutrients expressed as absolute amount or the sum of fat and protein content (Fig. 4.2). The most correlated features were the indexes obtained from the MI-OMM, i.e., AUC_{Ra}s and TGRs, with an absolute ρ spanning between 0.14 and 0.45, showing a higher correlation with the carbohydrates ($0.24 < |\rho| < 0.45$) compared to fat, protein, or the sum of their amounts ($0.14 < |\rho| < 0.29$). Outside of this block, the only relevant correlations were between MI-OMM parameters k_{abs}, d, c, and k_{max} with carbohydrates ($|\rho|$ = 0.30, 0.26, 0.23, and 0.21, respectively). Other features showed weak correlations with macronutrient composition, with values not exceeding 0.19.

Among the macronutrient transformations, the worst correlations were found between features and macronutrients expressed as percentages ($0.001 < |\rho| < 0.141$). These results suggested focusing our attention on macronutrients expressed as absolute amounts to classify meal composition, since they showed stronger monotonic relations with the extracted features detected with Spearman's correlation. Even if these relations were weak, one can expect that a more sophisticated methodology can combine and exploit them to perform the classification task successfully.

4.1.2 Repeated Measures Correlation

When applying repeated measures correlation between the extracted features and macronutrient content, the highest correlation coefficients were observed with macronutrients expressed as total amounts in grams (Table 4.1), consistent with the findings obtained using Spearman's rank correlation. Also in this case, AUC_{Ra}s and TGRs exhibited the strongest correlation with carbohydrate intake ($0.26 < |r_{rm}| < 0.48$).

The best correlated features with carbohydrate intake are AUC_{Ra}(60min), AUC_{Ra}(90min), AUC_{Ra}(30min) and AUC_{Ra}(120min) (r_{rm} = -0.47, -0.44, -0.41, −0.38, respectively).

In contrast, weaker correlations were generally observed between the extracted features and the intake of fat and protein ($0.01 < |r_{rm}| < 0.27$). An

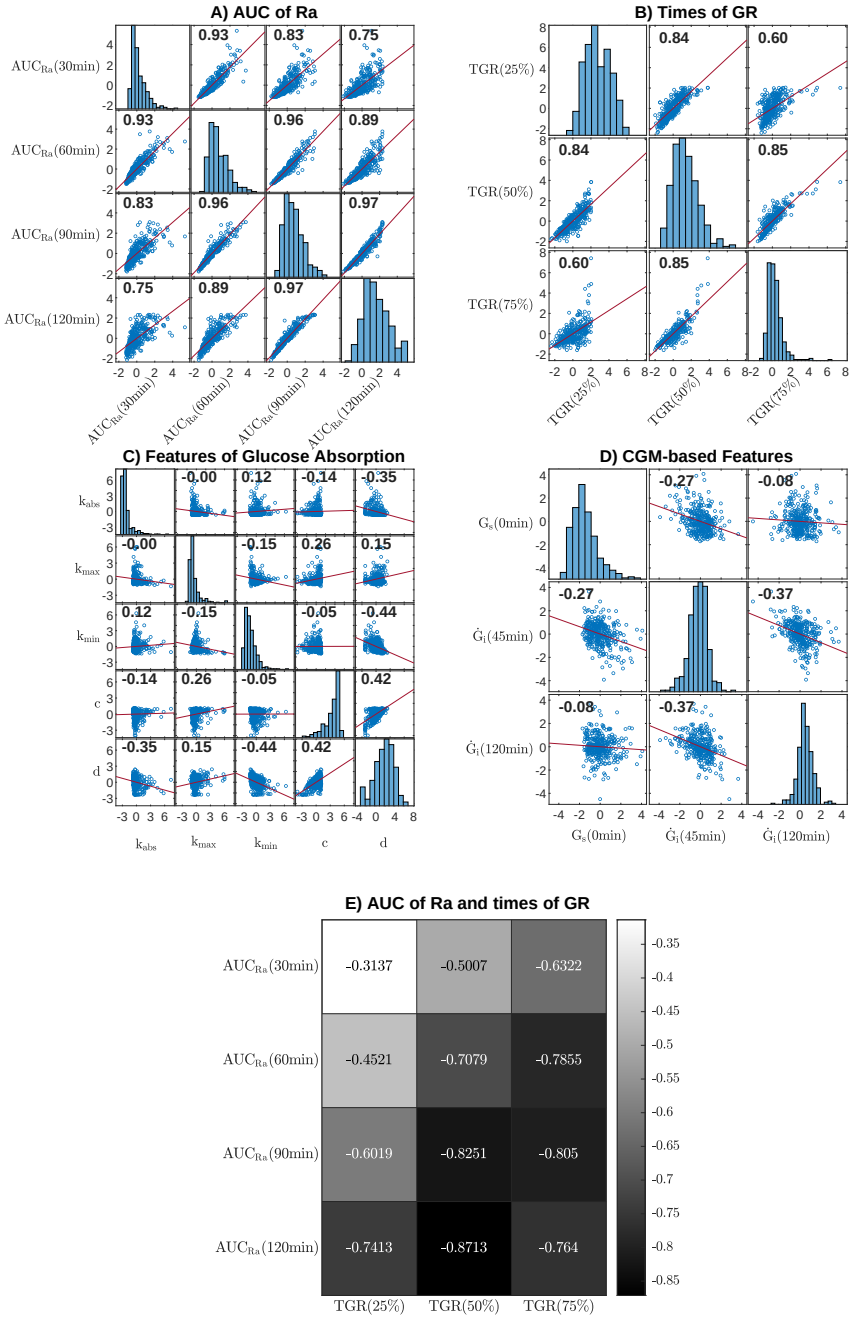

Figure 4.1: Correlation plot of features and their distributions. Numbers in each subplot represent Spearman's correlation, ρ. Panel A: Areas under the rate of glucose appearance curve (AUC_{Ra}s). Panel B: times to reach a certain gastric retention (TGRs). Panel C: MI-OMM parameters describing glucose absorption. Panel D: CGM-based features. Panel E: Correlation between TGRs and AUC_{Ra}s.

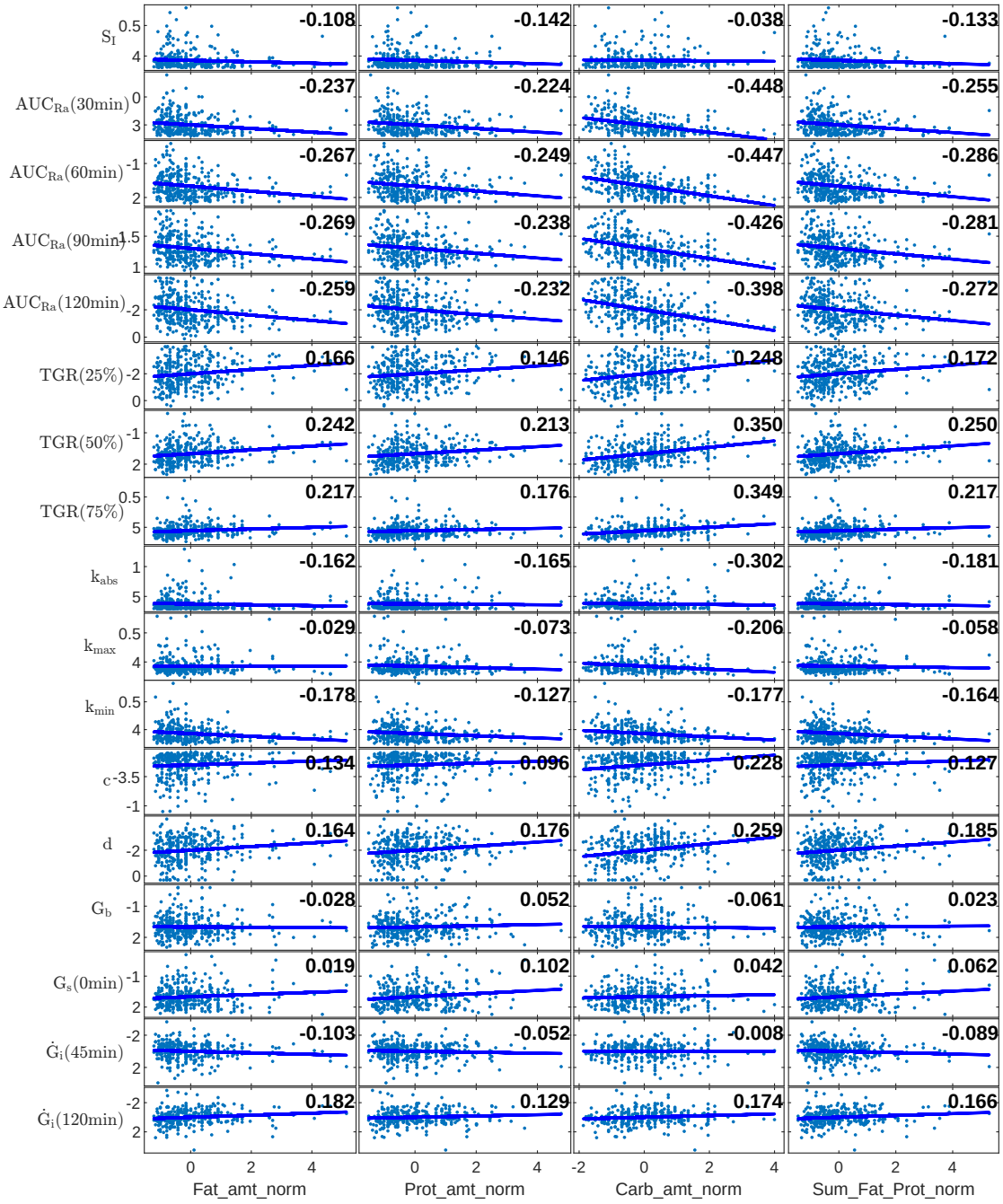

Figure 4.2: Correlation plot of extracted features (excluding demographics) and macronutrient transformations. Numbers in each subplot represent Spearman's ρ correlation. Fat_amt_norm = normalized absolute fat amount. Prot_amt_norm = normalized absolute protein amount. Carb_amt_norm = normalized absolute carbohydrate amount. Sum_fat_prot_norm = normalized sum of absolute fat and protein amounts.

exception to this trend was identified for the interstitial glucose rate of change at 45 minutes, $\dot{G}_i$(45min). This feature demonstrated higher absolute correlation with fat and protein intake ($r_{rm} = -0.24$ and -0.26, respectively) compared to carbohydrates ($r_{rm} = -0.15$). The negative sign indicates that fat and protein reduce the glucose rate of change around 45 min after the meal, likely caused by the slower absorption. This feature was also the one presenting the highest correlation with fat and protein amounts overall, suggesting that $\dot{G}_i$(45min) may serve as a relevant feature for meal classification tasks.

Table 4.1: Repeated measures correlation coefficients between extracted features, excluding demographics, and macronutrient expressed as their amount. Features with a r_{rm} above 0.15, in absolute value, are highlighted.

	Fat_amt_norm	Prot_amt_norm	Carb_amt_norm	Sum_Fat_Prot_norm
S_I	0.122	0.1531	0.1183	0.1451
AUC_{Ra}(30min)	-0.0673	-0.0795	-0.4105	-0.0777
AUC_{Ra}(60min)	-0.1316	-0.1281	-0.4741	-0.1392
AUC_{Ra}(90min)	-0.1233	-0.0939	-0.4357	-0.1183
AUC_{Ra}(120min)	-0.0906	-0.0449	-0.3803	-0.0756
TGR(50%)	0.1139	0.0632	0.2695	0.0982
TGR(50%)	0.1539	0.1074	0.3163	0.1429
TGR(75%)	0.1079	0.0587	0.27	0.0925
k_{abs}	-0.014	-0.0384	-0.0408	-0.0264
k_{max}	0.0302	-0.011	-0.1304	0.013
k_{min}	-0.148	-0.0922	-0.2283	-0.1323
c	0.0715	0.0679	0.255	0.0748
d	0.1101	0.1163	0.2638	0.1207
G_b	-0.0509	-0.0165	-0.1165	-0.0384
G_s(0min)	0.0503	0.1329	-0.0644	0.0925
$\dot{G}_i$(45min)	-0.2418	-0.2609	-0.145	-0.2678
$\dot{G}_i$(120min)	0.1763	0.1648	0.203	0.1833

The use of the repeated measures correlation approach provides valuable insight into within-subject variability in feature-macronutrient relationships. As explained in Section 3.3.2, unlike traditional correlation metrics, it enables the visualization of subject-specific associations that may not be evident in aggregated data, therefore offering a better understanding of underlying physiological responses to macronutrient intake. This is at the cost of detecting only linear relationships, whereas with Spearman's correlation, we could relax this assumption by requesting only the monotonicity between variables.

Below, we report the outcomes of this method for the most correlated features. Fig. 4.3 illustrates the relationship between $\dot{G}_i$(45min) and the amount of macronutrient intake. A clear negative association is evident between this

feature and the amount of fat (Panel A), protein (Panel B), and their sum (Panel D). A similar, but weaker, negative trend is observed for carbohydrate intake (Panel C), suggesting a smaller influence. On the other hand, Fig. 4.4 displays the correlation between $\dot{G}_i$(120min) and the amount of macronutrients, showing a positive correlation with all macronutrient transformations. As the total amount of fat, protein, and carbohydrate intake increases, the glucose derivative also increases 2 hours after the meal. These findings are consistent with those reported by Ekhlaspour and coworkers [46], confirming the potential relevance of the glucose rate of change in relation to meal composition, specifically by underlining a delay in the glycemic response in the presence of HF/HP meals.

The analysis of S_I in relation to macronutrient intake (Fig. 4.5) reveals a positive association with the content of fat, protein, and carbohydrates. The relations between macronutrients and TGR(50%) are shown in Fig. 4.6. An increase in the total macronutrient content corresponds to a delay in the time to halve GR, indicating a slower gastric emptying process in correspondence with meals with higher content of fat and protein. A similar trend is observed for AUC_{Ra}(60min), reported in Fig. 4.7. Meals with higher content of fat and protein are associated with a reduced AUC of Ra at 60 minutes, indicating a lag in glucose absorption caused by complex meals.

All these findings regarding the delay in the gastric emptying, the lag in the absorption of glucose, and the influence of the insulin sensitivity of HF/HP meals are consistent with what is reported in the literature [27, 32, 65].

4.2 Assessment of Supervised Learning Techniques

4.2.1 Performance of Linear Mixed Effects Modeling

As explained in Section 3.4.1, the aim of this analysis is to find which macronutrient transformation related to fat and protein intake is best explained by the extracted features, excluding demographics. The two metrics tested were ΔLL and ΔSSE.

A total of $2^{15} - 1 = 32767$ LME models for each macronutrient transformation of increasing complexity were exhaustively tested for each LME model type implemented (i.e., random intercept only, and random slope and intercept). The best model was chosen according to the best AIC for each macronutrient transformation.

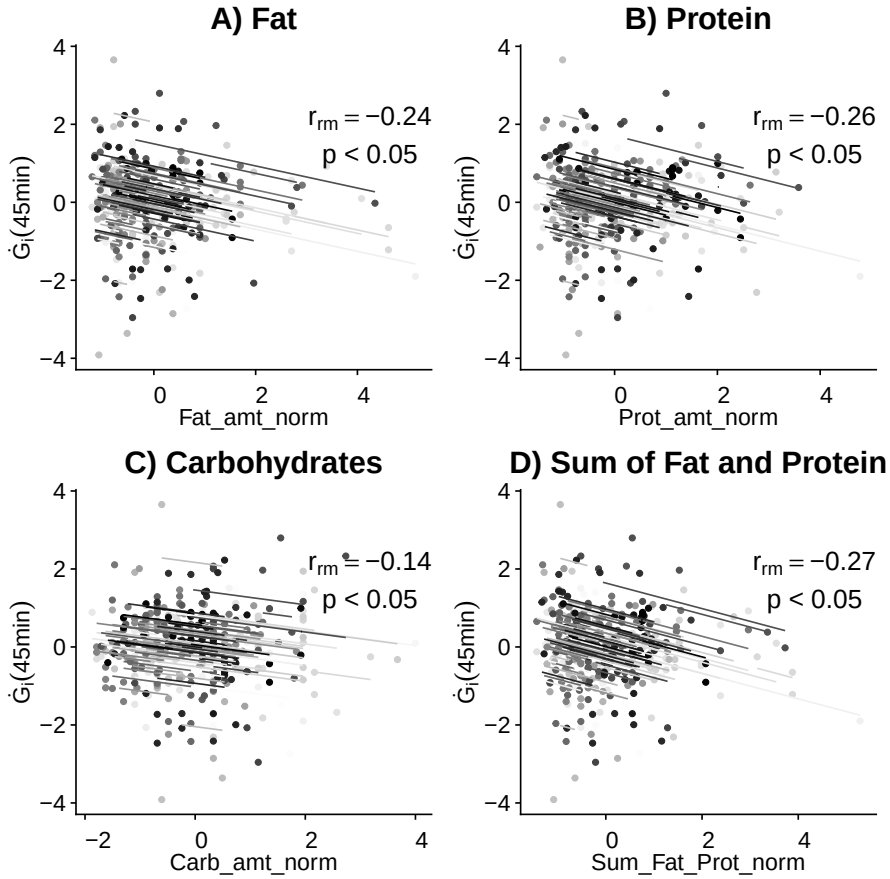

Figure 4.3: Repeated measures correlation between interstitial glucose rate of change at 45 minutes, $\dot{G}_i(45min)$, and macronutrient amounts. Panel A: Normalized amount of fat (Fat_amt_norm). Panel B: Normalized amount of protein (Prot_amt_norm). Panel C: Normalized amount of carbohydrates (Carb_amt_norm). Panel D: Normalized sum of fat and protein amount (Sum_Fat_Prot_norm). Gray solid lines represent subject-specific regression lines with common slope, whereas gray dots represent individual meal observations. The repeated measures correlation coefficient (r_{rm}) and p-value (p) are reported for each macronutrient.

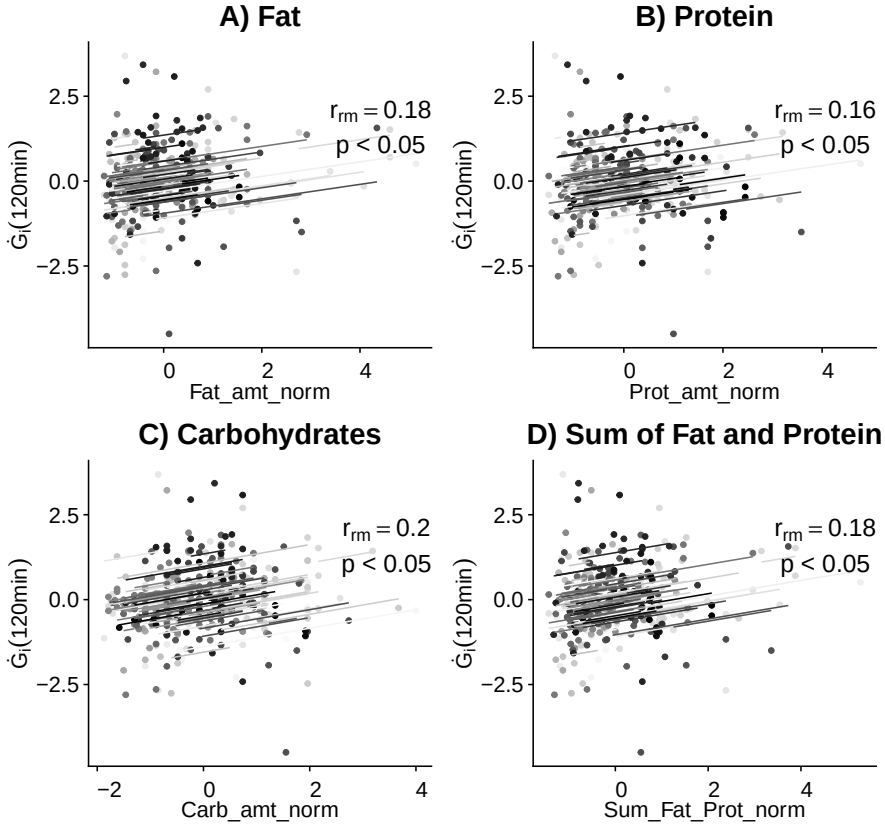

Figure 4.4: Repeated measures correlation between interstitial glucose rate of change at 120 minutes, $\dot{G}_i$(120min), and macronutrient amounts. Panel A: Normalized amount of fat (Fat_amt_norm). Panel B: Normalized amount of protein (Prot_amt_norm). Panel C: Normalized amount of carbohydrates (Carb_amt_norm). Panel D: Normalized sum of fat and protein amount (Sum_Fat_Prot_norm). Gray solid lines represent subject-specific regression lines with common slope, whereas gray dots represent individual meal observations. The repeated measures correlation coefficient (r_{rm}) and p-value (p) are reported for each macronutrient.

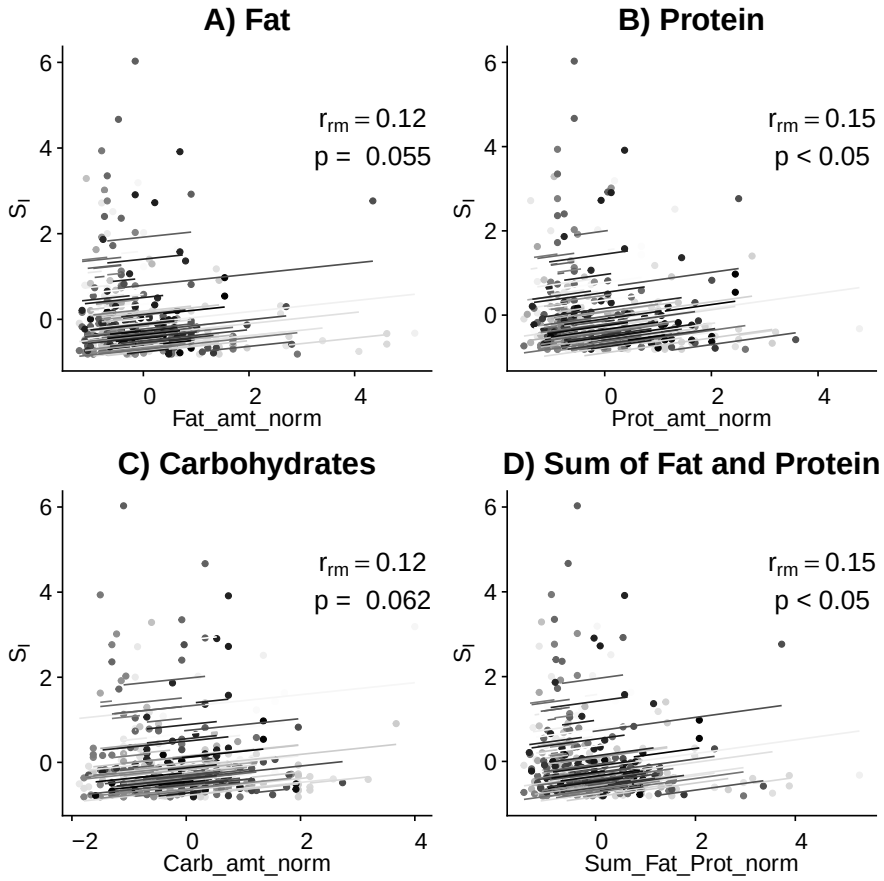

Figure 4.5: Repeated measures correlation between insulin sensitivity, S_I, and macronutrient amounts. Panel A: Normalized amount of fat (Fat_amt_norm). Panel B: Normalized amount of protein (Prot_amt_norm). Panel C: Normalized amount of carbohydrates (Carb_amt_norm). Panel D: Normalized sum of fat and protein amount (Sum_Fat_Prot_norm). Gray solid lines represent subject-specific regression lines with common slope, whereas gray dots represent individual meal observations. The repeated measures correlation coefficient (r_{rm}) and p-value (p) are reported for each macronutrient..

