



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN BIOINGEGNERIA

“Identificazione automatica di pasti da dati di monitoraggio in continua del glucosio nella terapia del diabete di tipo 1”

Relatore: Prof. Andrea Facchinetti

Laureanda: Valentina Camoglio

**Correlatori: Dott. Giacomo Cappon
Dott.ssa Elena Idi**

ANNO ACCADEMICO 2023 – 2024

Data di laurea: 11/04/2024

Indice

Abstract	4
1. Il Diabete di Tipo 1	6
1.1 Caratteristiche e diagnosi	6
1.2 Gestione del diabete e monitoraggio continuo del glucosio.....	8
1.3 Digital Twin	11
1.4 Scopi e obiettivi della tesi.....	12
2. Metodi di letteratura per l'identificazione automatica dei pasti	14
2.1 Approcci alla Rilevazione dei pasti nello stato dell'arte	14
2.1.1 Approcci euristici	14
2.1.2 Approcci basati sul machine learning.....	16
2.1.3 Approcci basati sui sistemi di controllo	16
2.2 Considerazioni e limitazioni delle applicazioni attuali	17
3. Dataset e preprocessing	18
3.1 Descrizione del dataset.....	18
3.2 Preprocessing dei dati.....	19
4. Algoritmo di identificazione dei pasti	22
4.1 Architettura dell'algoritmo	22
4.1.1 Attivazione della meal flag in corrispondenza dei boli di insulina	23
4.1.2 Distinzione tra i boli prandiali e i boli correttivi	23
4.1.3 Analisi degli eventi ipoglicemici	25
4.1.4 Analisi dei picchi del segnale CGM e della sua derivata prima.....	26
4.2 Dettagli implementativi dell'algoritmo	27
5. Risultati e discussione	30
5.1 Metriche di valutazione	30
5.2 Analisi dei risultati ottenuti.....	31
5.2.1 Caso solo MDI.....	32
5.2.2 Caso solo soggetti in pompa insulinica	35
5.2.3 Caso dataset completo (pompa + MDI).....	38
5.2.4 Considerazioni finali.....	40
Conclusione	43
Ringraziamenti	45
Bibliografia	46

Abstract

Il diabete di tipo 1 è una malattia cronica autoimmune derivante dall'assenza di secrezione di insulina da parte delle cellule β del pancreas. Questo disturbo comporta oscillazioni significative della concentrazione di glucosio nel sangue, nota come glicemia, e richiede un monitoraggio continuo e una gestione accurata al fine di limitare le complicanze a breve e lungo termine. La terapia del diabete di tipo 1 si basa sull'iniezione di insulina esogena e sull'adozione di un regime dietetico controllato. L'introduzione di nuove tecnologie, come il monitoraggio continuo del glucosio e l'utilizzo di penne smart o pompe insuliniche, ha reso la gestione della malattia sempre più accessibile e personalizzata.

Tra le nuove tecniche per l'ottimizzazione e per la personalizzazione della terapia troviamo metodi di "Digital Twin", ovvero un modello computazionale che rappresenta il comportamento glicemico del paziente nel mondo reale. Questi modelli si basano sull'acquisire in maniera completa dati relativi ai pasti, ai livelli di attività fisica e ai dosaggi insulinici, in modo da considerare e valutare scenari di terapia alternativi, riducendo al minimo i rischi per il paziente e i costi degli studi clinici.

Tuttavia, nonostante i progressi tecnologici, il controllo del diabete di tipo 1 rimane una sfida, soprattutto per i pazienti più giovani. Ad esempio, molti giovani pazienti possono avere difficoltà nel seguire rigorosamente il regime terapeutico prescritto, con frequenti omissioni nell'assunzione di insulina o errori nella stima dei carboidrati, il cui conteggio è necessario per eseguire un calcolo quanto più corretto possibile della quantità di insulina prandiale da iniettare. Questi comportamenti possono portare a fluttuazioni glicemiche e ad una gestione non ottimale della malattia. Inoltre, la richiesta al paziente di memorizzare sempre manualmente i pasti rappresenta un'azione aggiuntiva di difficile attuazione e che sarebbe preferibile evitare di dover richiedere.

L'obiettivo di questa tesi è di esplorare tecniche euristiche per rilevare automaticamente i pasti in modo retrospettivo utilizzando i dati provenienti da sensori per il monitoraggio continuo della glicemia integrati con informazioni sui boli di insulina.

L'algoritmo per l'identificazione di pasti sviluppato in questa tesi si è basato sull'analisi di dati reali, provenienti dal dataset T1DEXI, appartenenti ad una popolazione di 78 adolescenti, con dati di glicemia, insulina e pasti relativi ad una o più giornate per paziente, per un totale di 123 giornate. Esso si compone di 4 moduli sequenziali: il primo modulo associa i boli di insulina ai pasti, il

secondo distingue i boli correttivi dai boli prandiali, il terzo analizza gli eventi ipoglicemici ed il quarto modulo analizza il segnale CGM e la sua derivata.

I risultati sono promettenti, in quanto si sono ottenuti livelli di precisione del 90,97% per i soggetti in MDI e livelli di sensibilità del 88,62% per i pazienti in pompa insulinica. Sviluppi futuri consistono nell'ampliare il dataset, aumentando la numerosità del campione potenzialmente includendo la popolazione pediatrica, contribuendo così a migliorare ulteriormente l'efficacia e l'applicabilità della metodologia. L'utilizzo di questa tecnologia potrebbe essere affiancato all'applicazione di Digital Twin nella pratica clinica, consentendo di ridurre il numero di azioni richieste al paziente nell'utilizzo di nuovi strumenti per lo sviluppo di terapie personalizzate.

CAPITOLO 1

Il Diabete di Tipo 1

1.1 Caratteristiche e diagnosi

Il diabete mellito di tipo 1 (*Type 1 Diabetes Mellitus*, T1DM) è una malattia autoimmune cronica caratterizzata da una totale deficienza di secrezione di insulina, causata dalla distruzione autoimmune delle cellule β delle isole pancreatiche [1]. L'insulina è un ormone essenziale nella gestione dei livelli di glucosio nel sangue e svolge un ruolo fondamentale nell'assorbimento e nell'utilizzo del glucosio da parte delle cellule. La sua carenza porta ad un aumento dei livelli di glucosio nel sangue, noto come iperglicemia, poiché senza insulina le cellule non possono adeguatamente utilizzare il glucosio come fonte di energia e si accumula nel circolo sanguigno. L'iperglicemia cronica, condizione caratterizzata da livelli elevati di zucchero nel sangue nel tempo, può causare danni prolungati e disfunzioni in vari organi cruciali come gli occhi, i reni, i nervi, il cuore e i vasi sanguigni. Questi danni possono avere conseguenze gravi sulla salute a lungo termine. I sintomi tipici dell'iperglicemia includono poliuria (aumento