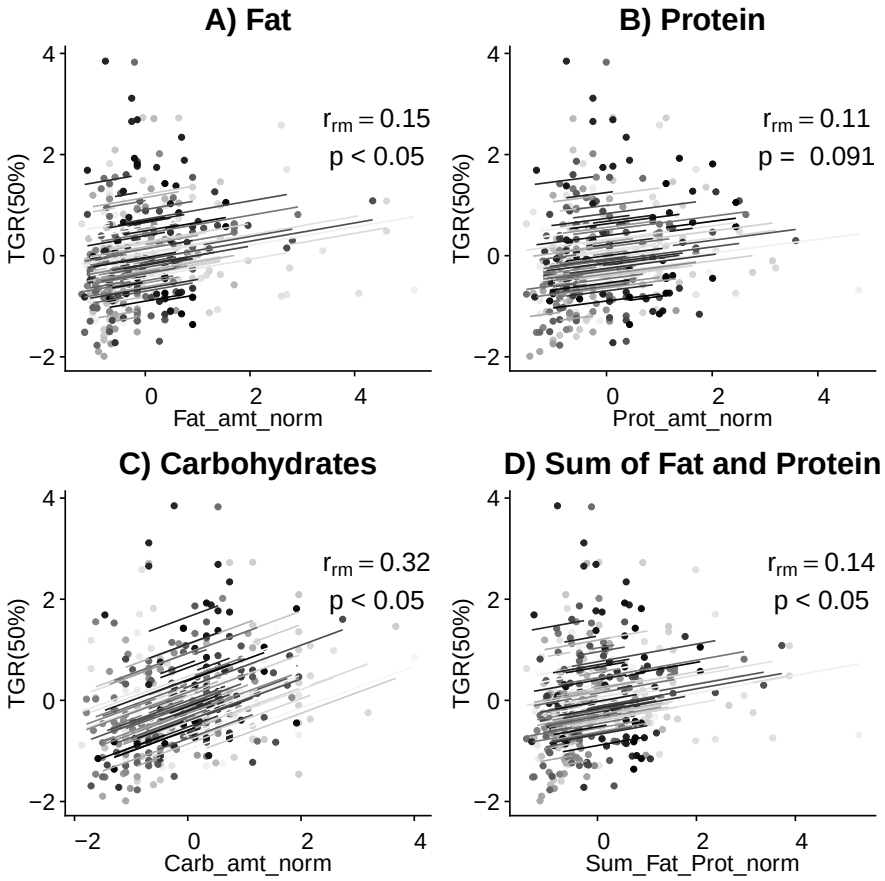

Figure 4.6: Repeated measures correlation between half-life of GR curve, TGR(50%) and macronutrient amounts. Panel A: Normalized amount of fat (Fat_amt_norm). Panel B: Normalized amount of protein (Prot_amt_norm). Panel C: Normalized amount of carbohydrates (Carb_amt_norm). Panel D: Normalized sum of fat and protein amount (Sum_Fat_Prot_norm). Gray solid lines represent subject-specific regression lines with common slope, whereas gray dots represent individual meal observations. The repeated measures correlation coefficient (r_{rm}) and p-value (p) are reported for each macronutrient.

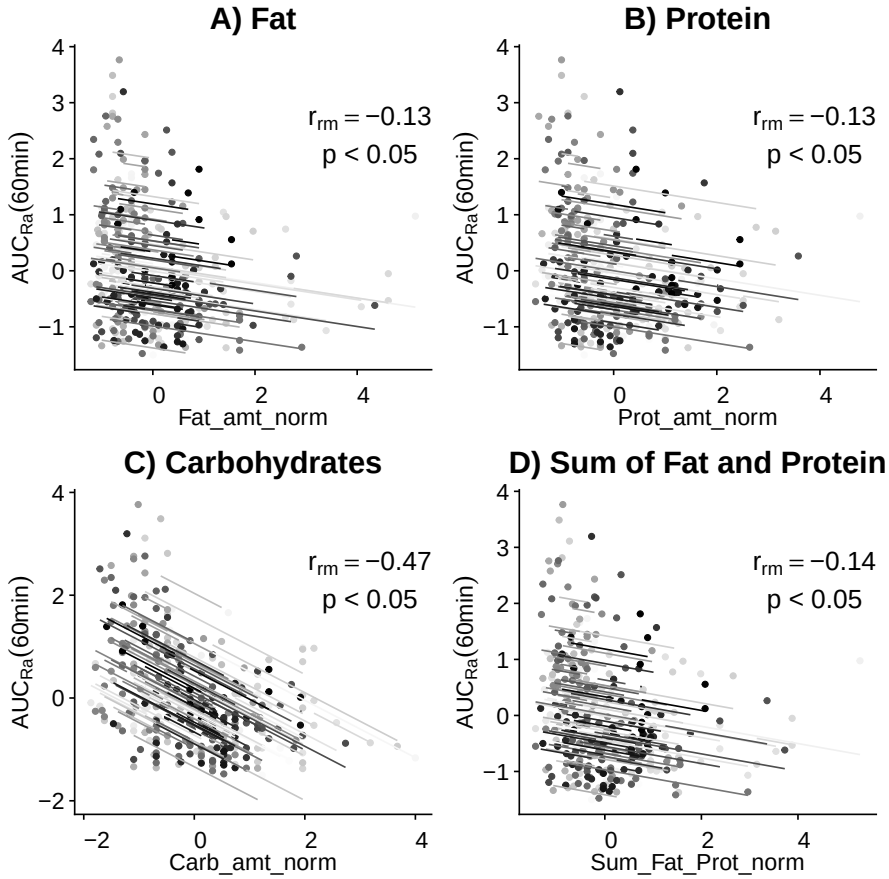

Figure 4.7: Repeated measures correlation between normalized area under the curve of glucose rate of appearance at 60 min, $AUC_{Ra}(60min)$, and macronutrient amounts. Panel A: Normalized amount of fat (Fat_amt_norm). Panel B: Normalized amount of protein (Prot_amt_norm). Panel C: Normalized amount of carbohydrates (Carb_amt_norm). Panel D: Normalized sum of fat and protein amount (Sum_Fat_Prot_norm). Gray solid lines represent subject-specific regression lines with common slope, whereas gray dots represent individual meal observations. The repeated measures correlation coefficient (r_{rm}) and p-value (p) are reported for each macronutrient.

Based on the ΔLL metric, the macronutrient transformations that is best predicted by the extracted are the ones expressed as their total amount (fat, protein and sum of fat and protein amounts). These findings are consistent with the preliminary correlation analysis, which showed stronger associations between the extracted features and the macronutrients when expressed as absolute amounts rather than percentages.

In contrast, according to the ΔSSE metric, the macronutrient transformations expressed as percentages perform better. In general, the ΔSSE metric favors transformations that reduce raw prediction errors (absolute residuals) at the cost of unmet assumptions on parameter variability and residual distribution. The ΔSSE, in other words, quantifies the reduction in the prediction error relative to the null model, and it can be helpful in depicting the performance improvements in relative prediction terms of the best model over the null one.

It is worth noting that macronutrients expressed as percentages, despite their high ΔSSE values, exhibited the lowest correlations in the preliminary analysis, both according to Spearman's rank correlation and to repeated measures correlation analysis. This may suggest that the ΔLL metric, by requiring proper modeling of parameter and residual variance, provides a more complete overview of model performance. Therefore, in our analysis, we deemed ΔLL a more suitable metric for evaluating the tested LME models. The ΔSSE metric is reported to facilitate interpretation, as it provides an indication of the predictive performance of the models.

Assessment of Linear Mixed Effect Models with Random Intercept

The results of the analysis are summarized in Table 4.2. Based on the ΔLL metric, the macronutrient transformation that is best predicted by the extracted features is the amount of protein, followed by the sum of fat and protein. For fat, the transformation with the highest performance is also its absolute amount.

The subset of features employed by the best LME model with random intercept is listed in Table 4.2, as well, for each tested macronutrient transformation. As expected, the AUC_{Ra}s appear frequently in the best models, both for protein and fat transformations, whereas the TGRs are mostly associated with transformations employing proteins.

In contrast, fat-related transformations tended to include features such as k_{min} and glucose rate of changes $\dot{G}_i$(45min) and $\dot{G}_i$(120min). Surprisingly, de-

spite their high r_{rm} coefficients, these features were selected only by models describing protein percentage, even if the LME models should stratify for subject occasions similar to repeated measures correlation.

Table 4.2: Summary of the performance of the best LME models with random intercept according to AIC for each macronutrient transformation. $-2 \cdot$LL = -2 Loglikelihood; $-2 \cdot \Delta$LL = -2 delta loglikelihood compared with the null model; SSE = sum of squares error; ΔSSE = delta sum of squares error compared to the null model. Features of the best LME models with random intercept for each macronutrient transformation are reported. Macronutrients are ranked by the best ΔLL.

	$-2 \cdot$**LL**	$-2 \cdot \Delta$**LL**	**SSE**	Δ**SSE**	**Best Model Features**
Prot_amt_norm	913.038	28.816	158.572	10.345	G_s(0min); AUC_{Ra}(120min); AUC_{Ra}(60min); TGR(50%); TGR(75%)
Sum_Fat_Prot_norm	923.402	28.454	171.246	12.135	G_s(0min); $\dot{G}_i$(120min); AUC_{Ra}(120min); AUC_{Ra}(60min); TGR(50%); TGR(75%)
Prot_kcal_perc_norm	970.15	22.842	254.679	21.344	G_b; G_s(0min); AUC_{Ra}(120min); TGR(25%); TGR(50%); TGR(75%); k_{max}
Prot_perc_norm	972.786	21.352	245.561	35.816	G_b; G_s(0min); $\dot{G}_i$(120min); AUC_{Ra}(120min); TGR(50%); c
Fat_amt_norm	944.732	19.84	194.273	8.815	G_s(0min); $\dot{G}_i$(120min); k_{min}
Sum_perc_Fat_Prot_norm	972.696	19.66	242.805	32.352	$\dot{G}_i$(45min); AUC_{Ra}(120min); S_I; TGR(50%); c
Fat_kcal_perc_norm	973.222	15.88	241.84	22.874	$\dot{G}_i$(120min); $\dot{G}_i$(45min); AUC_{Ra}(60min); k_{min}
Sum_kcal_perc_Fat_Prot_norm	978.82	15.15	250.631	32.015	$\dot{G}_i$(120min); $\dot{G}_i$(45min); AUC_{Ra}(60min)
Fat_perc_norm	971.458	14.75	234.438	20.991	$\dot{G}_i$(120min); $\dot{G}_i$(45min); AUC_{Ra}(60min); k_{min}

When assessing the overall predictive power of the LME models in estimating meal class or macronutrient composition, using the macronutrient transformations as continuous outcome variables, it is evident from the ΔSSE metric that the improvement over the null model is relatively low. Even in the best case, the best-performing LME model according to the ΔSSE, achieves an improvement of less than 13%, highlighting the limited marginal gain offered by including the full set of features in the model selection (after performing feature selection), and the limitation of LME models in performing the meal classification task.

The observation–prediction and residual plots are presented in Fig. 4.8,

where the performance of the best LME models is compared against the corresponding null models. For now, the contributions of the random effects were considered to calculate model predictions. The models take into account inter-individual variability by allowing subject-specific deviations from the population-level intercept. Despite accounting for the contribution of the random effects, the observation–prediction plots (Panels A and B) demonstrate only minimal improvement relative to the null model. This observation is further supported by the residual distributions (Panels C and D), which do not exhibit a tendency to contract towards zero but instead closely resemble the distribution observed under the null model.

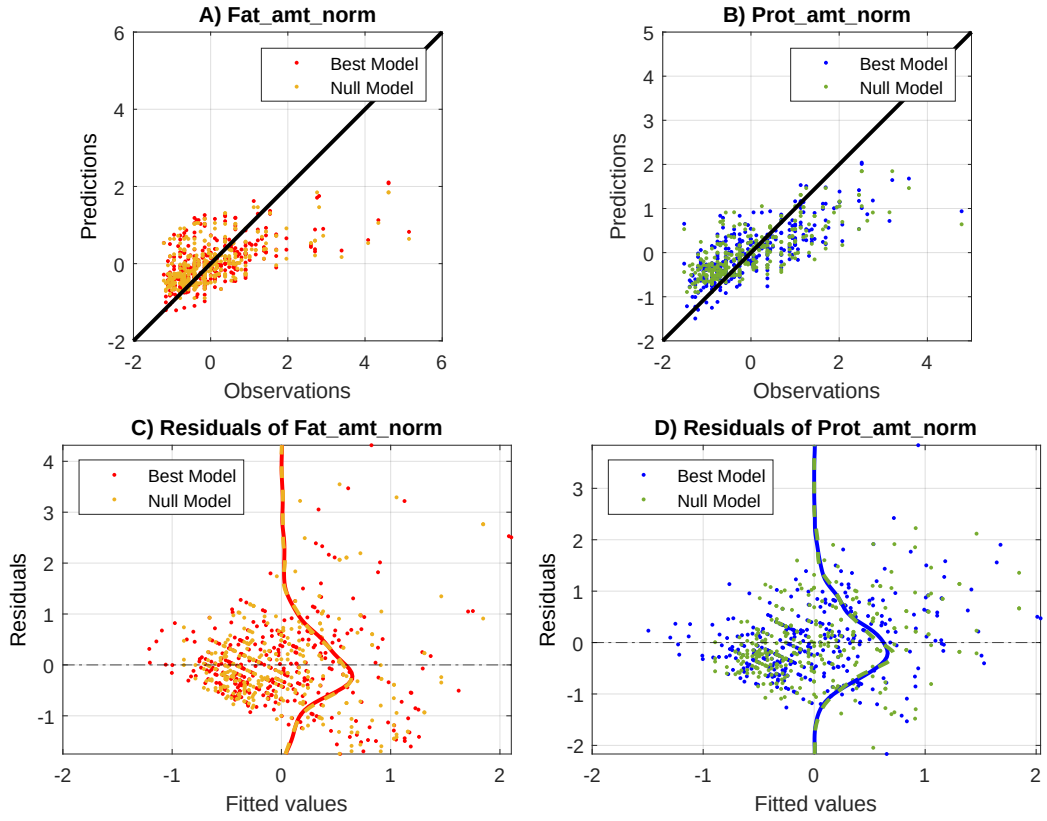

Figure 4.8: Observation-prediction plots and residual plots for the best LME models with random intercept of the best (red markers) and null (orange markers) model for fat and of the best (blue markers) and null (green markers) model for protein amount. Panel A: observation-prediction plot for fat amount. Panel B: observation-prediction plot for protein amount. Panel C: residuals and their distribution for the best (red solid line) and null (orange dashed line) model for fat amount. Panel D: residuals and their distribution for the best (blue solid line) and null (green dashed line) model for protein amount. Random effects were included to calculate model predictions.

Assessment of Linear Mixed Effect Models with Random Slope and Intercept

The performance metrics of the best LME models incorporating random slopes are presented in Table 4.3. Consistent with the findings from LME models including only random intercepts, and in line with the preliminary analysis, the most predictive macronutrient transformations are those expressed as absolute amounts of fat, protein, and the sum of the two. Unlike the previous analysis, which assessed LME models with only random intercepts, in this case, the ΔLL and ΔSSE metrics show strong concordance in defining the best macronutrient transformations.

As expected, allowing random slopes, therefore assuming that fat and protein have an individualized effect on each subject, led to substantial increases (in absolute value) in both ΔLL and ΔSSE.

Focusing on the ΔLL, which was used to choose the best macronutrient transformation, we observed that, for fat amount and the sum of fat and protein amount, more than doubled compared to their respective LME model with only random intercept, reflecting a meaningful improvement in model fit. In general, the SSE of the best LME models according to ΔLL for fat, protein, and their sum nearly halved, indicating that including random slopes significantly reduces the raw prediction error, thereby enhancing predictive performance.

Surprisingly, the best LME models of macronutrients expressed as percentages (whether by amount or caloric contribution) did not improve in terms of ΔLL compared to the respective models with only random intercept. A similar pattern was observed for ΔSSE, which also decreased for these macronutrient transformations, with the exception of the combined percentage of fat and protein. These findings further support the limited predictive utility of percentage-based macronutrient transformations in the task of meal classification.

Table 4.3 reports the features included in the best-performing LME models with random slopes for each macronutrient transformation. The number of selected features reduced when compared to the cases with only random intercept, ranging from a maximum of three for models related to macronutrient expressed by their total amounts, to as few as one for percentage-based macronutrient transformations. This reduction in model complexity for percentage-based transformations may partially explain the observed lack of improvement in predictive performance, as part of the variance of the outcome is now explained by

Table 4.3: Summary of the performance of the best LME models with random slope according to AIC for each macronutrient transformation. $-2 \cdot$LL = -2 Loglikelihood; $-2 \cdot \Delta$LL = -2 delta loglikelihood compared with the null model; SSE = sum of squares error; ΔSSE = delta sum of squares error compared to the null model. Features of the best LME models with random slope and intercept for each macronutrient transformation are reported. Macronutrients are ranked by the best ΔLL

	$-2 \cdot$**LL**	$-2 \cdot \Delta$**LL**	**SSE**	Δ**SSE**	**Best Model Features**
Fat_amt_norm	890.642	73.928	85.470	117.618	G_s(0min); AUC_{Ra}(120min); TGR(50%)
Sum_Fat_Prot_norm	886.078	65.778	77.678	105.704	G_s(0min); AUC_{Ra}(120min); TGR(50%)
Prot_amt_norm	892.424	49.430	84.971	83.946	G_s(0min); AUC_{Ra}(120min); TGR(50%)
Sum_perc_Fat_Prot_norm	975.934	16.422	221.642	53.514	$\dot{G}_i$(45min); AUC_{Ra}(120min)
Prot_perc_norm	984.348	9.790	264.418	16.959	AUC_{Ra}(60min)
Sum_kcal_perc_Fat_norm	984.712	9.256	261.103	21.543	$\dot{G}_i$(45min)
Fat_kcal_perc_norm	980.712	8.390	247.048	17.666	$\dot{G}_i$(45min)
Fat_perc_norm	978.236	7.972	240.942	14.486	$\dot{G}_i$(45min)
Prot_kcal_perc_norm	985.382	7.610	264.747	11.276	G_b

the random effects that include slopes and intercept. Thus, requiring a fewer amount of predictors to reach the best AIC during model selection process.

Among the features selected by the best LME models of the best macronutrient transformations, the AUC of Ra at 120 minutes (AUC_{Ra}(120min)) and the half-life of the GR curve (TGR(50%)) consistently emerged as selected features, underlining their relevance as identified in the preliminary analysis. Interestingly, the glycemia at mealtime (G_s(0min)) was also selected through this analysis.

The performances of LME models incorporating random slopes for the amounts of fat and protein is illustrated in Fig. 4.9. As done previously, random effects were included. For the protein amount, the observation-prediction plot (Panel B) and residuals distribution (Panel D) indicate a slight improvement in distribution compared to the null model, with a slight trend toward residuals centering more closely around zero. In contrast, for the fat amount (Panel A, C), the best LME model with random slopes does not demonstrate a clear enhancement in residual distribution relative to the null model.

Evaluation of Predictive Performances

Including random effects in predictions is not appropriate when assessing the suitability of LME models for the meal classification task, since random effects capture subject-specific deviations from the population-level estimates. When random effects are applied to predictions, the information about individual subjects that wouldn't be available in a population-level prediction, is exploited. For evaluating predictive performance at the population level, only the fixed effects should be used. Thus, checking the LME models employability for the objective of this thesis, so for the meal macronutrient inferring, requires solely the utilization of the fixed effects.

The predictive performance of the best LME models for the amounts of fat and protein is illustrated in Fig. 4.10 and 4.11, related, respectively, to LME models with random intercept, and random intercept and slope. Here the random effects, as individual corrections from the fixed effects, were not included to calculate model prediction, to evaluate the predictive capability of the LME models at a population level.

Starting from the assessment of predictive capability of LME model with random intercept, the inefficacy of the application of LME model at a population

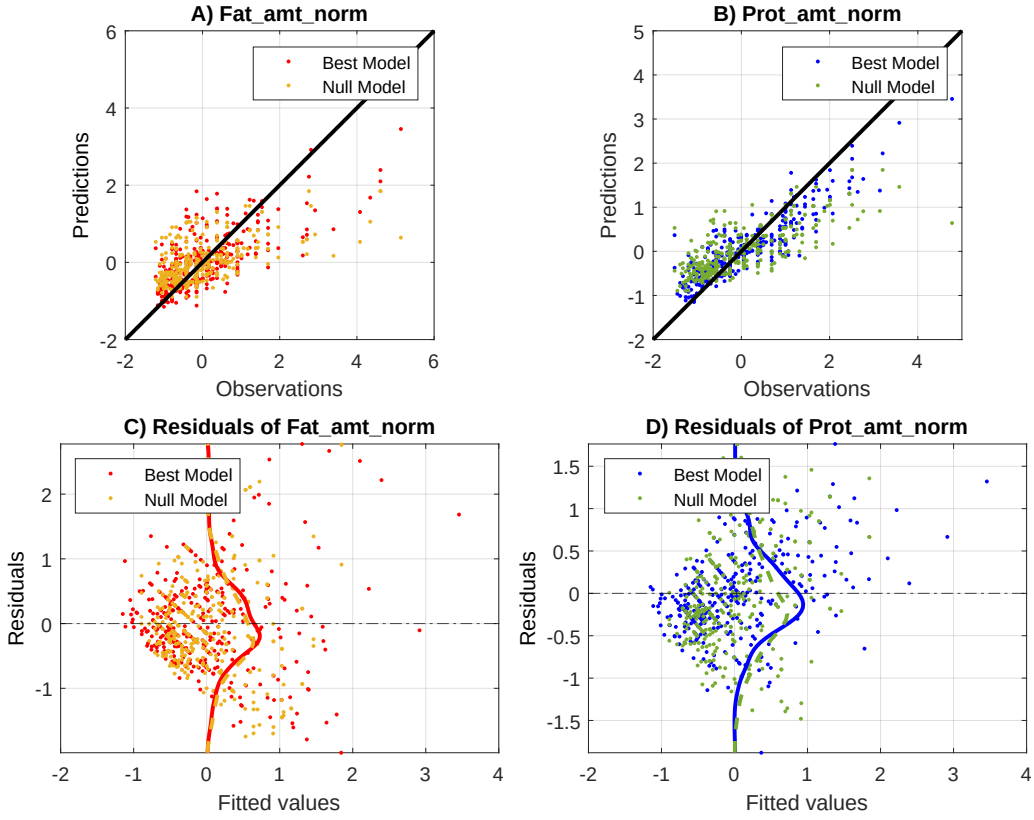

Figure 4.9: Observation-prediction plots and residual plots for the best LME models with random intercept and slope of the best (red markers) and null (orange markers) model for fat and of the best (blue markers) and null (green markers) model for protein amount. Panel A: observation-prediction plot for fat amount. Panel B: observation-prediction plot for protein amount. Panel C: residuals and their distribution for the best (red solid line) and null (orange dashed line) model for fat amount. Panel D: residuals and their distribution for the best (blue solid line) and null (green dashed line) model for protein amount. Random effects were excluded to calculate model predictions.

level is clear after an examination of the residual plots of Fig. 4.10. Specifically, the residuals (Panels B, D) for both macronutrients, despite meeting the assumptions of an LME model, do not converge toward zero and display distributions that remain dispersed. The improvement in model performance relative to the null model is minimal, highlighting the limitations of LME models with only random intercepts in effectively performing meal classification based on fat and protein content.

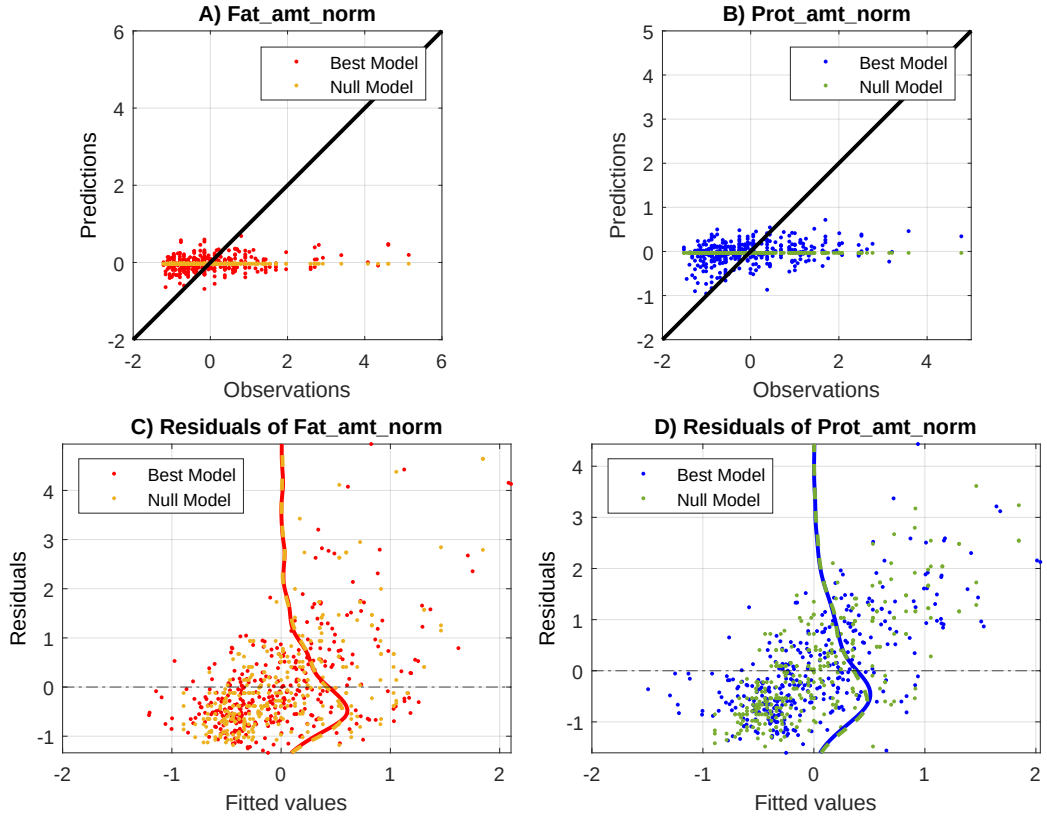

Figure 4.10: Observation-prediction plots and residual plots for the best LME models with random intercept of the best (red markers) and null (orange markers) model for fat and of the best (blue markers) and null (green markers) model for protein amount. Panel A: observation-prediction plot for fat amount. Panel B: observation-prediction plot for protein amount. Panel C: residuals and their distribution for the best (red solid line) and null (orange dashed line) model for fat amount. Panel D: residuals and their distribution for the best (blue solid line) and null (green dashed line) model for protein amount. Random effects were excluded to calculate model predictions.

The predictive capability of LME model with random intercept and slope for fat and protein amount are shown in Fig. 4.11. Similar conclusion can be drawn,

as in the observation–prediction plots (Panels A, B), comparing the null model with the best models, there is no substantial improvement in predictive capabilities. This observation is further supported by the residual plots (Panels C, D), which show no notable change in the distribution of residuals, despite meeting the statistical assumption. Specifically, the residuals and their distributions do not appear to shrink toward zero, indicating the limited benefit of employing LME models in the meal classification task at a population level.

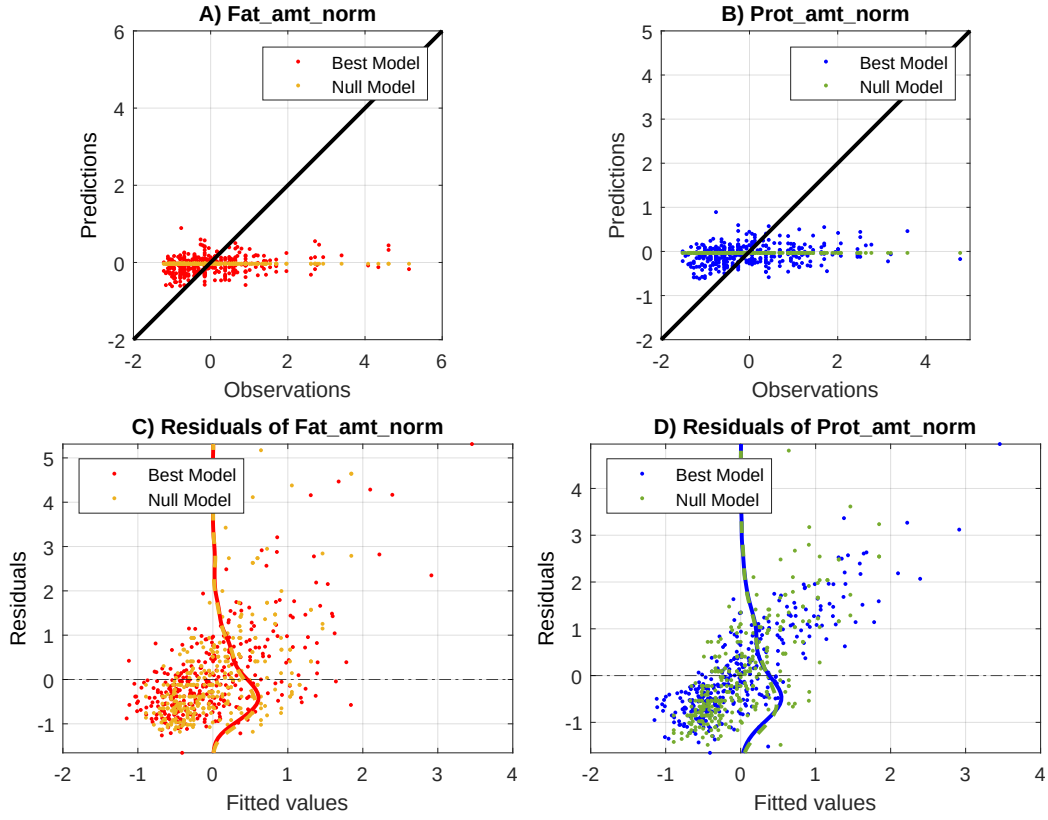

Figure 4.11: Observation-prediction plots and residual plots for the best LME models with random intercept and slope of the best (red markers) and null (orange markers) model for fat and of the best (blue markers) and null (green markers) model for protein amount. Panel A: observation-prediction plot for fat amount. Panel B: observation-prediction plot for protein amount. Panel C: residuals and their distribution for the best (red solid line) and null (orange dashed line) model for fat amount. Panel D: residuals and their distribution for the best (blue solid line) and null (green dashed line) model for protein amount. Random effects were excluded to calculate model predictions.

4.2.2 Logistic Regression

In Table 4.4, the prediction accuracies obtained for the best logistic model for each macronutrient transformation and threshold selection strategy are reported. As expected, when using the threshold derived from the ROC curve by fixing the FPR at 5%, the resulting prediction accuracies are consistently lower compared to those achieved using the optimal operating point of the ROC curve.

Table 4.4: Predictive accuracy of the best logistic model for each macronutrient transformation for the three different threshold selection strategies: maximum allowable false positive rate (FPR 5%), optimal operating point (OOP), and highest F1-score (Max F1-score)

	FPR 5%	**OOP**	**Max F1-score**
Fat_amt_norm	0.61	0.62	0.45
Prot_amt_norm	0.59	0.60	0.40
Sum_Fat_Prot_norm	0.62	0.69	0.68
Fat_perc_norm	0.53	0.65	0.65
Prot_perc_norm	0.53	0.55	0.49
Sum_perc_Fat_Prot_norm	0.53	0.58	0.51
Fat_perc_kcal_norm	0.46	0.57	0.56
Prot_perc_kcal_norm	0.52	0.54	0.49
Sum_perc_kcal_fat_prot_norm	0.45	0.55	0.54

Among all transformations, the sum of fat and protein amount demonstrated the highest predictive performance across all threshold strategies. The confusion matrices for the logistic model of the sum of fat and protein amount are shown in Fig. 4.12. Using the optimal ROC operating point, this macronutrient transformation achieved an accuracy of 69% (Panel A). It also performed the best among all the other macronutrient transformations when the threshold was set at FPR = 5%, obtaining 62% accuracy (Panel B) and when using the threshold that maximizes the F1-score, achieving 68% accuracy (Panel C). These results confirm the potential capabilities of this macronutrient transformation for meal classification tasks.

On the other hand, the worst predictive accuracies were observed when macronutrients were expressed as a percentage. In no case, these transformations exceeded 60% accuracy, regardless of the threshold strategy. The only exception is the percentage of fat, which reached 65% accuracy under both the optimal ROC threshold and the F1-maximizing threshold. This finding aligns

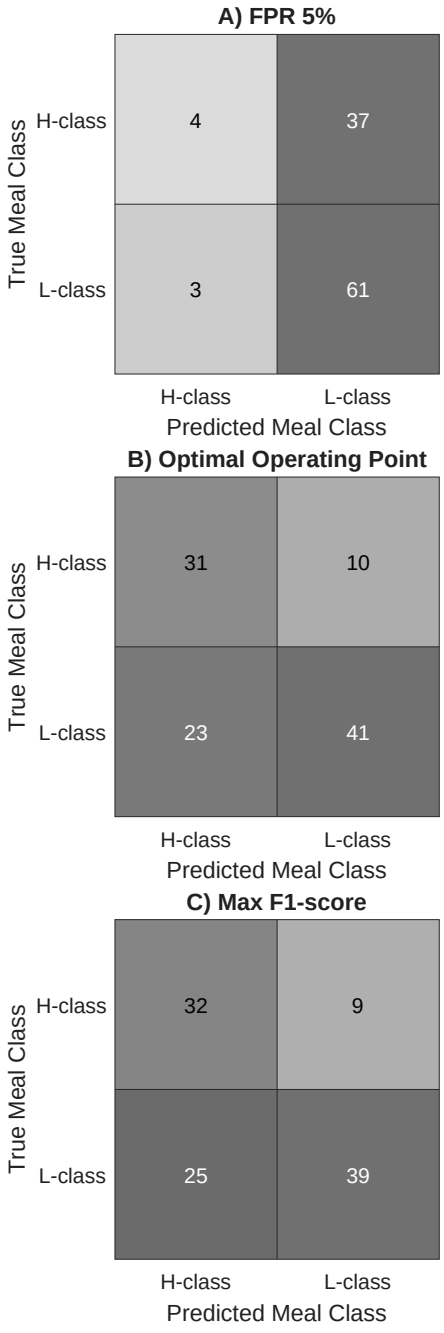

Figure 4.12: Confusion matrices of the best logistic regression model for the sum of fat and protein amount for the three different threshold strategies. Panel A: maximum acceptable false positive rate fixed at 5% (FPR 5%). Panel B: optimal operating point of the ROC. Panel C: maximum F1-score of the prediction-recall curve (Max F1-score).

with results coming from the LME models and the preliminary analysis, which also suggested that percentage-based macronutrient representations are potentially less informative for meal classification purposes.

Table 4.5 reports the area under the ROC and precision-recall curves for each macronutrient transformation. For the sum of fat and protein amount, that revealed to be the best macronutrient transformation in terms of predictive accuracy, the ROC and prediction-recall curve are depicted in Fig. 4.13. The sum of fat and protein amount not only showed the best classification accuracy but also achieved the highest area under the ROC (Panel A). However, the precision-recall curve did not show the expected performance (Panel B), with a relatively low AUC. Instead, the highest prediction-recall AUC values were observed for macronutrient transformations expressed as percentages. Despite this, their classification performance, particularly in terms of F1-score-based accuracy, remained relatively low.

Table 4.5: Areas under the receiver operating characteristic curve (AUC ROC) and precision-recall curve (AUC PR) of the best logistic model for each macronutrient transformation

	AUC ROC	**AUC PR**
Fat_amt_norm	0.55	0.42
Prot_amt_norm	0.50	0.42
Sum_Fat_Prot_norm	0.70	0.53
Fat_perc_norm	0.66	0.61
Prot_perc_norm	0.48	0.52
Sum_perc_Fat_Prot_norm	0.58	0.62
Fat_perc_kcal_norm	0.50	0.50
Prot_perc_kcal_norm	0.51	0.49
Sum_perc_kcal_fat_prot_norm	0.48	0.54

The features selected by each of the best logistic regression model for each macronutrient transformation are listed in Table 4.6. Compared to the LME models, the logistic models generally employed a higher number of features. For instance, the model related to the fat caloric contribution utilized nearly all the extracted features. This model, considering the *a priori* removal features, used all the remaining features except k_{max}. This showed that the logistic models selected through a stepwise backwards elimination approach appear to be less conservative compared to LME models, though this didn't necessarily translate to improved classification performance and better predictive accuracy.

Table 4.6: Features of the best logistic regression models for each macronutrient transformation.

	Best Model Features
Fat_amt_norm	S_I; AUC_{Ra}(60min); AUC_{Ra}(120min); TGR(25%); TGR(50%); TGR(75%); k_{max}; k_{min}; c; d; G_b; $\dot{G}_i$(45min); $\dot{G}_i$(120min)
Prot_amt_norm	S_I; AUC_{Ra}(60min); TGR(25%); TGR(50%); TGR(75%); k_{min}; c; d; G_b; G_s(0min); $\dot{G}_i$(45min); $\dot{G}_i$(120min)
Sum_Fat_Prot_norm	S_I; AUC_{Ra}(60min); AUC_{Ra}(120min); TGR(25%); TGR(50%); TGR(75%); k_{max}; k_{min}; d; G_b; G_s(0min); $\dot{G}_i$(120min)
Fat_perc_norm	S_I; AUC_{Ra}(60min); TGR(75%); k_{abs}; k_{max}; k_{min}; G_b; $\dot{G}_i$(45min); $\dot{G}_i$(120min)
Prot_perc_norm	S_I; AUC_{Ra}(120min); TGR(75%); k_{max}; k_{min}; c; G_s(0min)
Sum_perc_Fat_Prot_norm	S_I; AUC_{Ra}(120min); TGR(50%); TGR(75%); d; $\dot{G}_i$(45min); $\dot{G}_i$(120min)
Fat_perc_kcal_norm	S_I; AUC_{Ra}(60min); AUC_{Ra}(120min); TGR(25%); TGR(50%); TGR(75%); k_abs_norm; k_{min}; c; d; G_b; G_s(0min); $\dot{G}_i$(45min); $\dot{G}_i$(120min)
Prot_perc_kcal_norm	S_I; AUC_{Ra}(120min); TGR(25%); TGR(50%); TGR(75%); k_{max}; k_{min}; d; G_b; $\dot{G}_i$(120min)
Sum_perc_kcal_fat_prot_norm	S_I; AUC_{Ra}(120min); TGR(25%); TGR(50%); TGR(75%); d; G_s(0min); $\dot{G}_i$(45min)

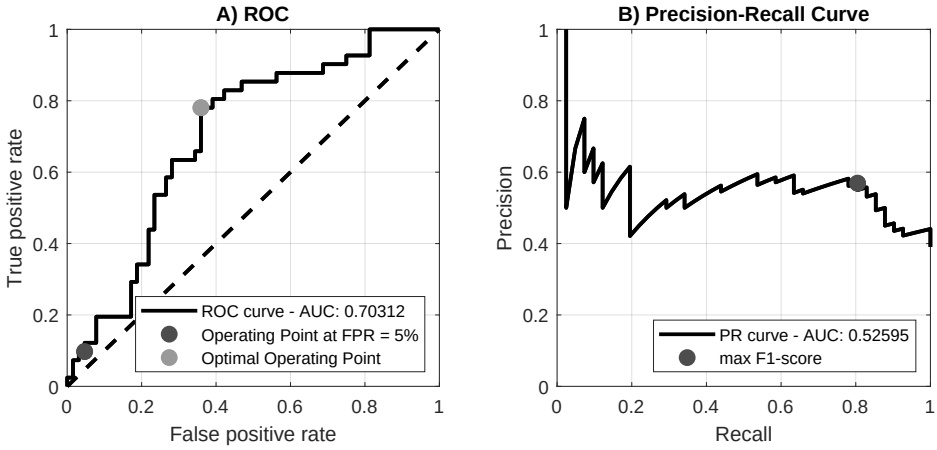

Figure 4.13: ROC curve and prediction-recall curve of the best logistic regression model for the sum of fat and protein amount. Panel A: ROC (solid black line) with optimal operating point (dark-gray marker) and operating point at FPR=5% (light-gray marker). Panel B: prediction-recall curve (solid black line) with max F1-score point (dark-gray marker).

Overall, all the accuracy metrics extracted utilizing the three thresholds and the areas under the ROC and prediction-recall curves highlight the limited applicability of the logistic models in the task of classifying meals based on macronutrient composition. While some transformations, especially the sum of fat and protein amounts, offer moderate accuracy, their results are far from being used in practice.

4.3 Assessment of Unsupervised Learning Techniques

4.3.1 Selected Features

Based on the maximization of the CV prediction accuracy (described in Section 3.5.5), the selected features utilized by each unsupervised learning method were:

- The AUC of Ra at 120 minutes, $AUC_{Ra}(120min)$;
- The half-life of GR curve, TGR(50%);
- The interstitial glucose derivative at 120 minutes, $\dot{G}_i(120min)$.

This feature set is consistent with the findings from the preliminary analysis. The inclusion of $\dot{G}_i(120min)$ is particularly expected, as it exhibited one of the

highest repeated measures correlation coefficients, especially for the macronutrients expressed as their amount. Similarly, AUC_{Ra}(120min) and TGR(50%) were consistently selected in the top-performing LME and logistic regression models predicting the amount of fat, protein, or the sum of the two.

4.3.2 Assessment of K-means Clustering

The optimal number of clusters, estimated with the silhouette approach, was found to be two, with a clear maximum in $\bar{S}$, as illustrated in Fig. 4.14. Feature importance assessed using the $\Delta\bar{S}_p$ metric identified AUC_{Ra}(120min) and TGR(50%) as the most discriminative features. The k-means clustering results, projected onto these two features, is illustrated in Fig. 4.15. From the projection, it is evident that the clustering algorithm partitioned the observations into two distinct groups: one characterized by a higher TGR(50%) and lower AUC_{Ra}(120min), and the other cluster with a lower TGR(50%) and a higher AUC_{Ra}(120min).

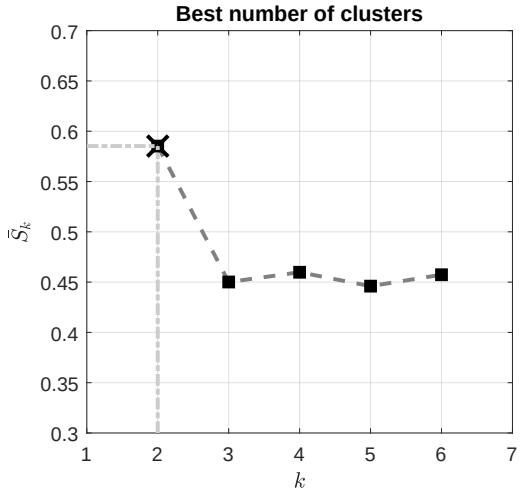

Figure 4.14: Mean Silhouette values, $\bar{S}_k$, for each of the k number of clusters tested (black squares). The best k (black cross) is highlighted.

In Fig. 4.16, the distributions of the three features across the two clusters are shown. It is possible to see that in cluster H, TGR(50%) and $\dot{G}_i$(120min) are generally higher, whereas AUC_{Ra}(120min) has a lower distribution with respect to cluster L.

The distributions of the macronutrient amounts of fat, protein, carbohydrates, and the sum of fat and protein within the two clusters are shown in Fig.

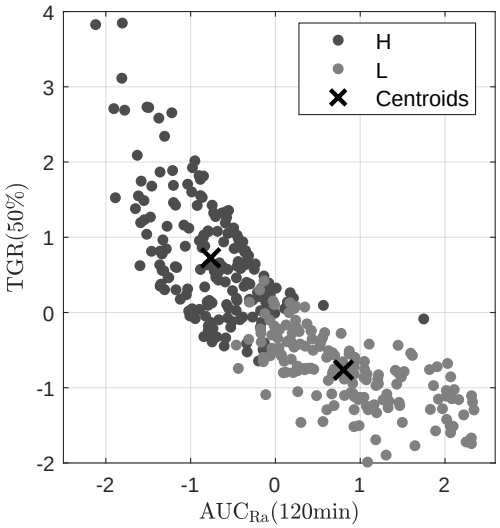

Figure 4.15: *k*-means clustering projected onto the two most important features according to $\Delta\bar{S}_p$ metric. Observations are partitioned between H- (dark gray markers) and L-clusters (light gray markers). Cluster centroids (black crosses) are also reported.

4.17. It can be observed that *k*-means partitioned data so that one cluster displays consistently higher medians across all macronutrients. This cluster was labeled as "cluster H", representing the meals with HF/HP and High-carbohydrate (HC) content, while the other cluster, with lower medians, was labeled "cluster L", representing the meals with LF/LP and Low-carbohydrate (LC) content. The clustering pattern on the features highlighted in Fig. 4.17 is consistent with the interpretation that meals with HF/HP/HC content are associated with distinct postprandial glycemic responses. Furthermore, these results align with the findings reported in the literature [27, 32, 46, 65], reinforcing the validity of the clustering outcome.

In Table 4.7, the p-values and hypothesis testing outcomes from the statistical comparison between the distributions of macronutrients across the two clusters are reported. The lowest p-values, indicating statistically significant differences between clusters, are achieved by the macronutrient transformation expressed as their amount: fat, protein, carbohydrate amount, and the sum of fat and protein amount. In these cases, the null hypothesis of equal distributions was rejected, suggesting that *k*-means clustering identified two physiologically meaningful meal classes.

In contrast, percentage-based macronutrient transformations achieved higher

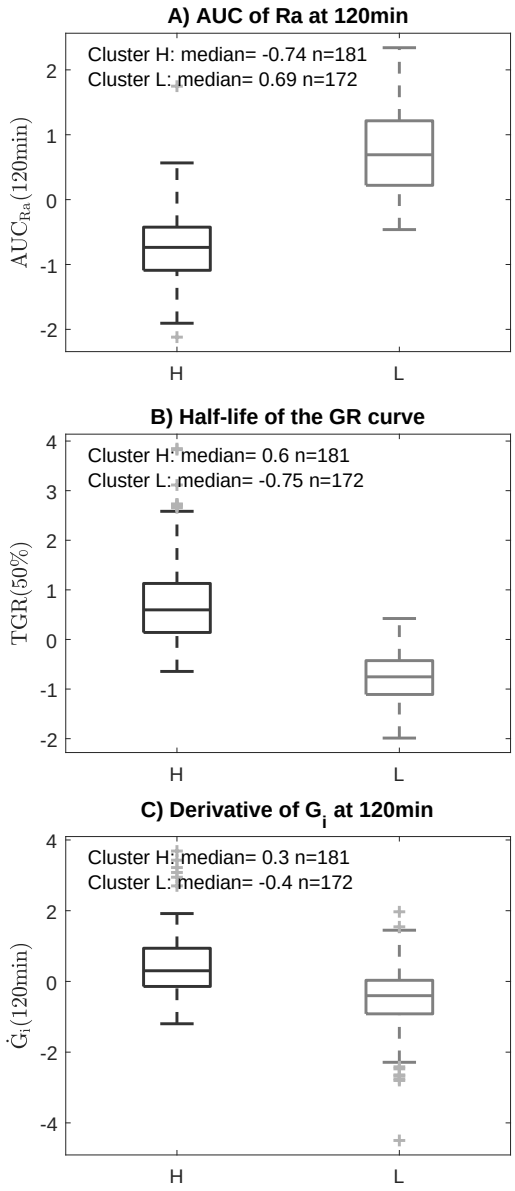

Figure 4.16: Boxplots of the distributions of the features of the cluster H and cluster L. Panel A: Area under the curve of glucose rate of appearance at 120min, $AUC_{Ra}(120min)$. Panel B: Time to reach half of the gastric retention, TGR(50%). Panel C: Interstitial glucose derivative at 120min, $\dot{G}_i(120min)$. The median and the number of observations (n) for each cluster are reported.

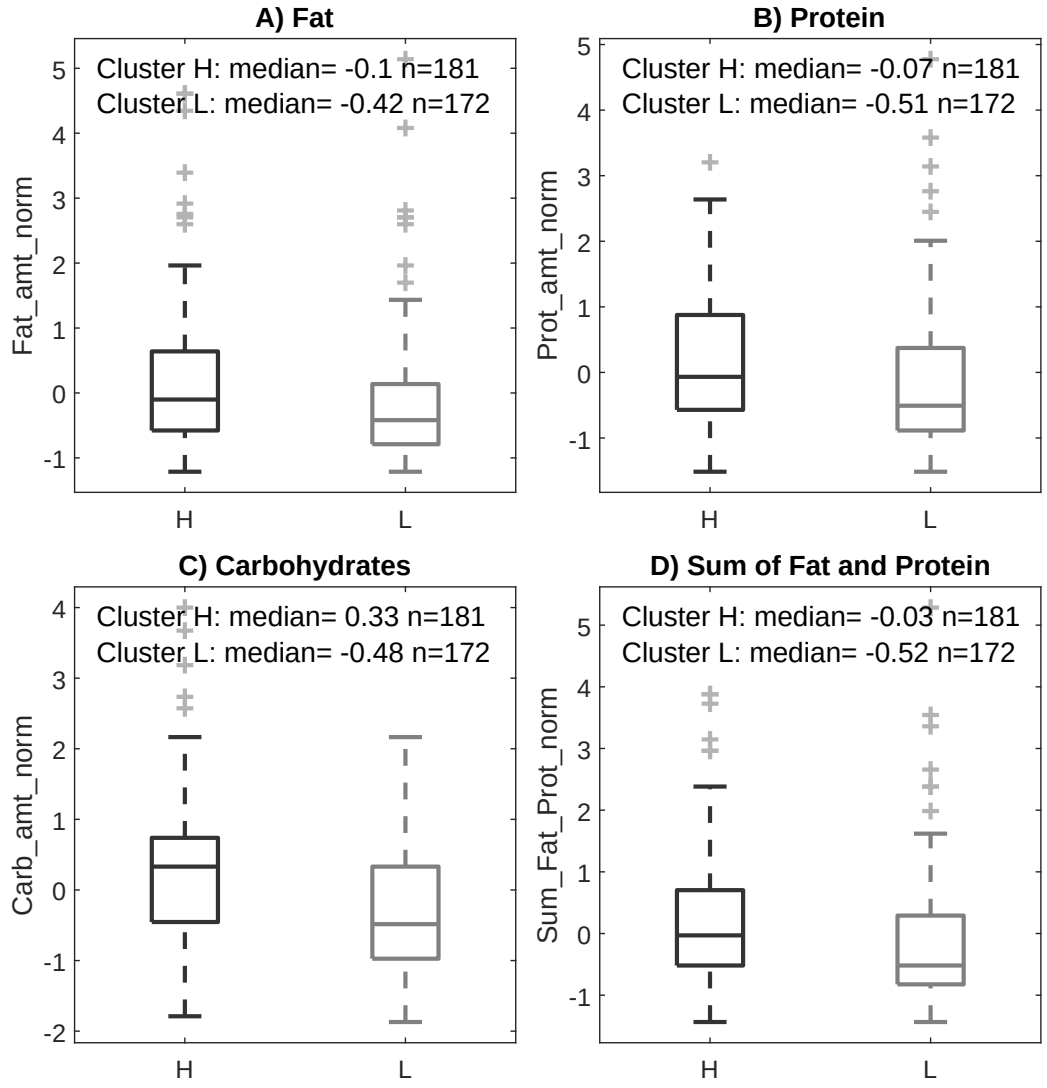

Figure 4.17: Boxplots of the distributions of the macronutrient transformations expressed as absolute amount in the two *k*-means clustering partitions. Panel A: Fat amount. Panel B: Protein amount. Panel C: Carbohydrate amount. Panel C: Sum of Fat and Protein amount. The clusters are labeled based on the rank of their median in Cluster H, with the highest medians, and Cluster L, with the lowest medians. The median and the number of observations (n) for each cluster are also reported in the panels.

p-values, without allowing the rejection of the null hypothesis, thus indicating that these macronutrient transformations do not differ significantly between the clusters and are not informative in distinguishing between meal classes.

Table 4.7: Statistical comparison of macronutrient transformation in the two clusters.

Transformation	**FAT**	**PROT**	**CARB**	**FAT+PROT**	**H_0 Rejected**
Amount	$1.81 \cdot 10^{-5}$	$2.59 \cdot 10^{-4}$	$6.54 \cdot 10^{-9}$	$9.41 \cdot 10^{-6}$	Yes
Amount contribution	0.31822	0.33861	0.90406	0.90406	No
Caloric contribution	0.31822	0.16871	0.76099	0.76060	No

P-values obtained with the Wilcoxon Rank Sum test across all macronutrient transformations (amount, amount contribution, and caloric contribution) for fat (FAT), protein (PROT), carbohydrates (CARB), and the sum of fat and protein (FAT+PROT). The null hypothesis H0 assumes identical distributions, and the rejection (p-value<0.05) indicates a significant difference between the medians of the two clusters.

4.3.3 Assessment of Hierarchical Clustering

The optimal number of clusters for the hierarchical clustering, determined using the Silhouette method, suggested the presence of two distinct clusters also in this case. This result is visually confirmed by the dendrogram of the Ward's linkage tree shown in Fig. 4.18. In the dendrogram, the branches are colored light blue and red to represent the two identified clusters. Within each cluster, the branches merge at relatively low heights, indicating a high degree of similarity among the observations in the same cluster. Since the height in a linkage tree reflects the dissimilarity between observations or clusters, the relatively low internal linkage heights support the homogeneity of each group, indicating a more appropriate clustering result. Conversely, the two major clusters merge at a substantially higher height, reflecting a relatively high dissimilarity between them.

In agreement with the *k*-means clustering, the most important features are AUC_{Ra}(120min) and TGR(50%). The hierarchical clustering projected onto these two features is depicted in Fig. 4.19. The two clusters, also in this case, are partitioned between one cluster with a higher TGR(50%) and lower AUC_{Ra}(120min) and the other cluster with a lower TGR(50%) and a higher AUC_{Ra}(120min). In Fig. 4.20, the distributions of the three features across the two clusters are

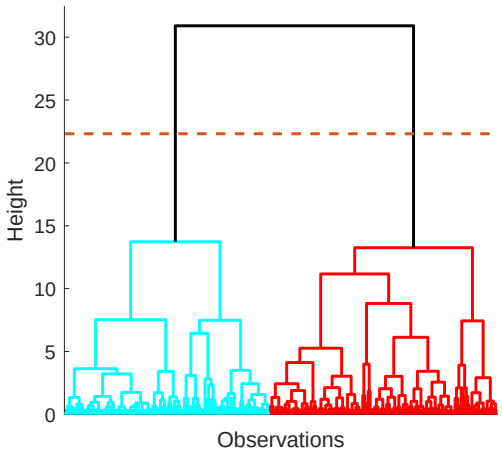

Figure 4.18: Dendrogram of the Ward's linkage tree. Solid red and light-blue lines represent the two clusters. Black solid lines represent the remaining branches of the tree. The orange dashed line represents the linkage tree optimal cut-off according to the Silhouette approach.

reported, confirming similar clustering partitioning as the one obtained with *k*-means.

The distributions of the macronutrient amounts of fat, protein, carbohydrates, and the sum of fat and protein within the two clusters partitioned with hierarchical clustering are shown in Fig. 4.21. Also in this case, two groups of physiological meaning are present: a "cluster H" referring to HF/HP/HC meals and a "cluster L" related to meals with LF/LP/LC content.

Table 4.8 reports the p-values and hypothesis testing outcomes from the Wilcoxon Rank Sum test comparing the distributions of macronutrient transformation variables across the two identified clusters with hierarchical clustering. The results are consistent with *k*-means clustering: the null hypothesis is rejected for fat, protein, carbohydrate amount, and the sum of fat and protein amount, indicating significantly different distributions of the macronutrients across the two clusters.

4.3.4 Assessment of Gaussian Mixture Model Clustering

As the number of clusters suggested by both *k*-means and hierarchical clustering was two, GMM clustering was performed by assuming the existence of two clusters. Consistent with the findings of the two previous unsupervised techniques, the employed features, in order of importance, were AUC_{Ra}(120min),

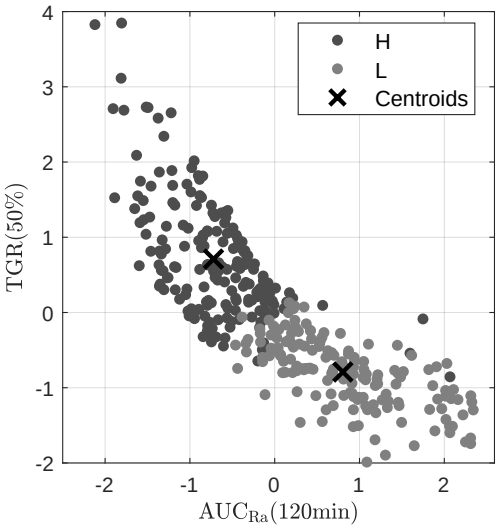

Figure 4.19: Hierarchical clustering projected onto the two most important features according to $\Delta\bar{S}_p$ metric. Observations are partitioned between H- (dark gray markers) and L-clusters (light gray markers). Cluster centroids (black crosses) are also reported.

TGR(50%), and $\dot{G}_i$(120min), confirming the results obtained before.

Fig. 4.22 illustrates the results of GMM clustering, specifically highlighting the posterior probabilities of cluster membership. Specifically, Panel A presents the soft-clustering outcome of the GMM clustering, visualized in the 3D space defined by the employed features. Each data point is colored according to its posterior probability of belonging to cluster L, as indicated by the colorbar. The figure demonstrates that the majority of the observations were assigned to a cluster, L or H, with high confidence, i.e., with a posterior probability close to 1 (observations most likely belonging to cluster L) or 0 (observations most likely belonging to cluster H). This indicates a strong separation between the clusters. Panel B displays the posterior membership probability across all observations that were ranked by the likelihood of belonging to cluster L. Only 32 (i.e., 9.07%) observations showed uncertain membership, i.e., with a posterior between 30% and 70%.

In Fig. 4.23, the GMM clustering results obtained by maximizing the posterior probability are represented in the space of the two most discriminative features according to $\Delta\bar{S}_p$ metric: TGR(50%) and AUC$_{Ra}$(120min). As before, one cluster shows a higher TGR(50%) and lower AUC$_{Ra}$(120min), and the other cluster a lower TGR(50%) and a higher AUC$_{Ra}$(120min).

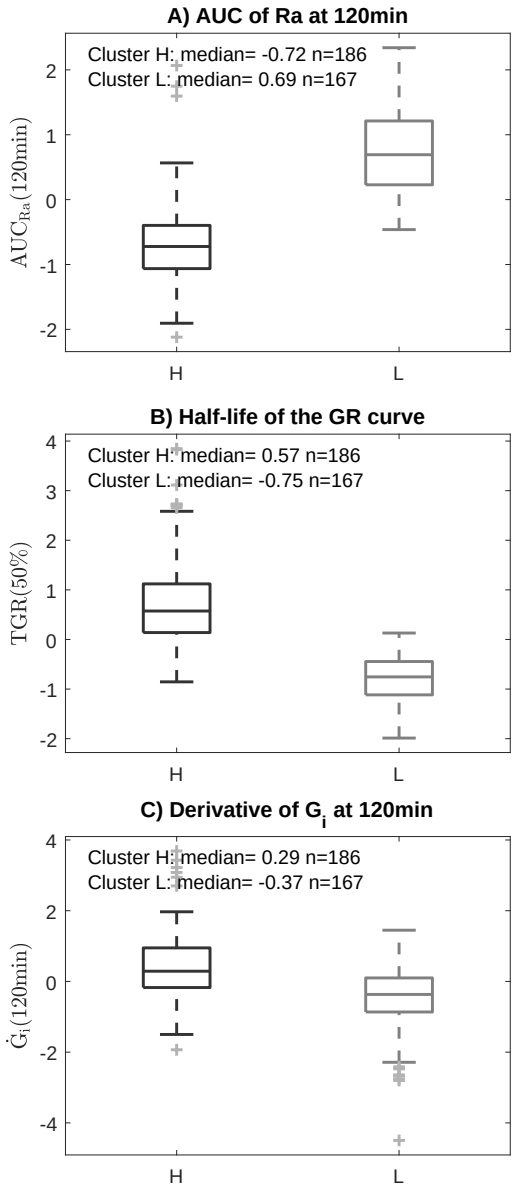

Figure 4.20: Boxplots of the distributions of the features of the cluster H and cluster L. Panel A: Area under the curve of glucose rate of appearance at 120min, $AUC_{Ra}(120min)$. Panel B: Time to reach half of the gastric retention, TGR(50%). Panel C: Interstitial glucose derivative at 120min, $\dot{G}_i(120min)$. The median and the number of observations (n) for each cluster are reported.

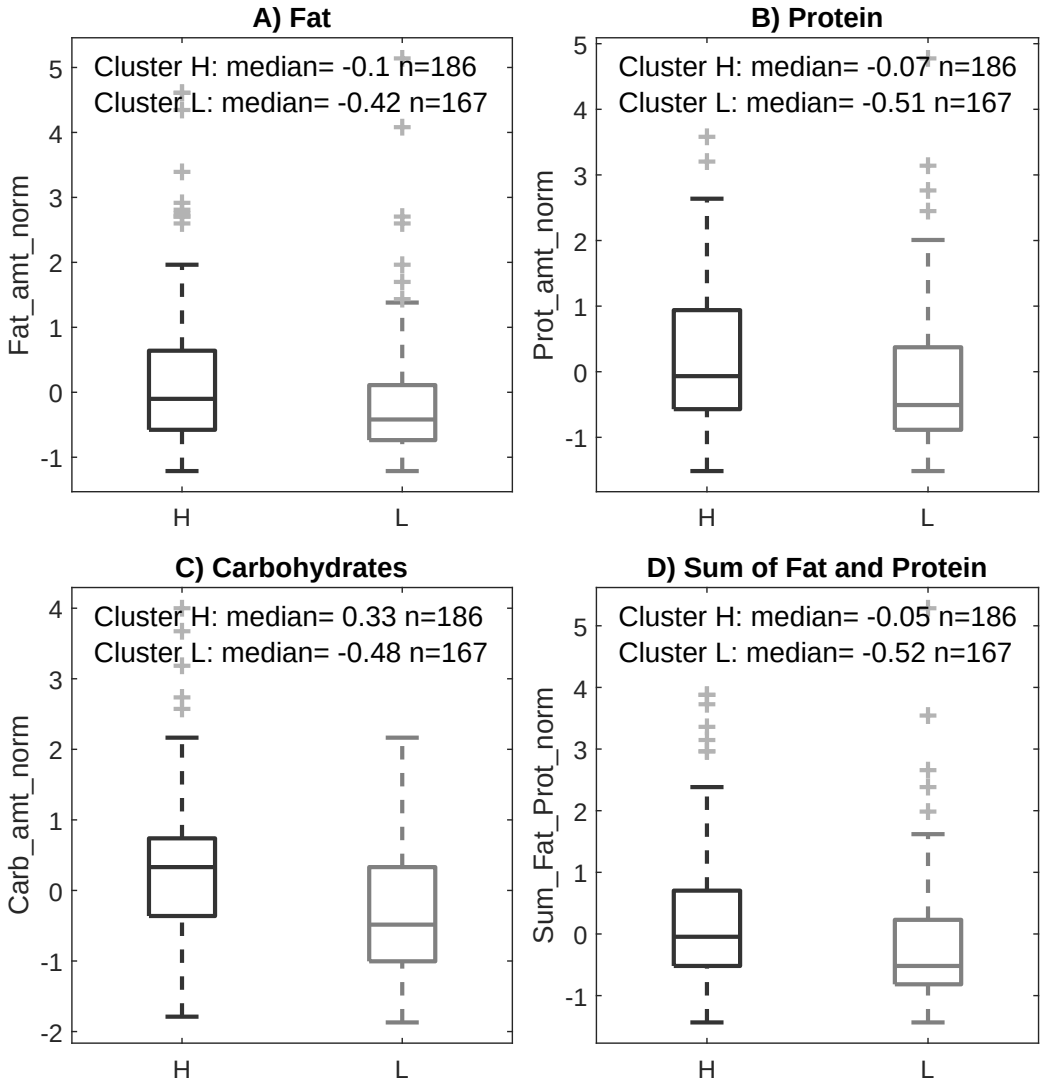

Figure 4.21: Boxplots of the distributions of the macronutrient transformations expressed as absolute amount in the two hierarchical clustering partitions. Panel A: Fat amount. Panel B: Protein amount. Panel C: Carbohydrate amount. Panel C: Sum of Fat and Protein amount. The clusters are labeled based on the rank of their median in Cluster H, with the highest medians, and Cluster L, with the lowest medians. The median and the number of observations (n) for each cluster are also reported in the panels.

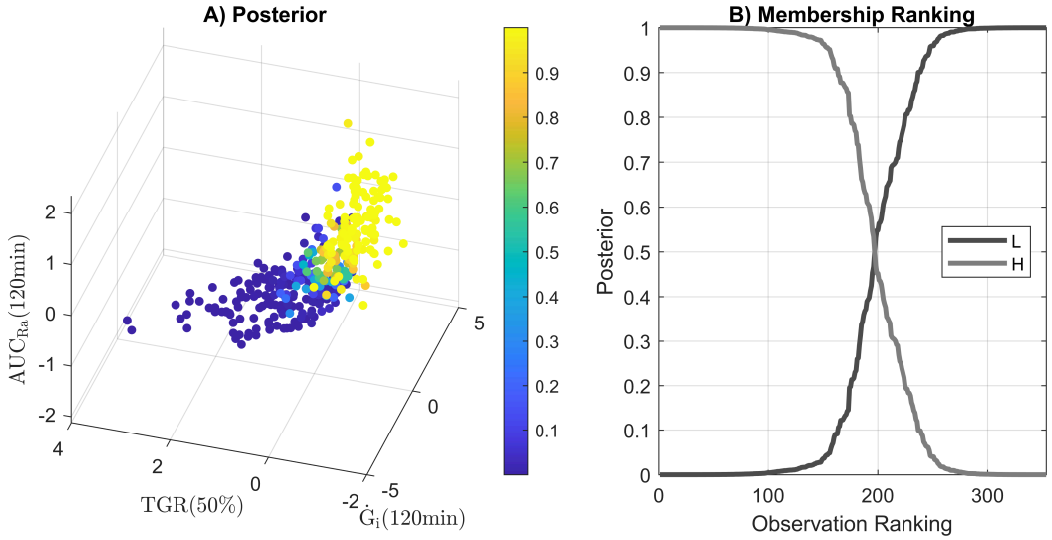

Figure 4.22: Visualization of the posterior probability of GMM clustering. Panel A: Posterior of the cluster L of the observations projected in 3D onto the three features employed: yellow markers=high posterior, blue markers=low posterior. Panel B: Posterior values of cluster H (solid light-gray line) and cluster L (solid dark-gray line) with observations ranked by the increasing posterior of belonging to cluster L.

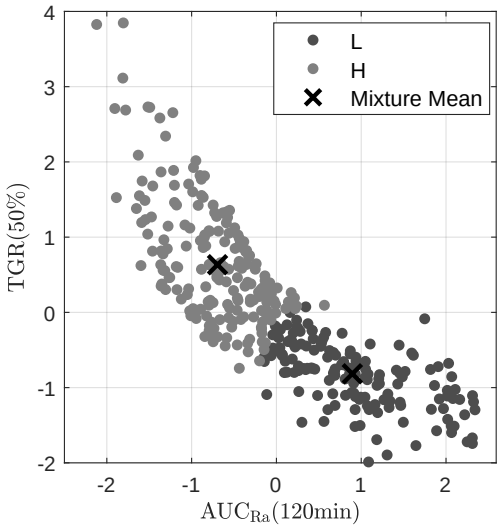

Figure 4.23: GMM clustering projected onto the two most important features according to $\Delta\bar{S}_p$ metric. Observations are partitioned between H- (light gray markers) and L-clusters (dark gray markers). Gaussian mixture means (black crosses) are also reported.

Table 4.8: Statistical comparison of macronutrient transformation in the two clusters.

Transformation	FAT	PROT	CARB	FAT+PROT	H_0 Rejected
Amount	$4.87{\cdot}10^{-5}$	$1.94{\cdot}10^{-4}$	$6.55{\cdot}10^{-9}$	$1.62{\cdot}10^{-5}$	Yes
Amount contribution	0.48429	0.40388	0.97916	0.97916	No
Caloric contribution	0.48429	0.24283	0.88537	0.88496	No

P-values obtained with the Wilcoxon Rank Sum test across all macronutrient transformations (amount, amount contribution, and caloric contribution) for fat (FAT), protein (PROT), carbohydrates (CARB), and the sum of fat and protein (FAT+PROT). The null hypothesis H0 assumes identical distributions, and the rejection (p-value<0.05) indicates a significant difference between the medians of the two clusters.

In Fig. 4.24, the distributions of the three features used for clustering across the two partitions are reported, with similar partitioning as the ones achieved with both *k*-means and hierarchical clustering.

Fig. 4.25 shows the distributions of macronutrients expressed as their total amount compared for the two clusters. GMM clustering partitioned data into two physiologically meaningful groups: the group referring to HF/HP/HC meals (cluster H) and the one indicating LF/LP/LC meals (cluster L).

Table 4.9 reports the p-values and hypothesis testing outcomes from the Wilcoxon Rank Sum test comparing the distributions of the macronutrient transformation across the two identified clusters with GMM clustering. Results are analogous to *k*-means and hierarchical clustering, demonstrating the potential of GMM clustering in distinguishing meal classes with different macronutrient amounts.

4.3.5 Selection of the Best Unsupervised Technique

All three unsupervised learning techniques performed clustering by efficiently finding a hidden pattern in the data and splitting them into two groups that differed in terms of macronutrient content. Table 4.10 presents the prediction accuracies obtained through the 5-fold CV algorithm for each clustering technique and macronutrient transformation. Among the evaluated methods, *k*-means clustering achieved the highest performance, with a CV prediction accuracy of 100%, indicating a stable assignment across all folds. Closely following is GMM clustering, which achieved a CV prediction accuracy of 99.7%. This is

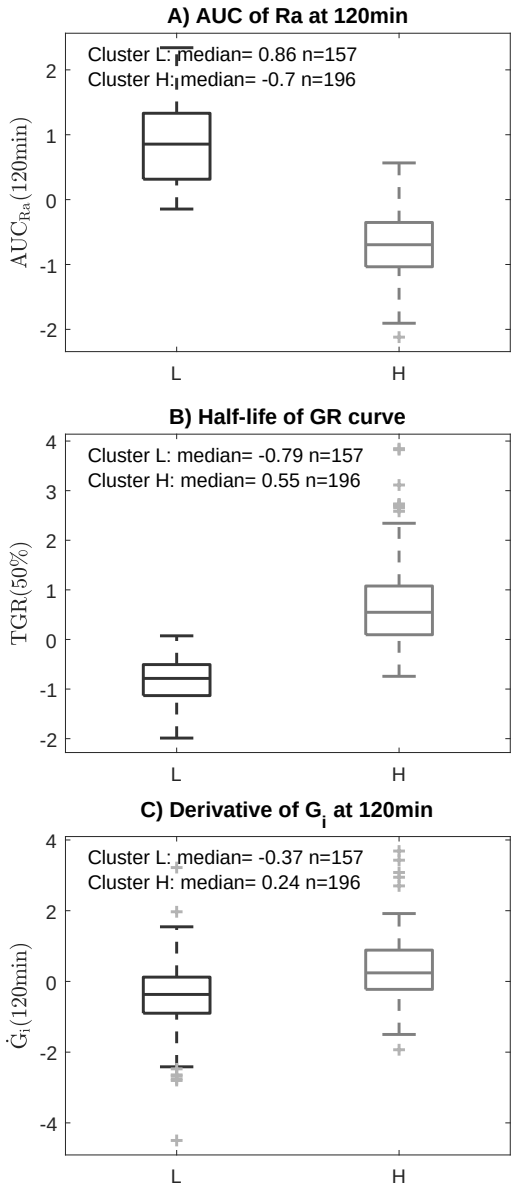

Figure 4.24: Boxplots of the distributions of the features of the cluster H and cluster L. Panel A: Area under the curve of glucose rate of appearance at 120min, $AUC_{Ra}(120min)$. Panel B: Time to reach half of the gastric retention, TGR(50%). Panel C: Interstitial glucose derivative at 120min, $\dot{G}_i(120min)$. The median and the number of observations (n) for each cluster are reported.

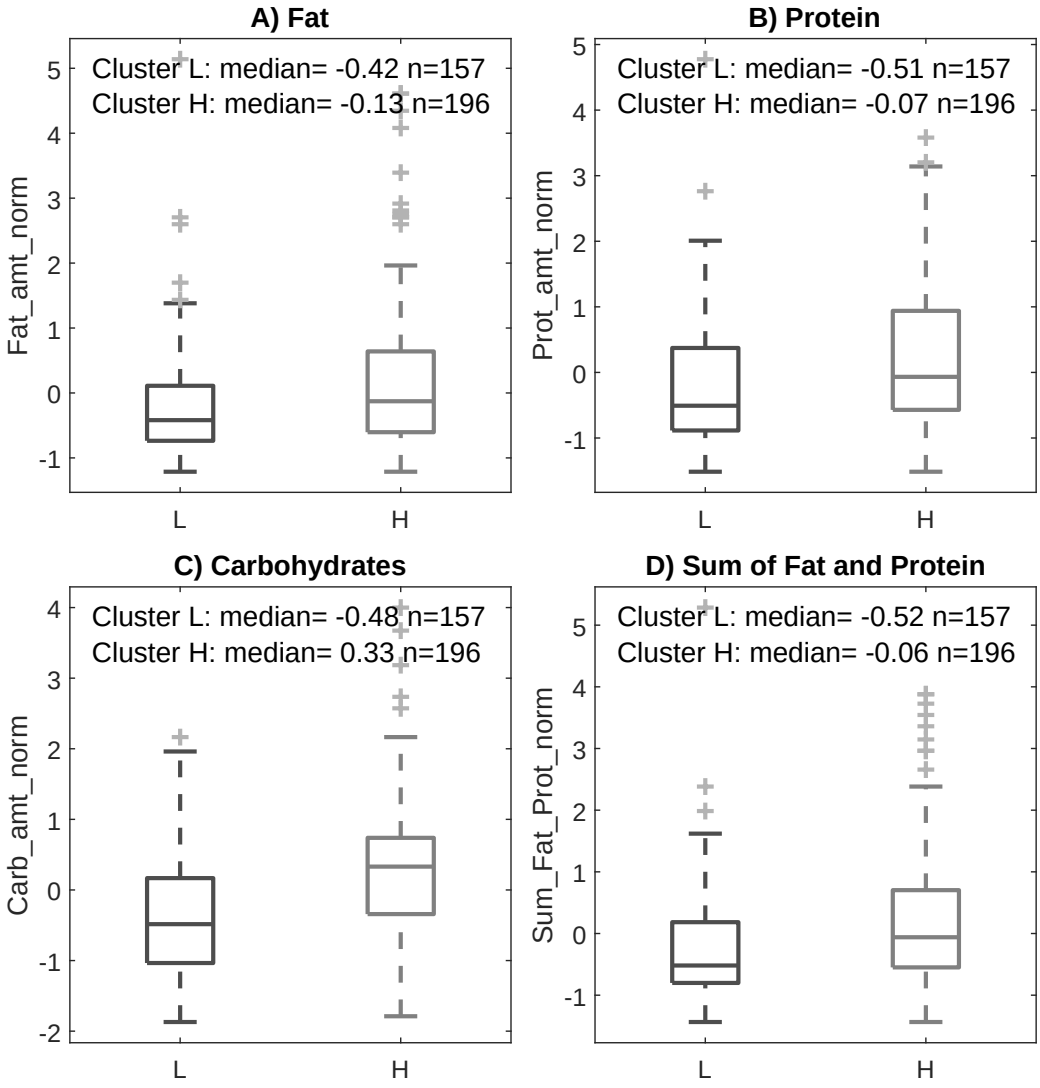

Figure 4.25: Boxplots of the distributions of the macronutrient transformations expressed as absolute amount in the two GMM partitions. Panel A: Fat amount. Panel B: Protein amount. Panel C: Carbohydrate amount. Panel C: Sum of Fat and Protein amount. The clusters are labeled based on the rank of their median in Cluster H, with the highest medians, and Cluster L, with the lowest medians. The median and the number of observations (n) for each cluster are also reported in the panels.

Table 4.9: Statistical comparison of macronutrient transformation in the two clusters.

Transformation	**FAT**	**PROT**	**CARB**	**FAT+PROT**	**H_0 Rejected**
Amount	$6.96{\cdot}10^{-5}$	$1.09{\cdot}10^{-4}$	$1.2{\cdot}10^{-10}$	$1.67{\cdot}10^{-5}$	Yes
Amount contribution	0.86457	0.30391	0.67153	0.67153	No
Caloric contribution	0.86457	0.17946	0.71845	0.71884	No

P-values obtained with the Wilcoxon Rank Sum test across all macronutrient transformations (amount, amount contribution, and caloric contribution) for fat (FAT), protein (PROT), carbohydrates (CARB), and the sum of fat and protein (FAT+PROT). The null hypothesis H0 assumes identical distributions, and the rejection (p-value<0.05) indicates a significant difference between the medians of the two clusters.

consistent with the expectations, as the GMM formalization employed for this analysis is in practice a fuzzy *k*-means clustering. In contrast, hierarchical clustering yielded the lowest performance, although still satisfactory, with a CV prediction accuracy of 89% across all macronutrient transformations. Overall, the high CV prediction accuracies confirm the solidity of the internal structure of the clusters.

Table 4.10: Prediction accuracy over 5-fold cross-validation for each clustering technique and macronutrient transformation.

	Fat_amt_norm	**Prot_amt_norm**	**Carb_amt_norm**	**Sum_Fat_Prot_norm**
k-means	1.000	1.000	1.000	1.000
Hierachical	0.890	0.890	0.890	0.890
GMM	0.997	0.997	0.997	0.997

The table reports the cross-validation prediction accuracy for *k*-means, hierarchical, and Gaussian mixture model (GMM) clustering for fat amount (Fat_amt_norm), protein amount (Prot_amt_norm), carbohydrate amount (Carb_amt_norm), and sum of fat and protein amount (Sum_Fat_Prot_amt).

In conclusion, these findings support our decision in selecting *k*-means as the most robust and efficient technique that exploits physiologically-based features for partitioning and classifying meals based on their macronutrient composition. In Fig. 4.26 the 3D projection of the *k*-means clustering partitions across the three employed features, AUC_{Ra}(120min), TGR(50%) and $\dot{G}_i$(120min), is shown.

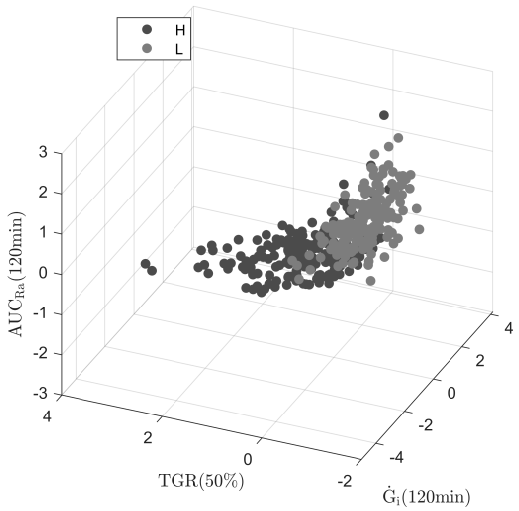

Figure 4.26: *k*-means clustering projected in 3D onto the three employed features. Observations are partitioned between H- (dark gray markers) and L-clusters (light gray markers).

4.4 Results of Validation on Simulated Meals

After evaluating both supervised and unsupervised learning techniques, *k*-means clustering was identified as the most promising method. In this Section, the results of applying this method to classify meals and consequently adjust insulin therapy are reported.

4.4.1 Alignment of Simulated Meals on the Feature Space

Fig. 4.27 illustrates the alignment of simulated meals within the *k*-means feature space of Dataset 2a with a focus on their positioning relative to the cluster centroids. The two most important features were utilized to visualize the two clusters. In Panel A, the features extracted from the simulated meals of Dataset 2a are noticeably misaligned with the *k*-means centroids defined using Dataset 1. As a result, the classification based on the closest-centroid became unreliable. The majority of observations cluster near the L-centroid, leading to a potential misclassification of these meals that would be classified as HF/HP. This problem was also present in the other simulated meals composing Dataset 2b, 2c, and 2d.

This difference was caused by the gap between the simulated and real meals.

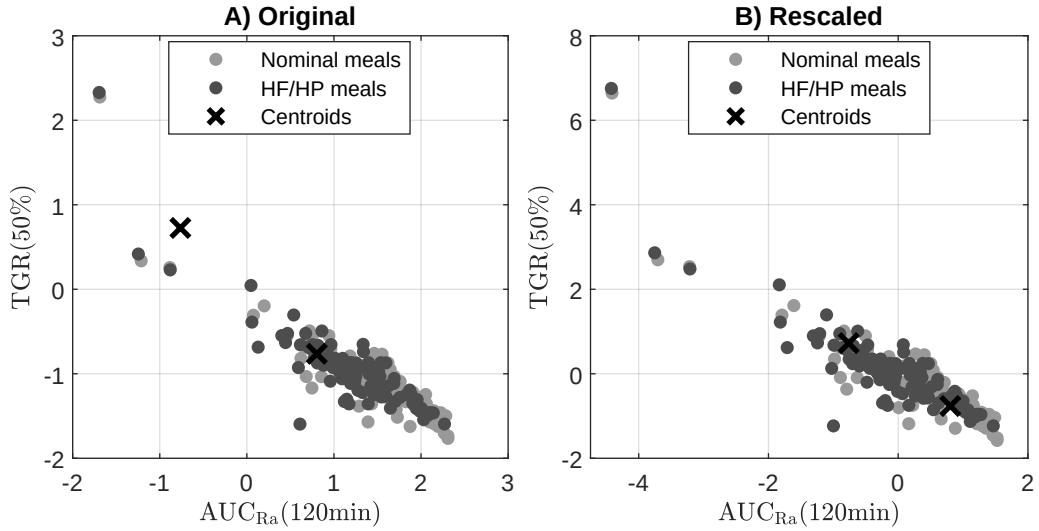

Figure 4.27: Simulated nominal (light-gray markers) and HF/HP (dark-gray markers) meals of Dataset 2a projected onto the two most important features according to $\Delta\bar{S}_p$ metric. Cluster centroids (black crosses) are also reported. Panel A: original projection of the simulated meals. Panel B: projection of the simulated meals after shifting and rescaling.

In fact, simulated postprandial profiles reproduce the shape of the original meal tolerance tests on which the parameters of the simulator were first identified, where subjects ate a jell-O meal, which was much faster than those eaten in daily life.

To compensate for this gap and still be able to validate our method using the UVa/Padova T1D simulator, a shift and rescaling transformation was applied to the features extracted from the simulated meals of Datasets 2a-d. As shown in Panel B of Fig. 4.27, this transformation successfully redistributed the observations within the *k*-means feature space, aligning the simulated meals with the two centroids. Consequently, the classification task becomes more balanced and, possibly, reliable.

4.4.2 Evaluation of Predictive Capability

Assessment of the Methodology on Dataset 2a

The squared Euclidean and Mahalanobis distances were computed between each aligned simulated meal and both *k*-means centroids. The mean distances, stratified by meal type and distance metric, are presented in Table 4.11. As

shown, the mean distances to the H-centroid are shorter, on average, for HF/HP meals than for nominal meals, regardless of the distance metric used. Conversely, nominal meals exhibit shorter mean distances to the L-centroid, whereas HF/HP meals are, on average, farther from it. This pattern is consistent with the expected separation between meal types in the clustering space.

Table 4.11: Mean ± SD distances of Dataset 2a of nominal and HF/HP meals to the H- and L- centroids computed with Squared Euclidean and Mahalanobis distance

Squared Euclidean Distance		
Meal Type	**H- Centroid**	**L- Centroid**
Nominal	4.9848 ± 6.2992	4.1739 ± 9.8594
HF/HP	3.6439 ± 5.8824	4.7402 ± 10.0574
Mahalanobis Distance		
Meal Type	**H- Centroid**	**L- Centroid**
Nominal	2.7201 ± 1.5622	2.2397 ± 1.8657
HF/HP	2.2742 ± 1.3993	2.3881 ± 1.9460

In Fig. 4.28, the confusion matrices reporting the classification results according to (Panel A) squared Euclidean and (Panel B) Mahalanobis distances are shown. Despite the limitations encountered with the use of the simulator, the squared Euclidean distance achieves an accuracy of 59%, meanwhile, the accuracy obtained is 57% when employing the Mahalanobis distance. Given the slightly superior performance of the squared Euclidean distance and considering that this distance metric was also employed during the training phase of the *k*-means clustering on Dataset 1, as well as in the unsupervised CV algorithm, it was selected for the meal classification task.

In Fig. 4.29, the classification results are visualized in the feature space. Panel A shows the true class of each simulated meal, while Panel B displays the assigned meal class based on the smallest Euclidean distance to the centroid of the cluster. By visually inspecting the plots, it's possible to note that the *k*-means clustering meal classification captures the general structure of the true meal classes in the feature space, as shown by the structural similarity between true and assigned class distributions in the feature space. However, the performances are far from being perfect: HF/HP and nominal simulated meals appear to overlap in the feature space, making the classification task non-trivial.

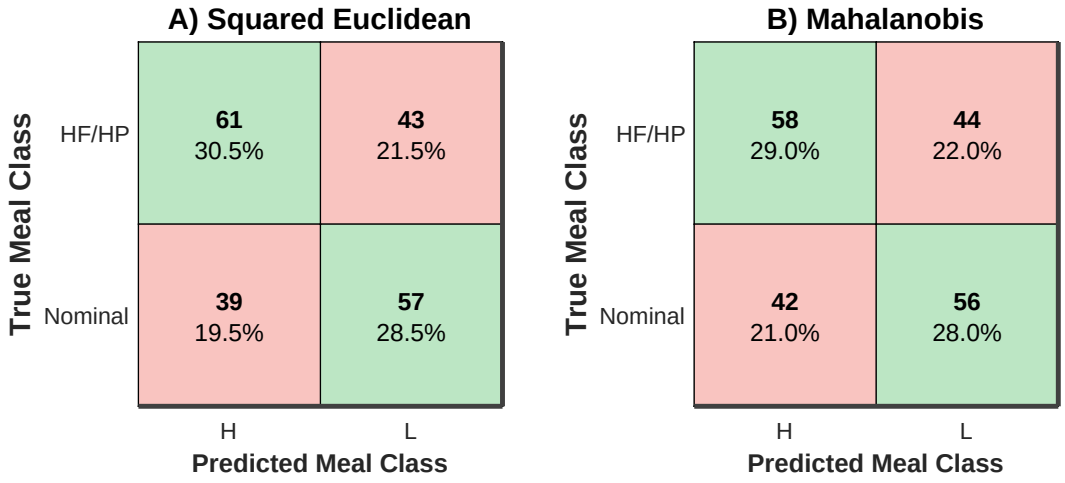

Figure 4.28: Confusion matrices of Dataset 2a of meal assignment for the two tested metrics. Panel A: Meal classification with squared Euclidean distance. Panel B: Meal classification with Mahalanobis distance.

Assessment of the Methodology on Datasets 2b-d

To assess whether the simulated meals were insufficiently differentiated between HF/HP and nominal categories, and to evaluate the potential for improved classification accuracy using the T1D simulator, three sets of HF/HP and nominal meals were generated, creating Dataset 2b, 2c, and 2d. In each set, 100 HF/HP and 100 nominal meals are present, as for Dataset 2a. In these newly simulated HF/HP meals, the parameters governing glucose absorption were progressively accentuated, thereby creating increasingly extreme HF/HP profiles with the aim of enhancing their distinction from nominal meals.

Table 4.12 reports the mean ± SD distances of the observations in Datasets 2b–d from the two *k*-means clustering centroids. Consistent with the results observed for Dataset 2a, all HF/HP meals in Datasets 2b–d are, on average, closer to the H-centroid and farther from the L-centroid. Conversely, nominal meals are, on average, closer to the L-centroid and farther from the H-centroid. The largest difference in average distance between nominal and HF/HP meals with respect to the L-centroid is observed in Dataset 2d. This outcome is expected, as this dataset includes the most accentuated HF/HP meals. However, the HF/HP meals in this dataset exhibit only a minimal difference in average distance relative to nominal meals with respect to the H-centroid.

Fig. 4.30 presents the confusion matrices for meal assignment classification

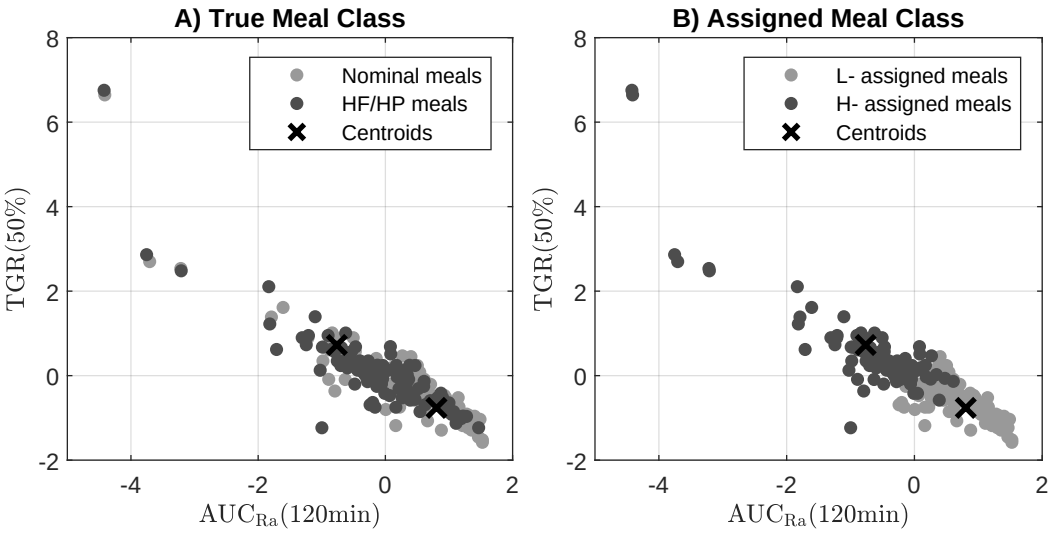

Figure 4.29: Simulated meals of Dataset 2a projected onto the two most important features according to $\Delta\bar{S}_p$ metric. Cluster centroids (black crosses) are also reported. Panel A: True meal classes: nominal meals (light gray markers) and HF/HP meals (dark gray markers). Panel B: Assigned meal classes: L- (light gray markers) and H- (dark gray markers).

Table 4.12: Mean ± SD distances of Datasets 2b-d of nominal and HF/HP meals to the H- and L-centroids computed with Squared Euclidean distance

Dataset	**Meal Type**	**L- Centroid**	**H- Centroid**
2b	Nominal	3.8487 ± 7.9898	4.8286 ± 5.2746
	HF/HP	5.0654 ± 10.3401	3.8000 ± 6.0692
2c	Nominal	3.7477 ± 8.4386	4.8714 ± 5.5433
	HF/HP	5.1664 ± 9.9375	3.7572 ± 5.8650
2d	Nominal	3.1908 ± 6.0744	4.3241 ± 4.2113
	HF/HP	5.7233 ± 12.1798	4.3045 ± 6.8120

across Datasets 2b–d. As anticipated, given the stronger separation between HF/HP and nominal meals, classification accuracy improves relative to Dataset 2a, reaching 62% for Dataset 2c, which also displayed the clearest distinction in average distances between the H- and L-centroids across the two meal types. Nevertheless, the overall classification accuracy remains low, and the improvement is less pronounced than expected, even with the more emphasized HF/HP meals.

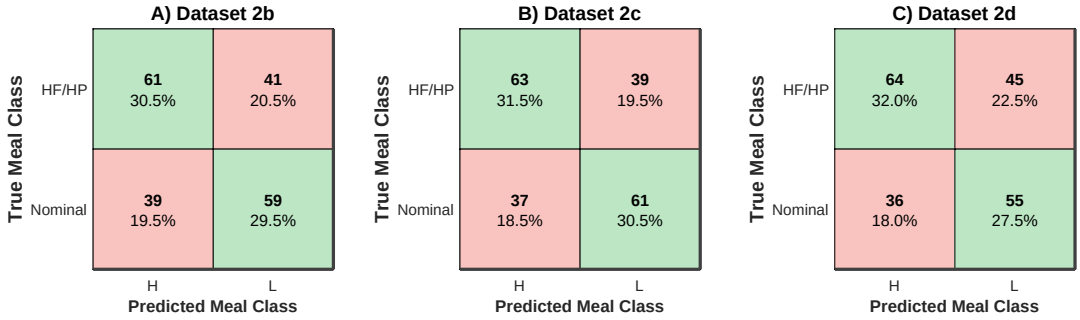

Figure 4.30: Confusion matrices of Dataset 2b-d meal assignment classification with squared Euclidean distance. Panel A: Dataset 2b. Panel B: Dataset 2c. Panel C: Dataset 2d.

In Fig. 4.31, the classification results for the best-performing simulated dataset (Dataset 2c) are illustrated. Panel A shows the clustering in the feature space defined by the two most important features according to the Silhouette method, while Panel B displays the true meal classes. As evident from this comparison, the low accuracy reported in the confusion matrix arises from the inability of the T1D simulator to *in silico* generate meals with a clear distinct pattern between HF/HP and nominal meals in the selected features, resulting in two overlapping groups of observations. While HF/HP meals show some elongation along the H-centroid, this effect is insufficient, as nominal and HF/HP meals remain highly overlapping in the feature space. Consequently, classification based on *k*-means centroid distances on these simulated meals proves to be ineffective, since the simulated meals fail to exhibit distinct population-level patterns necessary for reliable classification.

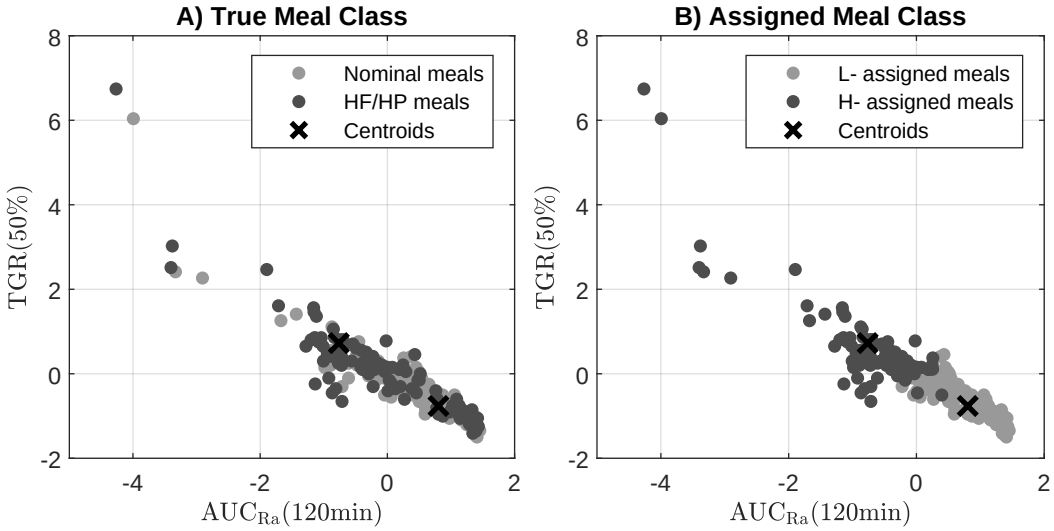

Figure 4.31: Simulated meals of Dataset 2c projected onto the two most important features according to $\Delta\bar{S}_p$ metric. Cluster centroids (black crosses) are also reported. Panel A: True meal classes: nominal meals (light gray markers) and HF/HP meals (dark gray markers). Panel B: Assigned meal classes: L- (light gray markers) and H- (dark gray markers).

4.4.3 Evaluation of the Adjusted Insulin Therapy

Assessment of the Adjusted Insulin Therapy on Dataset 2a

Fig. 4.32 illustrates the median CGM profiles and corresponding Interquartile Ranges (IQR) resulting from the standard insulin therapy (Dataset 2.1a) and the insulin therapy assigned based on clustering classification (Dataset 2.2a). As shown in Panel A, the HF/HP meals CGM profiles recorded with the standard therapy (i.e., based on the optimal carbohydrate counting rule) demonstrate a higher postprandial glycemia when compared to nominal meals. Whereas in Panel B, where HF/HP meals were treated, for the majority, with the DWB strategy, the resulting CGM profiles start to resemble those observed for nominal meals. This alignment between the CGM profiles of nominal and HF/HP meals indicates that our clustering approach was able to correctly detect some of the meals that needed a therapy adjustment and that DWB therapy was able to mitigate part of the glycemic impact of those meals.

Table 4.13 summarizes the CGM metrics obtained in each case, presenting the median and IQR for each metric, stratified by insulin therapy (standard, i.e., Dataset 2.1a, and adjusted, i.e., Dataset 2.2a), and meal type (nominal and

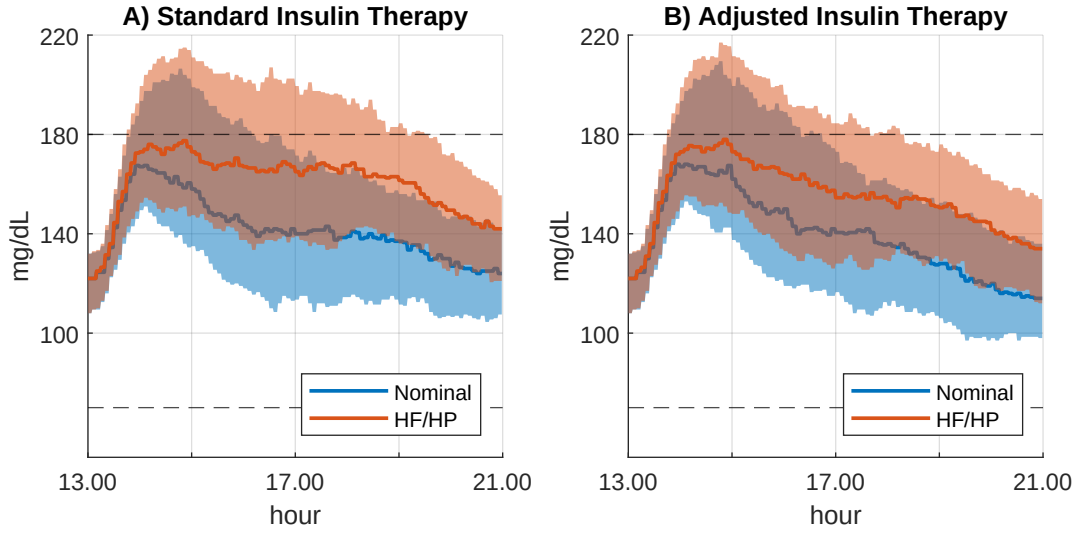

Figure 4.32: Median (solid lines) and interquartile ranges (colored areas) of postprandial CGM profiles of nominal (blue) and HF/HP meals (orange). Horizontal dashed lines represent hypo- and hyperglycemic thresholds. Panel A: Standard insulin therapy (Dataset 2.1a). Panel B: Insulin therapy assigned based on clustering results (Dataset 2.2a).

HF/HP). As anticipated from the postprandial glycemic curves in Fig. 4.32, the median TIR for HF/HP meals improved from 77.6% under standard therapy to 79.17% with the adjusted therapy. In contrast, the TIR for nominal meals decreased from 87.5% under standard therapy to 80.73%, since, due to misclassification, 39 out of 100 nominal meals were classified as HF/HP, and were treated with the DWB.

A similar trend was observed for TAR. For HF/HP meals, the median TAR decreased from 20.31% (standard therapy) to 18.75% (adjusted therapy), suggesting improved postprandial glycemic control. These improvements are consistent with the CGM profiles visualized in Fig. 4.32, where glycemia under assigned therapy resembled more closely the postprandial CGM profiles of nominal meals.

Additionally, the median CGM over the 8 hours, $\overline{\text{CGM}}$, for HF/HP meals decreased, from 162.68 mg/dl under the standard therapy to 157.26 mg/dl with the adjusted therapy. The median and the IQR of TBR remained stable at 0, indicating no increase in hypoglycemia risk. One adverse effect observed was an increase in the number of HE for HF/HP meals, rising from 1 event (standard therapy) to 4 events (adjusted therapy). While this is clearly a worsening of that

Table 4.13: Performances of CGM Metrics for standard (Dataset 2.1a) and adjusted (Dataset 2.2a) insulin therapies for nominal and HF/HP meals. P-values of the Wilcoxon signed rank sum test are reported.

Therapy	Meal Type	Median	IQR (Q1 – Q3)	p-value
TIR [%]				
Standard	Nominal	87.5	[64.06 – 100]	$2.0210 \cdot 10^{-5}$
	HF/HP	77.6	[41.67 – 100]	$3.2018 \cdot 10^{-5}$
Adjusted	Nominal	80.73	[62.5 – 100]	–
	HF/HP	79.17	[51.04 – 100]	–
TBR [%]				
Standard	Nominal	0	[0 – 0]	0.0085
	HF/HP	0	[0 – 0]	0.2500
Adjusted	Nominal	0	[0 – 0]	–
	HF/HP	0	[0 – 0]	–
TAR [%]				
Standard	Nominal	5.21	[0 – 30.21]	$6.2936 \cdot 10^{-14}$
	HF/HP	20.31	[0 – 57.29]	$2.9802 \cdot 10^{-8}$
Adjusted	Nominal	5.21	[0 – 31.77]	–
	HF/HP	18.75	[0 – 46.88]	–
$\overline{\textbf{CGM}}$ [mg/dl]				
Standard	Nominal	142.42	[122.36 – 165.83]	$6.4739 \cdot 10^{-25}$
	HF/HP	162.68	[140.56 – 185.21]	$1.3878 \cdot 10^{-17}$
Adjusted	Nominal	138.56	[120.76 – 161.61]	–
	HF/HP	157.26	[134.21 – 177.98]	–
#HE		**Amount**		
Standard	Nominal	13	–	–
	HF/HP	1	–	–
Adjusted	Nominal	21	–	–
	HF/HP	4	–	–

The table reports the median and IQR for TIR, TBR, TAR and $\overline{\text{CGM}}$. The number of HE is also reported. The performance metrics are stratified for meal type and insulin therapy.The p-values in the last column come from Wilcoxon Signed Rank tests: (i) Nominal Standard vs. HF/HP Adjusted; (ii) HF/HP Standard vs. HF/HP Adjusted.

particular metric, it is worth noting that the number of HE for nominal meals was 13 under standard therapy. This suggests that the adjusted insulin therapy closely replicates the performance of the standard therapy for nominal meals, which corresponds to the optimal insulin bolus possible, except for the presence of the carb-counting error.

It is noteworthy that 39% of nominal meals were misclassified as HF/HP meals. This misclassification caused an erroneous assignment of the DWB insulin dosing strategy to nominal meals, which resulted in a relevant increase in HE, rising from 13 under standard therapy to 21 under adjusted therapy.

The Wilcoxon Signed Rank paired two-sided test was employed to detect differences between the median of the metrics obtained in Dataset 2.1a and Dataset 2.2a. A significance level of 5% was used to determine whether to reject the null hypothesis. First, statistical differences were examined between nominal meals under standard therapy and HF/HP meals under adjusted therapy. Second, differences between HF/HP meals under standard and adjusted therapy were assessed. The p-values of these two statistical tests are reported in the last column of Table 4.13.

When comparing CGM metrics derived from HF/HP meals between Dataset 2.1a and Dataset 2.2a, the null hypothesis was consistently rejected across all metrics except TBR, where the null hypothesis was accepted, indicating statistically significant differences between the reference and assigned therapy, confirming the efficacy of the DWB strategy to treat HF/HP meals in our simulations.

In contrast, when comparing nominal meals from Dataset 2.1a to the HF/HP meals with adjusted dosing (Dataset 2.2a), the null hypothesis was rejected for all CGM metrics. This outcome suggests that, although visual inspection of the CGM profiles indicates a potential improvement in glycemic control, statistically significant differences remained between the nominal meals treated with the standard insulin therapy and HF/HP meals treated with the assigned dosing, meaning that there is still room for improving TAR, TIR, and $\overline{\text{CGM}}$. This improvement is likely to be achieved by using an individualized DWB therapy, optimized based on the subject and on the type of meal.

Assessment of the Adjusted Insulin Therapy on Datasets 2b-d

Fig. 4.33, 4.34, and 4.35 illustrate the median CGM profiles and corresponding IQR obtained under standard insulin therapy (Panel A), comprising Datasets

2.1b, 2.1c, and 2.1d, respectively, as well as under insulin therapy assigned on the basis of *k*-means clustering classification (Panel B), forming Datasets 2.2b, 2.2c, and 2.2d, respectively (i.e., applying a DWB when a meal was classified as HF/HP and otherwise administering standard insulin therapy).

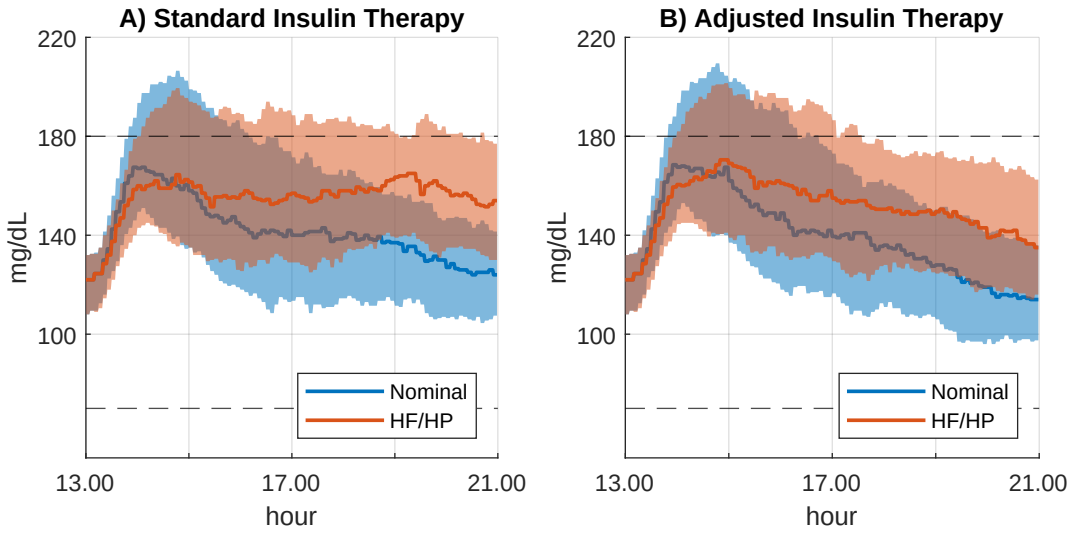

Figure 4.33: Median (solid lines) and interquartile ranges (colored areas) of postprandial CGM profiles of nominal (blue) and HF/HP meals (orange). Horizontal dashed lines represent hypo- and hyperglycemic thresholds. Panel A: Standard insulin therapy (Dataset 2.1b). Panel B: Insulin therapy assigned based on clustering results (Dataset 2.2b).

Comparable patterns are observed in the CGM profiles of simulated meals in Dataset 2.2a. Also in this case, the adjusted therapy assigned to HF/HP meals resulted in a notable decrease of late postprandial hyperglycemia. The median CGM profiles of HF/HP meals increasingly resembled those of nominal meals, indicating that applying the DWB strategy to a subset of HF/HP meals contributed to reducing glycemic variability and improving glycemic control. Nonetheless, this improvement is less evident in the most extreme HF/HP simulated meals (Dataset 2.2d).

Table 4.14 reports the CGM metrics of the best-performing dataset, i.e., Dataset 2c. The table presents the median and IQR for each metric, stratified by insulin therapy and meal type. The CGM metrics are provided for both insulin therapies (standard, i.e., Dataset 2.1c, and adjusted, i.e., Dataset 2.2c) and for both meal types (nominal and HF/HP). In contrast to what observed in Dataset 2a, in Dataset 2c, the TIR of HF/HP meals decreased from 89.06% under

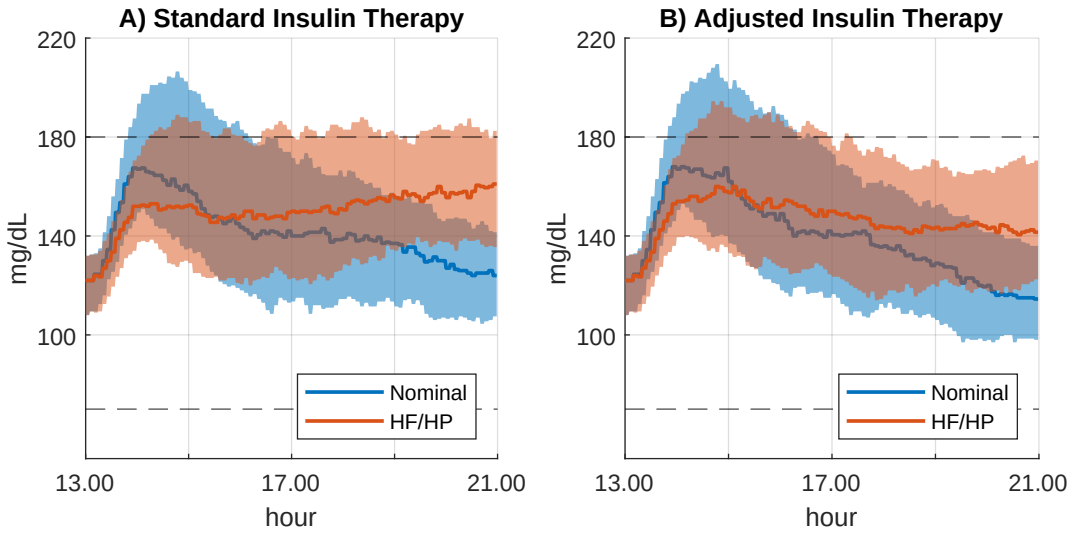

Figure 4.34: Median (solid lines) and interquartile ranges (colored areas) of postprandial CGM profiles of nominal (blue) and HF/HP meals (orange). Horizontal dashed lines represent hypo- and hyperglycemic thresholds. Panel A: Standard insulin therapy (Dataset 2.1c). Panel B: Insulin therapy assigned based on clustering results (Dataset 2.2c).

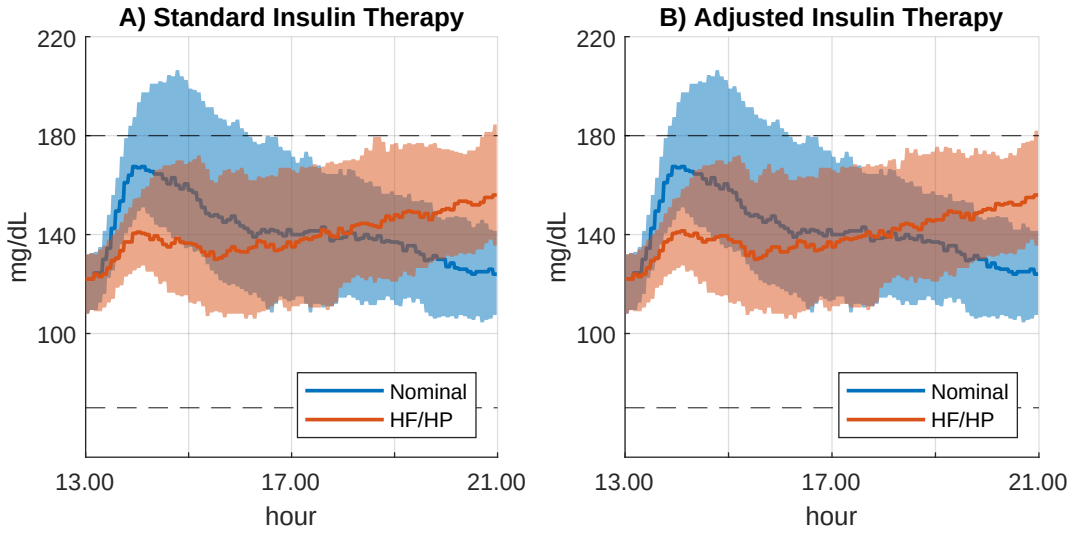

Figure 4.35: Median (solid lines) and interquartile ranges (colored areas) of postprandial CGM profiles of nominal (blue) and HF/HP meals (orange). Horizontal dashed lines represent hypo- and hyperglycemic thresholds. Panel A: Standard insulin therapy (Dataset 2.1d). Panel B: Insulin therapy assigned based on clustering results (Dataset 2.2d).

standard therapy to 85.94% under adjusted therapy. This was accompanied by a reduction in the TIR of nominal meals, from 87.50% with standard therapy to 80.73% with adjusted therapy. The decline is due to the misclassification of 37 nominal meals out of 100 that were incorrectly categorized as HF/HP. Interestingly, in these simulated meals, the TIR of HF/HP meals under standard therapy (i.e., based on carbohydrate counting only) was higher than that of nominal meals. This result contradicts the assumption made when generating HF/HP meals CGM profiles with the T1D simulator. The TIR would normally be expected to be lower for HF/HP meals compared to nominal ones, given that HF/HP meals treated with standard insulin therapy are typically affected by early hypoglycemia and/or delayed hyperglycemia, thus causing a degradation of this metric.

Other CGM metrics showed limited variation between nominal meals under standard and adjusted therapy. A slight reduction in $\overline{\text{CGM}}$ was observed, from 142.42 mg/dl to 140.10 mg/dl, accompanied by an increase in the number of HE that rose from 13 to 20. In contrast, median TAR and TBR remained unchanged. For HF/HP meals, the median TAR and TBR also remained stable, whereas $\overline{\text{CGM}}$ decreased from 150.78 mg/dl to 147.15 mg/dl, with the number of HE increasing from 8 under standard therapy to 10 under adjusted therapy.

Table 4.14 also reports the p-values from the Wilcoxon Signed Rank paired two-sided test, conducted to assess whether meaningful differences exist between CGM profiles of HF/HP meals under adjusted therapy compared with CGM tracks of nominal meals and HF/HP meals under standard therapy. As before, statistical significance was set at the 5% level. When comparing CGM profiles of HF/HP meals with adjusted therapy against nominal meals with standard therapy, the null hypothesis was accepted for TIR, TBR, and TAR, indicating no significant differences between the two profiles. This suggests that HF/HPmeals treated with the adjusted therapy, most of which were assigned a DWB strategy, produced CGM profiles resembling those of nominal meals. Conversely, when comparing HF/HP meals with adjusted therapy to those with standard therapy, the null hypothesis was rejected for TAR and $\overline{\text{CGM}}$, indicating significant differences between treatments, while it was accepted for TIR and TBR.

Table 4.14: Performances of CGM Metrics for standard (Dataset 2.1c) and adjusted (Dataset 2.2c) insulin therapies for nominal and HF/HP meals. P-values of the Wilcoxon signed rank sum test are reported.

Therapy	**Meal Type**	**Median**	**IQR (Q1 – Q3)**	**p-value**
TIR [%]				
Standard	Nominal	87.5	[64.06 – 100]	0.3494
	HF/HP	89.06	[55.73 – 100]	0.1588
Adjusted	Nominal	80.73	[62.5 – 100]	–
	HF/HP	85.94	[85.94 – 100]	–
TBR [%]				
Standard	Nominal	0	[0 – 0]	0.4292
	HF/HP	0	[0 – 0]	0.1250
Adjusted	Nominal	0	[0 – 0]	–
	HF/HP	0	[0 – 0]	–
TAR [%]				
Standard	Nominal	5.21	[0 – 30.21]	0.5635
	HF/HP	4.17	[0 – 43.75]	0.0121
Adjusted	Nominal	5.21	[0 – 31.77]	–
	HF/HP	4.17	[0 – 34.38]	–
$\overline{\textbf{CGM}}$ [mg/dl]				
Standard	Nominal	142.42	[122.36 – 165.83]	0.0158
	HF/HP	150.78	[132.97 – 173.38]	$6.9389 \cdot 10^{-18}$
Adjusted	Nominal	140.1	[120.76 – 161.61]	–
	HF/HP	147.15	[126 – 170.84]	–
#HE		**Amount**		
Standard	Nominal	13	–	–
	HF/HP	8	–	–
Adjusted	Nominal	20	–	–
	HF/HP	10	–	–

The table reports the median and IQR for TIR, TBR, TAR, and $\overline{\text{CGM}}$. The number of HE is also reported. The performance metrics are stratified for meal type and insulin therapy. The p-values in the last column come from Wilcoxon Signed Rank tests: (i) Nominal Standard vs. HF/HP Adjusted; (ii) HF/HP Standard vs. HF/HP Adjusted.

4.5 Evaluation of the Impact of the Model-based Features

Fig. 4.36 illustrates the results of k-means clustering applied using only CGM-derived features: CGM(0min), $\dot{\text{CGM}}$(45min), $\dot{\text{CGM}}$(120min). In particular, meals from Dataset 1 are visualized in Fig. 4.36 according to the two most informative CGM-based features, as determined by the $\Delta\bar{S}_p$ metric: the CGM derivative at 45 minutes after mealtime, $\dot{\text{CGM}}$(45min), and the CGM value at mealtime, CGM(0min). For consistency with the notation used previously, the two clusters obtained were named H- and L-cluster, where H showed generally a higher amount of macronutrients. In this cluster, $\dot{\text{CGM}}$(45min) was lower compared to the L cluster, whereas $\dot{\text{CGM}}$(120min) and CGM(0min) were higher. This, in general, supports the results presented in Section 4.1. Nonetheless, the partitioning strength is weak, and the overall cluster structure appears poorly defined.

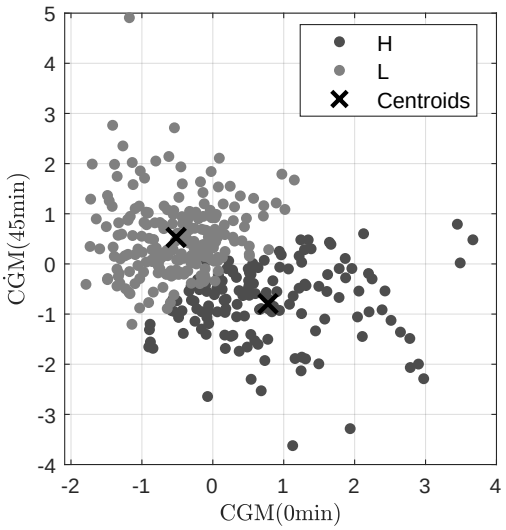

Figure 4.36: k-means clustering projected onto the two most important CGM-derived features according to $\Delta\bar{S}_p$ metric. Observations are partitioned between H- (dark gray markers) and L-clusters (light gray markers). Cluster centroids (black crosses) are also reported.

The distributions of the macronutrient amounts of fat, protein, carbohydrates, and the sum of fat and protein within the two clusters are shown in Fig. 4.37. Unlike the results obtained with model-based features, i.e., those

derived from the identification of a mechanistic model, here, the macronutrient distributions showed minimal differences between the two CGM-based clusters.

This lack of clear separation is further supported by the results of the Wilcoxon Rank Sum test, reported in Table 4.15, which failed to reject the null hypothesis of equal distributions at the 5% significance level for all the macronutrients, indicating that the two macronutrient distributions in the H- and L-clusters do not have significantly different medians. These findings suggest that clustering based solely on CGM-derived features is less effective in capturing meaningful structure in the data compared to clustering using model-based features, at least in terms of partitioning based on meal composition.

Table 4.15: P-values for *k*-means clustering on CGM features using Wilcoxon Rank Sum Test of macronutrient amounts

Transformation	**FAT**	**PROT**	**CARB**	**FAT+PROT**	**H_0 Rejected**
Amount	0.075352	0.21557	0.1881	0.10071	No

These p-values are from the Wilcoxon Rank Sum test comparing the distributions of two clusters across macronutrient amount. The null hypothesis H_0 assumes identical distributions and the rejection indicates a significant difference between the medians of the two clusters.

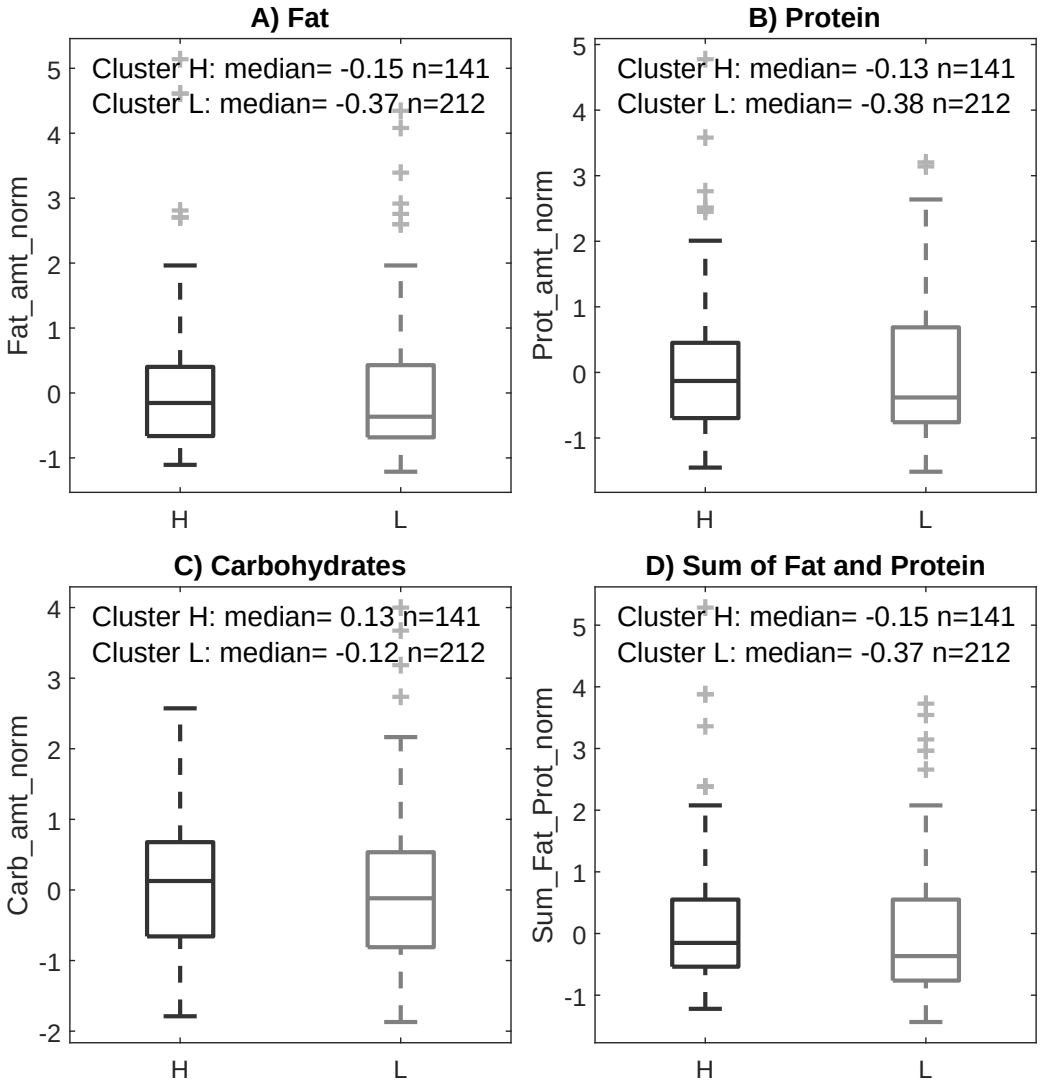

Figure 4.37: Boxplots of the distributions of the macronutrient transformations expressed as absolute amount in the two *k*-means clusters. Panel A: Fat amount. Panel B: Protein amount. Panel C: Carbohydrate amount. Panel C: Sum of Fat and Protein amount. The clusters are labeled based on the rank of their median in Cluster H, with the highest medians, and Cluster L, with the lowest medians. The median and the number of observations (n) for each cluster are also reported in the panels.

5

Discussion

Current insulin therapies are based solely on the carbohydrate amount [66, 67]. However, glucose dynamics are affected by the presence of proteins and fat. Therefore, individuals with T1D would benefit from the information regarding the macronutrient content in a meal, especially when dealing with HF/HP meals. These meals are known to delay glucose absorption and increase insulin resistance. Thus, if not accounted for, their presence can lead to early hypoglycemia and/or delayed and sustained postprandial hyperglycemia [28, 68].

The MI-OMM, a physiologically-based model that exploits minimally-invasive data acquired from CGM sensor measurements and IP data under free-living conditions, proved to describe accurately the behavior of Ra, GR and insulin sensitivity [33] and therefore could provide a set of features of physiological meaning to aid the task of meal classification.

This thesis aims to develop a robust model-based methodology to infer meal macronutrient composition, especially identifying complex meals, such as HF/HP meals, that are characterized by special insulin requirements. In addition, this work aims to demonstrate the capability of the DWB to effectively manage such meals, thereby improving glycemic control in scenarios where standard insulin strategies fail. Finally, the thesis aimed to establish that the model-based framework utilized for this work allows performing meal classification, as the extracted features are directly linked to mechanistic physiological processes, providing both predictive power and interpretability.

5.1 Development of the Methodology

In the first part of this work, we developed an algorithm to detect challenging meals, using a set of features derived either from the identification of the MI-OMM or directly from CGM profiles. These features were selected based on their potential relevance in predicting macronutrient composition.

As a preliminary step, we conducted a correlation analysis using both Spearman's rank correlation and repeated measures correlation, r_{rm}. The analysis of Spearman's rank correlation among the extracted features revealed strong associations between adjacent TGRs and AUC_{Ra}s, indicating substantial mutual dependence.

Correlations between extracted features and macronutrient transformations were generally modest in both Spearman's rank and r_{rm} analyses. The strongest associations were observed between AUC_{Ra}s and TGRs with carbohydrate amounts, whereas correlations with fat and protein amounts were weaker. Transformations expressed as percentages showed the weakest correlations overall.

Repeated measures correlation suggested intra-individual relationships between MI-OMM features, particularly of AUC_{Ra}s and TGRs with macronutrient amounts. Specifically, AUC_{Ra}s tended to decrease with increasing macronutrient amount, whereas TGRs exhibited the opposite trend. These findings are consistent with patterns reported in the literature, as in [27, 32, 65].

Moreover, our results support the observations of Ekhlaspour and coworkers [46], where $\dot{G}_i$(45min) and $\dot{G}_i$(120min) showed to decrease and increase, respectively, when increasing the amount of fat or protein in a meal. Furthermore, $\dot{G}_i$(45min) and $\dot{G}_i$(120min) reported among the highest r_{rm} values for these two macronutrients.

Interestingly, the repeated measures correlation between S_I and macronutrients was positive despite the negative Spearman's rank correlation. One plausible explanation is that Spearman's rank, which assesses monotonic relationships across all subjects, captured a different association pattern than r_{rm}, which focuses on within-subject variation. This suggests that r_{rm} may have detected intra-individual trends not visible in the Spearman analysis, or vice versa, that the relation is not linear and thus does not meet the assumptions required by r_{rm}. This behavior needs further investigation.

Following the correlation analysis, we assessed the predictive performance of LME models. Two LME structures were considered: models with random inter-

cepts only, and models with both random intercepts and slopes. In agreement with the preliminary results, the best-performing macronutrient transformations were those expressed in absolute amounts (fat amount, protein amount, and their sum), whereas percentage-based transformations performed poorly. In many of the best performing models, AUC_{Ra}s and TGRs features were consistently selected according to ΔLL, reinforcing the assumption of their potential relevance for meal classification.

Surprisingly, $\dot{G}_i$(45min) and $\dot{G}_i$(120min) were not included in all top LME models, despite high r_{rm} values, appearing mainly in fat-related models. Given that r_{rm} can be represented as an LME model with a single predictor and random intercept, these features were expected to appear more frequently. Their exclusion likely reflects the fact that LME models rely on multiple predictors, and, despite the potential better fit that can be achieved, the contribution of other features, such as $\dot{G}_i$(45min) and $\dot{G}_i$(120min) may be insufficient to reduce AIC due to the increased model complexity. This is more plausible when dealing with complex models with a high number of predictors, therefore distancing the LME models from the r_{rm} conceptual structure. This behavior was confirmed in cases where the best LME models contained only one predictor, such as the best models with random intercept and slope related to fat-percentage transformations, where $\dot{G}_i$(45min) was indeed selected.

When including both random intercepts and slopes, LME models generally achieved better ΔLL for macronutrient amounts, despite incorporating fewer predictors. This is likely reflecting the AIC penalty for higher model complexity, since the increased number of random effects causes the capturing of inter-individual variability. Therefore, the inclusion of more features, especially fixed effects, was redundant, since the random effects already absorbed the majority of the variance.

Across these models, AUC_{Ra}(120min), TGR(50%), and G_s(0min) were consistently selected, again aligning with their high correlations in the preliminary analysis, especially regarding the potential role of AUC_{Ra}s and TGRs as important features for the meal classification task.

Regarding predictive performance, the LME models, both with random intercepts only and random intercepts and slopes, were not able to accurately predict macronutrient transformations treated as continuous outcomes. The model predictions, including subject-specific random effects to individualize predictions, did not result in satisfactory performance. Only a slight improvement

was visible from the residual distribution for the best LME model with random intercepts and slopes for protein amount. Using fixed-effects-only predictions, i.e., at the population level, the LME models performed only marginally better than the null (intercept-only) model, indicating little gain over a baseline that ignores predictors altogether.

Following the LME analysis, we assessed logistic regression models for the task of meal classification. In this framework, the continuous labels that were used in the LME models were binarized into a low-class (L-class) and high-class (H-class) categories to assess whether label discretization could improve classification performance.

Consistent with the LME results, the most effective macronutrient transformations were those expressed as absolute amounts, with the sum of fat and protein amount achieving the highest accuracy (69%) at the optimal point of the ROC curve. However, when fixing FPR at 5%, predictive performance declined significantly. The true positive counts decreased, and accuracy dropped to 62%. This decline reflects the trade-off between sensitivity and specificity. The max F1-score given on the precision-recall curve did not lead, in general, to better performances.

Compared with the top-performing LME models, the best logistic regression models incorporated a larger number of predictors. This did not yield a corresponding improvement in predictive accuracy, suggesting possible overfitting issues. The inclusion of too many predictors may have reduced the ability of the model to generalize, leading to poorer performance on unseen test data. It's interesting to observe that, even if the model selection followed a backwards stepwise approach, the improvement in AIC stopped after removing just a few features, and in the case of the model for fat caloric contribution, immediately after eliminating a single feature. This is explained by the fact that logistic regression has a fixed-effects-only structure that lacks the explanatory variance introduced with random effects in LME models. In this case, logistic regression must rely only on the deterministic and constant fixed effects, requiring a higher number of features to cope with the missing descriptive power carried by the random component in the LME case.

A key limitation emerged when employing both continuous and binarized labels is the difficulty of defining macronutrient-based meal classes in a way that is robust across individuals. Currently, no consensus exists in the literature for the classification of meals into categories such as HF/HP, HF/LP, LF/HP

or LF/LP meals, as a standardized definition has not yet been defined. Even if such meal classes were defined, inter-individual variability would remain a challenge. For example, a meal classified as HF for a low BW individual might not be as challenging for an individual with a high BW, still without considering other subject-specific physiological factors, such as individual diets and food exposure, that can influence the post-prandial glycemic profile of complex meals.

Moreover, the binarization of the meal classes, utilized to support the logistic regression models, raised some concerns. The label assignment based on the sample mean of each macronutrient, as done in this work, can lead to counterintuitive and misleading classifications. Two meals consumed by the same individual, with substantially different fat content, may both be classified as L-class if they both fall below the population mean, despite their substantial relative difference for that person. Similarly, individuals with higher BW may reasonably consume meals assigned to the H-class, even when some of those meals would be relatively low in fat for them. Furthermore, this definition of the labels is highly dataset-dependent. For example, considering another population with different demographics, such as with a different age distribution, can result in a completely different labeling outcome depending on the position of the sample mean. These issues highlight the difficulty of applying supervised learning to this problem, as the approach requires predefined labels that may not reflect physiologically meaningful differences.

Another concern relates to the measurement of macronutrients in the meals. Even when recorded with the support of medical personnel, these measurements are still subject to error. Consequently, the resulting labels may not only be conceptually imperfect but also uncertain, as inaccuracies in estimating the macronutrient composition may have propagated in the analysis conducted. This aspect worsens if we consider that carbohydrates, fat, and protein are pools of different macronutrients, all containing sub-categories that may act differently on the postprandial glucose excursion (e.g., simple vs complex carbohydrates, saturated vs non-saturated fats, liquid vs solid food, etc.).

Given the limited success of both LME and logistic regression models under these labeling constraints, we explored unsupervised learning approaches, which, on the contrary, did not require predefined labels. The idea is that these approaches may better capture the hidden structure or patterns in data without the biases and limitations introduced by label definition.

When transitioning to unsupervised learning approaches, feature selection

via the CV algorithm identified AUC_{Ra}(120min), TGR(50%), and $\dot{G}_i$(120min) as the feature set yielding the highest CV accuracy. The inclusion of AUC_{Ra}(120min) and TGR(50%) was expected, as these features were frequently selected in the best-performing LME models and exhibited both high Spearman's ρ and high r_{rm} values. The selection of $\dot{G}_i$(120min) was also anticipated, given that it achieved the highest r_{rm} values for fat amount, protein amount, and their sum. The presence of two model-based features among the top features highlights the potential added predictive value of incorporating physiologically-based model-derived features into the meal classification task.

For both *k*-means and hierarchical clustering, the optimal number of clusters, determined by the Silhouette method, was two. In hierarchical clustering, this result was supported by the dendrogram, which displayed two distinct groups merging at a relatively large height, with within-group cohesion evident at low merge heights.

The two clusters divided the dataset into one group with higher AUC_{Ra}(120min) and lower TGR(50%) and $\dot{G}_i$(120min), and another group with the opposite trend, i.e., a lower AUC_{Ra}(120min) and higher TGR(50%) and $\dot{G}_i$(120min). These partitions are consistent with literature expectations, as in [27, 32, 46, 65], suggesting that the clustering partitioned the observations into two different meal types. Wilcoxon Rank Sum tests confirmed significant differences between the partitioned observations in the two clusters in terms of absolute amounts of fat, protein, carbohydrates, and the sum of fat and protein, rejecting the null hypothesis of equal distributions (p-values < 0.05). In contrast, the distributions of percentage-based macronutrient transformation (both caloric and amount contributions) were not significantly different (p-values > 0.05) between clusters, in agreement with findings from both the preliminary and supervised analyses.

Visual inspection of the macronutrient amount distributions within clusters indicated that the group with higher AUC_{Ra}(120min) and lower TGR(50%) and $\dot{G}_i$(120min) represented LF/LP/LC meals (L group), whereas the other partition corresponded to HF/HP/HC meals (H group).

GMM clustering supported these results, with only ~9% of observations showing uncertain cluster membership (i.e., with a posterior probability of cluster membership between 0.3 and 0.7). This suggests that the dataset could be partitioned robustly. Further, the 5-fold unsupervised CV confirmed the applicability of clustering methods: *k*-means achieved 100% CV accuracy in every fold across all macronutrients expressed in absolute amount, followed by GMM

(99.7%) and hierarchical clustering (89%).

Nevertheless, some limitations should also be acknowledged. First, clustering results in unsupervised learning cannot be validated against ground truth labels. In this work, ground truth labels were defined based on the rank of the medians of macronutrient distributions within each of the clusters using all the available observations for the training process. This label assignment, despite being plausible, cannot be externally validated to prove its correctness. Furthermore, the unsupervised CV algorithm was designed to assign cluster labels retrospectively, based on the median values of macronutrient transformations within each cluster generated from the training set. It then evaluated whether the left-out test observations were assigned the same cluster labels as those in the reference clustering used to define the ground truth. This process solely validates internal cluster consistency and in no case can substitute an external validation. Thus, for *k*-means, a 100% CV accuracy simply proves that the same label assignment is recovered when training on a subset of the data rather than using all observations, not that the labels are correctly corresponding to physiologically meaningful meal classes.

Second, even assuming that the labels assigned to each cluster were precise and univocally linked to a modification in the postprandial glycemic excursion, the physiological interpretation of the resulting meal types is not fully aligned with the original classification aim. Actually, the identified clusters correspond to HF/HP/HC and LF/LP/LC meals, reflecting a separation between complex and simple meals, rather than differentiating meals based on the fat and protein content. Consequently, classification tasks seeking to isolate fat and protein differences at fixed carbohydrate intake may suffer from reduced accuracy and robustness. Nonetheless, in view of applying this methodology in real life, the ability to detect a complex meal rather than estimating its exact composition is yet a promising achievement.

Despite these limitations, the clustering approaches demonstrated greater practical utility than the supervised techniques in this context, mainly because they did not require predefined labels, which proved ineffective for the supervised models due to measurement uncertainties and inter-individual variability.

5.2 In Silico Assessment of the Methodology

The most successful meal classification methodology, *k*-means clustering, was assessed *in silico* using simulated meals from Datasets 2a-d, comprising 100 nominal and 100 HF/HP meals each. This validation task was non-trivial and revealed several methodological issues due to the simulated environment and the proper generation of synthetic meals.

To enable a fair comparison, observations from Dataset 2a-d were z-scored using the mean and standard deviation from Dataset 1 before being projected onto the *k*-means feature space. Upon projection, the simulated observations clustered predominantly around the L-centroid, with minimal spread around the H-centroid. This distribution would clearly result in an erroneous classification, as both nominal and HF/HP meals would mostly be assigned to the L group, without making any intervention on the insulin therapy.

The underlying cause was identified as a mismatch between real and simulated meal dynamics. Simulated meals showed to have significantly faster dynamics than real meals because they were generated from data collected during a mixed meal tolerance test, performed in a hospitalized setting, and consisting of an ingestion of a standardized jell-O meal [69]. This mismatch likely prevented the simulated meals from aligning appropriately with the real meal centroids. To mitigate this, an *ad hoc* affine transformation was applied to the simulated meals to achieve the alignment expected for real meals. Nevertheless, such misalignment hampered the use of the simulated environment for our purpose.

In addition, and in particular with the simulated meals of Dataset 2a, the interstitial glucose derivative at 45 min, $\dot{G}_i(45\text{min})$, showed an opposite trend to that observed in real meals and reported in [46]: instead of decreasing for HF/HP meals relative to nominal meals, it increased. This potentially useful feature became unusable for our validation purposes. This encouraged the generation of HF/HP meals with progressively slower dynamics (Datasets 2b-d).

After addressing these limitations, HF/HP meals clustered, on average, near the H-centroid, while nominal meals clustered, on average, near the L-centroid. Dataset 2a, which exhibited the fastest CGM profiles for HF/HP meals, achieved only modest accuracy. The best-performing dataset was Dataset 2c, where HF/HP meals were delayed with respect to nominal ones. In this case, accuracy

improved to 62%, which, however, was still unsatisfactory.

The critical issue was identified in the inter-individual variability, which was much higher than the impact of meal composition on Ra and CGM profiles. In other words, the feature trends between nominal and HF/HP meals of the same subject were consistent, i.e., AUC_{Ra}(120min) decreased while TGR(50%) and $\dot{G}_i$(45min) increased. However, the distance between nominal and HF/HP meals within the same subject was smaller than the distance between nominal and HF/HP meals across different subjects. This overlap between meal types is the main reason for the limited effectiveness of *k*-means clustering classification in the simulated scenarios.

Nevertheless, to prototype our methodology *in silico* and demonstrate the potential of correctly identifying complex meals to suggest a better-suited treatment, a DWB bolus was assigned to the meals classified as HF/HP.

Starting from Dataset 2a, when a DWB bolus was applied to HF/HP meals, TIR, TAR, TBR, and $\overline{CGM}$ improved compared to standard therapy, based on carbohydrate content only. Specifically, median TIR increased, and median TAR decreased, indicating better postprandial glycemic control. Median $\overline{CGM}$ for HF/HP meals decreased as well, while TBR remained consistently at 0% (25th-75th percentile).

The number of HE for HF/HP meals increased. However, this increment is expected, since the DWB adjustment aims to make HF/HP meals physiologically similar to nominal meals under standard therapy, and nominal meals under standard therapy recorded a total of 13 HE. Thus, the observed similarity in CGM profiles between HF/HP meals under adjusted therapy and nominal meals treated with the standard therapy suggests that the DWB intervention was successful. Ultimately, as in [29, 30, 31], we proved, *in silico*, that the DWB insulin strategy is capable of successfully treating HF/HP meals by lowering the risks of both early hypoglycemia and delayed and sustained hyperglycemia.

Of note, the DWB insulin strategy applied in this work was not individualized. An individualized DWB strategy, ideally based on individual characteristics and clustering results, would yield better outcomes. Furthermore, since 39% of nominal meals of Dataset 2a were misclassified as HF/HP, the DWB was also applied to these meals, which resulted in the slight worsening of CGM metrics: the TIR decreased, while the number of HE increased. TAR, TBR and $\overline{CGM}$, on the other hand, remained nearly unchanged.

When repeating the simulation in Dataset 2c, further issues emerged. In

HF/HP meals, TIR was higher with the standard than with the DWB. This behavior contradicts expectations, as HF/HP meals are generally more susceptible to early hypoglycemia and/or delayed and sustained hyperglycemia [28], which should result in a lower TIR when treated with carbohydrate-based insulin therapy. A second issue was observed when applying the adjusted therapy to meals classified as HF/HP. Although Dataset 2a demonstrated the efficacy of the DWB strategy for treating HF/HP meals, in Dataset 2c the HF/HP meals exhibited excessively accentuated characteristics, leading to a decrease in TIR between standard and adjusted therapy.

A further concern regards the suitability of these CGM metrics in evaluating the outcomes of a therapy acting on a single meal scenario. Although they are widespread and consolidated [64], they seem inappropriate to characterize a single postprandial CGM profile. In future studies, we will assess the use of other metrics, such as the time to peak or decrease from baseline [31].

Taken together, these results suggested that the current simulator is inadequate to support the assessment of the meal classification algorithm and underscore the need for future improvements to the UVa/Padova T1D simulator, with the aim of generating meals with more physiologically plausible characteristics.

5.3 Contribution of Model-derived Features

In this work, some of the features are derived from model parameters obtained after the numerical identification of the MI-OMM. The relevance of these features and the potential added predictive value given to the task of meal classification was clear from the preliminary analysis as especially the AUC_{Ra}s and TGRs showed high correlation with macronutrients. The important relevance of these model-derived features was then confirmed as they were often included in the best LME models and in the clustering feature selection process.

To further demonstrate the utility of this physiologically-based model-informed approach, the most successful classification method, i.e., the *k*-means clustering, was applied to a subset of features extracted from CGM measurements only: CGM(0min), $\dot{CGM}$(45min) and $\dot{CGM}$(120min). These features were chosen as direct counterparts of the MI-OMM–derived CGM-based features, G_s(0min), $\dot{G}_i$(45min) and $\dot{G}_i$(120min).

Despite the MI-OMM counterparts, $\dot{G}_i$(45min) and $\dot{G}_i$(120min), exhibiting

high repeated measure correlations with macronutrient intake, clustering exploiting CGM(0min), CĠM(45min) and CĠM(120min) yielded noisy and poorly separable groups. The resulting clusters lacked a clear structure, and the Wilcoxon Rank Sum test failed to detect significant differences in macronutrient composition (fat, protein, and carbohydrate amounts, or the sum of fat and protein amounts). Visual inspection of macronutrient distributions confirmed this finding, as the boxplots of the two clusters showed highly overlapping distributions.

These results highlight an important insight into the power of this work. When relying solely on raw CGM-derived features, meal classification becomes substantially more challenging. Neither adding more CGM-based features, such as peak values, peak times, or other CGM derivatives, allowed to improving the clustering performance, especially with unsupervised learning. On the contrary, increasing dimensionality may hide physiologically meaningful structures and drive clustering toward data patterns without direct meal type interpretations. In contrast, features derived from the MI-OMM encapsulate part of the mechanistic knowledge of glucose absorption and insulin kinetics, providing more robust and physiologically-based meal classification. This, likely, helped the clustering algorithm to find meaningful partitions of the data successfully representing different meal classes.

5.4 Conclusions

In conclusion, this thesis presented an explorative analysis of machine learning approaches for inferring meal macronutrient composition, exploiting features derived from the MI-OMM parameters. The results showed that unsupervised learning techniques, particularly k-means clustering, can successfully partition meals into two groups: those with high content and those with lower content of fat, protein, and carbohydrates. This partitioning aligns with existing literature, as complex meals, such as HF/HP meals, are characterized by slower gastric emptying, delayed Ra, and a higher interstitial glucose derivative 120 minutes from mealtime. The clinical relevance of this classification was demonstrated by showing that assigning to complex meals an adjusted insulin therapy, the DWB, improves glycemic control. Overall, the model-based framework developed in this work demonstrated the potential of leveraging physiologically

meaningful features for meal classification. By providing a methodology to infer meal macronutrient composition, this work provides the basis for moving towards a more suitable insulin therapy that accounts not only for carbohydrate content, but also for the effects of fat and protein, potentially without increasing the burden on the patient, which are both goals of growing importance in diabetes research.

6

Conclusions

Despite major technological advances in T1D treatment through minimally invasive devices such as CGM and IP, a large proportion of patients still fail to achieve recommended glycemic targets, largely due to suboptimal postprandial glucose control [25]. Current insulin dosing strategies are primarily based on carbohydrate counting [7], which, although simple, neglects the substantial effects of fat and protein on glucose dynamics. These macronutrients are known to delay glucose absorption and increase insulin resistance, thereby predisposing patients to early hypoglycemia and delayed, sustained hyperglycemia if not properly accounted for [27, 28].

Several studies have shown that DWB delivery can improve outcomes for complex meals [32]. However, a robust methodology for inferring meal macronutrient composition from minimally invasive data under free-living conditions remains lacking.

In this thesis, we addressed this gap by exploiting the MI-OMM, a physiologically based model capable of estimating Ra, GR, and insulin sensitivity from CGM and IP data [33]. We demonstrated that features derived from MI-OMM parameters provide physiologically meaningful information that significantly enhances the classification of meal macronutrient composition. Finally, the utility of meal classification in adjusting T1D insulin therapy was explored *in silico* through the application of the DWB strategy to HF/HP simulated meals.

6.1 Summary of the Findings

This work proposed and evaluated a methodology to infer meal macronutrient composition from features derived either directly from CGM profiles or through identification of the MI-OMM. This thesis project demonstrated that model-derived features, particularly the AUC of Ra, AUC_{Ra}s and times of the GR curve, TGRs, showed an important predictive role across correlation analysis (Spearman's and repeated measures), LME models, and clustering techniques. In contrast, features based on CGM profiles showed less pronounced discriminative power, even if the interstitial glucose derivatives, $\dot{G}_i(45min)$ and $\dot{G}_i(120min)$, showed high repeated measures correlation.

The supervised learning techniques tested, LME and logistic regression models, achieved only modest predictive performance. Their effectiveness was limited by meal class label definition. Which were found both when employing the continuous nature of macronutrient outcomes in LME models, and also when defining physiologically meaningful categorical labels representing meal classes for logistic regression. Furthermore, logistic regression showed issues of overfitting potentially caused by the inclusion of too many predictors.

Across all analyses, the most informative macronutrient transformation was consistently the representation based on absolute amounts (i.e., fat, protein, and carbohydrate amounts), which outperformed percentage-based transformations (e.g., caloric contribution or amount contribution expressed as percentage of the total sum of fat, protein and carbohydrates intake). Amount-based transformations yielded the highest correlation values, were systematically preferred by LME and logistic regression models, and were also aligned with patterns identified through unsupervised methods (*k*-means, hierarchical clustering, and GMM), which consistently partitioned meals according to macronutrient amounts.

In general, unsupervised methods, particularly *k*-means clustering, provided more robust results by circumventing the need for predefined labels. The optimal feature set, comprising $AUC_{Ra}(120min)$, TGR(50%), and $\dot{G}_i(120min)$, consistently separated observations into two physiologically interpretable groups referring to two meal classes, corresponding to LF/LP/LC and HF/HP/HC meals.

Validation on simulated datasets generated with the UVa/Padova T1D simulator, despite the several challenges encountered, further indicated the potential

clinical utility of this approach in trying to infer meal macronutrient composition, especially in assigning an improved insulin strategy such as the DWB to the detected HF/HP meals, that showed to improve the postprandial glycemic curve by increasing the time spent within the safety range (70-180 mg/dl).

Overall, the findings highlight the added value of incorporating mechanistic, model-derived features into the classification tasks, offering a better descriptive physiologically-based methodology for meal characterization compared with a non-model-based approach exploiting exclusively CGM-derived features.

6.2 Limitations

Several limitations should be acknowledged. First, the supervised learning techniques tested relied on labels that were both conceptually imperfect and, in the case of logistic regression, dataset-dependent. The use of sample mean–based threshold for meal class binarization introduced inconsistencies across individuals, and inter-individual variability remained a challenge in defining meal classes such as HF/HP and LF/LP. The lack of universally accepted definitions for macronutrient-based meal categories further complicated the construction and definition of labels.

Second, the macronutrient estimation of meals, even when recorded with the support of medical personnel, was subject to measurement errors. These uncertainties propagated into the supervised learning analyses, as model training could be affected by imprecise macronutrient values, thereby reducing reliability and adding an obstacle in the already challenging task of meal classification label definition.

Third, while clustering approaches yielded encouraging results, they could not be externally validated in the absence of ground-truth labels. As previously noted, such labels are difficult, or even impossible, to define. The clustering validation procedure adopted in this work therefore assessed only internal clustering consistency, rather than verifying the correctness of meal partitioning. Moreover, the resulting clusters tended to distinguish broad meal types (i.e., HF/HP/HC versus LF/LP/LC), rather than isolating the specific contribution of fat and protein at a fixed carbohydrate intake, which would have been the more desirable outcome.

Fourth, validation of k-means clustering, trained on real meals, with sim-

ulated meals was constrained by discrepancies between *in silico* and real dynamics. Simulated meals exhibited overall a faster absorption kinetics with respect to real meals, and attempts to mitigate this by artificially reducing the rate of glucose absorption did not solve all the issues. Specifically, the reduced intra-individual variability in the feature set, compared with inter-individual variability, made impossible to obtain accurate classification.

Finally, the proposed insulin intervention strategy (DWB) was not individualized in this work. Individualized strategies could provide more accurate treatment adaptation. Furthermore, meal misclassification, particularly when nominal meals were incorrectly classified and labeled as HF/HP, resulted in inappropriate therapy assignment and a consequent deterioration of CGM-based control metrics.

6.3 Future Developments

Several directions and future prospects could emerge from this pivotal work.

First, the *in silico* validation of the clustering methodology remains to be completed. Improving the dynamics of HF/HP simulated meals generated with the UVA/Padova T1D simulator is a critical step, particularly to reduce the mismatch between inter-individual and intra-individual variability that currently is making difficult to get an accurate meal type classification. Enhancing simulator accuracy will strengthen the reliability of the *in silico* validation. Furthermore, the improvement of the simulator could help in developing better insulin strategies or improving the current DWB to address HF/HP meals insulin requirements.

Second, *in vivo* validation represents the natural next step. A clinical trial, involving individuals with T1D under SAP therapy or hybrid AP systems consuming standardized LF/LP and HF/HP meals, could be designed. During the trial, minimally-invasive data, namely the CGM postprandial profiles, the IP infusion data and SMBG measurements will be gathered. By applying the MI-OMM and extracting the same feature set used in the current work, the *k*-means clustering methodology could be validated under real-world conditions.

Third, beyond offline validation, the methodology could be implemented online in individuals with T1D under SAP therapy or AP system integrated with MPC algorithm. In this ideal system, the MPC would forecast postpran-

dial glucose trajectories within the prediction horizon, the MI-OMM would be identified with the inclusion of predicted CGM profile other than the previous information, and the set of features extracted in real time. Each meal could then be projected into the trained *k*-means feature space to determine its proximity to the H- or L-centroid. Based on this classification, a DWB strategy could be administered, reducing the risk of delayed hyperglycemia associated with HF/HP meals.

Taken together, these steps outline a potential route that started from methodology development, carried out in this thesis work, through the imminent next step of improving simulation and generation of meals with different macronutrient amounts, and reaching clinical validation, toward the final objective of real-time application.

This thesis represents the first step in this pipeline. By providing a methodology to infer meal macronutrient composition, it establishes the basis for moving towards more suitable insulin therapy that accounts not only for carbohydrate content, but also for the effects of fat and protein. This represents a goal that is currently highly requested in the field of diabetes research.

References

[1] American Diabetes Association. "Diagnosis and classification of diabetes mellitus". In: *Diabetes Care* 33 Suppl 1.Supplement_1 (2010), S62–9.

[2] American Diabetes Association Professional Practice Committee. "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2022". In: *Diabetes Care* 45.Supplement_1 (2021), S17–S38. ISSN: 0149-5992.

[3] Jessica Lucier and Priyanka M Mathias. "Type 1 diabetes". In: *StatPearls*. StatPearls Publishing, 2025.

[4] Richard I G Holt, J Hans DeVries, Amy Hess-Fischl, Irl B Hirsch, M Sue Kirkman, Tomasz Klupa, Barbara Ludwig, Kirsten Nørgaard, Jeremy Pettus, Eric Renard, et al. "The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)". In: *Diabetologia* 64.12 (2021), pp. 2609–2652.

[5] M Derouich and A Boutayeb. "The effect of physical exercise on the dynamics of glucose and insulin". In: *J. Biomech.* 35.7 (2002), pp. 911–917.

[6] C E Lloyd, P H Dyer, R J Lancashire, T Harris, J E Daniels, and A H Barnett. "Association between stress and glycemic control in adults with type 1 (insulin-dependent) diabetes." In: *Diabetes Care* 22.8 (1999), pp. 1278–1283. ISSN: 0149-5992.

[7] American Diabetes Association. "6. Glycemic Targets: Standards of Medical Care in Diabetes—2021". In: *Diabetes Care* 44.Supplement_1 (2020), S73–S84. ISSN: 0149-5992.

[8] *Type 1 diabetes: Hyperglycemia and hypoglycemia in type 1 diabetes*. Institute for Quality and Efficiency in Health Care (IQWiG), 2021.

[9] American Diabetes Association Professional Practice Committee. "12. Retinopathy, Neuropathy, and Foot Care: Standards of Care in Diabetes—2025". In: *Diabetes Care* 48.Supplement_1 (2024), S252–S265. ISSN: 0149-5992.

[10] Cari Berget, Laurel H Messer, and Gregory P Forlenza. "A clinical overview of insulin pump therapy for the management of diabetes: Past, present, and future of intensive therapy". In: *Diabetes Spectr.* 32.3 (2019), pp. 194–204.

[11] Novo Nordisk A/s. *NovoPen 6 - User Manual*. `https://www.novonordisk.com.au/content/dam/nncorp/au/en/pdfs/au-novo-pen-6-user-guide.pdf`. Product photograph. Last Accessed: 02-09-2025.

[12] Insulet Corp. *Omnipod DASH Insulin Management System - Product Image*. `https://www.omnipod.com/en-gb/what-is-omnipod/omnipod-dash`. Official product images and media. Last Accessed: 02-09-2025.

[13] Andrej Janež, Cristian Guja, Asimina Mitrakou, Nebojsa Lalic, Tsvetalina Tankova, Leszek Czupryniak, Adam G Tabák, Martin Prazny, Emil Martinka, and Lea Smircic-Duvnjak. "Insulin therapy in adults with type 1 diabetes mellitus: A narrative review". In: *Diabetes Ther.* 11.2 (2020), pp. 387–409.

[14] Ascensia Diabetes Care AG (formerly Bayer). *CONTOUR TS Blood Glucose Meter — Product Information and User Manual*. `https://medaval.ie/docs/manuals/Bayer-Contour-TS-Manual.pdf`. Product photograph. Last Accessed: 02-09-2025.

[15] Dexcom Inc. *Dexcom G7 Continuous Glucose Monitoring (CGM) System - Product Image*. `https://www.dexcom.com/en-us/g7-cgm-system`. Official product images and media. Last Accessed: 02-09-2025.

[16] Kranti Shreesh Khadilkar, Tushar Bandgar, Vyankatesh Shivane, Anurag Lila, and Nalini Shah. "Current concepts in blood glucose monitoring". In: *Indian J. Endocrinol. Metab.* 17.Suppl 3 (2013), S643–9.

[17] Thomas K Mathew, Muhammad Zubair, and Prasanna Tadi. "Blood glucose monitoring". In: *StatPearls*. StatPearls Publishing, 2025.

[18] S F Clarke and J R Foster. "A history of blood glucose meters and their role in self-monitoring of diabetes mellitus". In: *Br. J. Biomed. Sci.* 69.2 (2012), pp. 83–93.

[19] Richard M. Bergenstal, William V. Tamborlane, Andrew Ahmann, John B. Buse, George Dailey, Stephen N. Davis, Carol Joyce, Tim Peoples, Bruce A. Perkins, John B. Welsh, et al. "Effectiveness of Sensor-Augmented Insulin-Pump Therapy in Type 1 Diabetes". In: *New England Journal of Medicine* 363.4 (2010), pp. 311–320.

[20] Thomas Peyser, Eyal Dassau, Marc Breton, and Jay Skyler. "The artificial pancreas: Current status and future prospects in the management of diabetes". In: *Annals of the New York Academy of Sciences* 1311 (Apr. 2014).

[21] Garry Steil, Kerstin Rebrin, Christine Darwin, Farzam Hariri, and Mohammed Saad. "Feasibility of Automating Insulin Delivery for the Treatment of Type 1 Diabetes". In: *Diabetes* 55 (2007), pp. 3344–50.

[22] Su Lim Kang, Yoo Na Hwang, Ji Yean Kwon, and Sung Min Kim. "Effectiveness and safety of a model predictive control (MPC) algorithm for an artificial pancreas system in outpatients with type 1 diabetes (T1D): systematic review and meta-analysis". In: *Diabetol. Metab. Syndr.* 14.1 (2022), p. 187.

[23] Michael Tsoukas, Dorsa Majdpour, Jean-Francois Yale, Anas El Fathi, Natasha Garfield, Joanna Rutkowski, Jennifer Rene, Laurent Legault, and Ahmad Haidar. "A fully artificial pancreas versus a hybrid artificial pancreas for type 1 diabetes: a single-centre, open-label, randomised controlled, crossover, non-inferiority trial". In: *The Lancet Digital Health* 3 (2021).

[24] Iván Contreras and Josep Vehí. "Artificial Intelligence for Diabetes Management and Decision Support: Literature Review". In: *Journal of Medical Internet Research* 20 (2018).

[25] American Diabetes Association. "Postprandial Blood Glucose". In: *Diabetes Care* 24.4 (2001), pp. 775–778. ISSN: 0149-5992.

[26] Adnan Jafar and Melissa-Rosina Pasqua. "Postprandial glucose-management strategies in type 1 diabetes: Current approaches and prospects with precision medicine and artificial intelligence". In: *Diabetes Obes. Metab.* 26.5 (2024), pp. 1555–1566.

[27] Carmel E.M. Smart, Bruce R. King, and Prudence E. Lopez. "Insulin Dosing for Fat and Protein: Is it Time?" In: *Diabetes Care* 43.1 (2019), pp. 13–15. ISSN: 0149-5992.

[28] Kirstine J. Bell, Carmel E. Smart, Garry M. Steil, Jennie C. Brand-Miller, Bruce King, and Howard A. Wolpert. "Impact of Fat, Protein, and Glycemic Index on Postprandial Glucose Control in Type 1 Diabetes: Implications for Intensive Diabetes Management in the Continuous Glucose Monitoring Era". In: *Diabetes Care* 38.6 (2015), pp. 1008–1015. ISSN: 0149-5992.

[29] H P Chase, S Z Saib, T MacKenzie, M M Hansen, and S K Garg. "Postprandial glucose excursions following four methods of bolus insulin administration in subjects with type 1 diabetes". In: *Diabet. Med.* 19.4 (2002), pp. 317–321.

[30] Ewa Pańkowska, Piotr Ładyżyński, Piotr Foltyński, and Karolina Mazurczak. "A randomized controlled study of an insulin dosing application that uses recognition and meal bolus estimations". In: *J. Diabetes Sci. Technol.* 11.1 (2017), pp. 43–49.

[31] Kirstine J Bell, Elena Toschi, Garry M Steil, and Howard A Wolpert. "Optimized mealtime insulin dosing for fat and protein in type 1 diabetes: Application of a model-based approach to derive insulin doses for open-loop diabetes management". In: *Diabetes Care* 39.9 (2016), pp. 1631–1634.

[32] Mariam Metwally, Tin Oi Cheung, Roslyn Smith, and Kirstine J Bell. "Insulin pump dosing strategies for meals varying in fat, protein or glycaemic index or grazing-style meals in type 1 diabetes: A systematic review". In: *Diabetes Res. Clin. Pract.* 172.108516 (2021), p. 108516.

[33] Edoardo Faggionato, Michele Schiavon, Laya Ekhlaspour, Bruce A Buckingham, and Chiara Dalla Man. "The minimally-invasive Oral glucose Minimal Model: Estimation of gastric retention, glucose rate of appearance, and insulin sensitivity from type 1 diabetes data collected in real-life conditions". In: *IEEE Trans. Biomed. Eng.* 71.3 (2024), pp. 977–986.

[34] Roberto Visentin, Enrique Campos-Náñez, Michele Schiavon, Dayu Lv, Martina Vettoretti, Marc Breton, Boris P Kovatchev, Chiara Dalla Man, and Claudio Cobelli. "The UVA/Padova Type 1 diabetes simulator goes from single meal to single day". In: *J. Diabetes Sci. Technol.* 12.2 (2018), pp. 273–281.

[35] Jennifer L. Sherr, Bruce A. Buckingham, Gregory P. Forlenza, Alfonso Galderisi, Laya Ekhlaspour, R. Paul Wadwa, Lori Carria, Liana Hsu, Cari Berget, Thomas A. Peyser, et al. "Safety and Performance of the Omnipod

Hybrid Closed-Loop System in Adults, Adolescents, and Children with Type 1 Diabetes Over 5 Days Under Free-Living Conditions". In: *Diabetes Technology & Therapeutics* 22.3 (2020), pp. 174–184.

[36] Edoardo Faggionato, Michele Schiavon, and Chiara Dalla Man. "Simulation of High-Fat High-Protein Meals Using the UVA/Padova T1D Simulator". In: *IFAC-PapersOnLine* 59 (2025), pp. 103–108.

[37] Martina Vettoretti, Andrea Facchinetti, Giovanni Sparacino, and Claudio Cobelli. "Type-1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments". In: *IEEE Trans. Biomed. Eng.* 65.6 (2018), pp. 1281–1290.

[38] Chiara Dalla Man, Michael Camilleri, and Claudio Cobelli. "A system model of oral glucose absorption: validation on gold standard data". In: *IEEE Trans. Biomed. Eng.* 53.12 Pt 1 (2006), pp. 2472–2478.

[39] C. Della Man, A. Caumo, and C. Cobelli. "The oral glucose minimal model: Estimation of insulin sensitivity from a meal test". In: *IEEE Transactions on Biomedical Engineering* 49.5 (2002), pp. 419–429.

[40] Claudio Cobelli and Ewart Carson. "8 - Parametric models—the estimation problem". In: *Introduction to Modeling in Physiology and Medicine (Second Edition)*. Ed. by Claudio Cobelli and Ewart Carson. Second Edition. Academic Press, 2019, pp. 227–231. ISBN: 978-0-12-815756-5.

[41] Chiara Dalla Man, Robert A. Rizza, and Claudio Cobelli. "Meal Simulation Model of the Glucose-Insulin System". In: *IEEE Transactions on Biomedical Engineering* 54.10 (2007), pp. 1740–1749.

[42] Edoardo Faggionato, Michele Schiavon, and Chiara Dalla Man. "Modeling between-subject variability in subcutaneous absorption of a fast-acting insulin analogue by a nonlinear mixed effects approach". In: *Metabolites* 11.4 (2021), p. 235.

[43] Andrea Facchinetti, Simone Del Favero, Giovanni Sparacino, and Claudio Cobelli. "Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices". In: *Medical Biological Engineering Computing* 53 (2015), pp. 1259–1269.

[44] Martina Vettoretti, Andrea Facchinetti, Giovanni Sparacino, and Claudio Cobelli. "A model of self-monitoring blood glucose measurement error". In: *J. Diabetes Sci. Technol.* 11.4 (2017), pp. 724–735.

[45] The MathWorks Inc. *MATLAB Version: R2024b Update 4 (24.2.0.2833386)*. Natick, Massachusetts, United States, 2024. URL: https://www.mathworks.com.

[46] Laya Ekhlaspour, Michele Schiavon, Ryan Kingman, Bruce Buckingham, and Chiara Dalla Man. "Effect of Meal Composition on Glucose Rate of Change in Patients with Type 1 Diabetes Using a Closed-Loop System". In: *Diabetes Technology & Therapeutics* 23 (2021), A96–A96.

[47] Jonathan Z Bakdash and Laura R Marusich. "Repeated measures correlation". In: *Front. Psychol.* 8 (2017), p. 456.

[48] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: https://www.R-project.org/.

[49] Jonathan Z. Bakdash and Laura R. Marusich. *rmcorr: Repeated Measures Correlation*. R package version 0.7.0. 2024. URL: https://CRAN.R-project.org/package=rmcorr.

[50] J.C. Pinheiro and D. Bates. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer, 2009. ISBN: 9781441903174.

[51] Shengxin Zhu and Andrew J Wathen. *Essential formulae for restricted maximum likelihood and its derivatives associated with the linear mixed models*. 2018.

[52] Tom Fawcett. "ROC Graphs: Notes and Practical Considerations for Researchers". In: *Machine Learning* 31 (Jan. 2004), pp. 1–38.

[53] Yutaka Sasaki. "The truth of the F-measure". In: *Teach Tutor Mater* (Jan. 2007).

[54] Jesse Davis and Mark Goadrich. "The relationship between Precision-Recall and ROC curves". In: *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 233–240. ISBN: 1595933832.

[55] David Arthur and Sergei Vassilvitskii. "K-Means++: The Advantages of Careful Seeding". In: vol. 8. Jan. 2007, pp. 1027–1035.

[56] Peter J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427.

[57] Frank Nielsen. "Hierarchical Clustering". In: 2016, pp. 195–211. ISBN: 978-3-319-21902-8.

[58] Joe H Ward Jr. "Hierarchical grouping to optimize an objective function". In: *J. Am. Stat. Assoc.* 58.301 (1963), pp. 236–244.

[59] Geoffrey J McLachlan, Sharon X Lee, and Suren I Rathnayake. "Finite mixture models". In: *Annu. Rev. Stat. Appl.* 6.1 (2019), pp. 355–378.

[60] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1 (1977), pp. 1–38. ISSN: 00359246. (Visited on 08/07/2025).

[61] Roberto Rocci, Stefano Antonio Gattone, and Roberto Di Mari. *A data driven equivariant approach to constrained Gaussian mixture modeling*. 2016.

[62] Raphael Araujo Sampaio, Joaquim Dias Garcia, Marcus Poggi, and Thibaut Vidal. "Regularization and optimization in model-based clustering". In: *Pattern Recognition* 150 (2024), p. 110310. ISSN: 0031-3203.

[63] Jörg Lücke and Dennis Forster. "k-means as a variational EM approximation of Gaussian mixture models". In: *Pattern Recognition Letters* 125 (2019), pp. 349–356. ISSN: 0167-8655.

[64] Tadej Battelino, Thomas Danne, Richard M. Bergenstal, Stephanie A. Amiel, Roy Beck, Torben Biester, Emanuele Bosi, Bruce A. Buckingham, William T. Cefalu, Kelly L. Close, et al. "Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range". In: *Diabetes Care* 42.8 (2019), pp. 1593–1603. ISSN: 0149-5992.

[65] José Manuel García-López, María González-Rodriguez, Marcos Pazos-Couselo, Francisco Gude, Alma Prieto-Tenreiro, and Felipe Casanueva. "Should the Amounts of Fat and Protein Be Taken into Consideration to Calculate the Lunch Prandial Insulin Bolus? Results from a Randomized Crossover Trial". In: *Diabetes Technology & Therapeutics* 15.2 (2013), pp. 166–171.

[66] Michele Schiavon, Chiara Dalla Man, and Claudio Cobelli. "Insulin sensitivity index-based optimization of insulin to carbohydrate ratio: In silico study shows efficacious protection against hypoglycemic events caused by suboptimal therapy". In: *Diabetes Technol. Ther.* 20.2 (2018), pp. 98–105.

[67] American Diabetes Association. "6. Glycemic Targets: Standards of Medical Care in Diabetes—2018". In: *Diabetes Care* 41.Supplement$_1$ (2017), S55–S64. ISSN: 0149-5992.

[68] Carmel E M Smart, Megan Evans, Susan M O'Connell, Patrick McElduff, Prudence E Lopez, Timothy W Jones, Elizabeth A Davis, and Bruce R King. "Both dietary protein and fat increase postprandial glucose excursions in children with type 1 diabetes, and the effect is additive". In: *Diabetes Care* 36.12 (2013), pp. 3897–3902.

[69] Rita Basu, Barbara Di Camillo, Gianna Toffolo, Ananda Basu, Pankaj Shah, Adrian Vella, Robert Rizza, and Claudio Cobelli. "Use of a novel triple-tracer approach to assess postprandial glucose metabolism". In: *American Journal of Physiology-Endocrinology and Metabolism* 284.1 (2003), E55–E69.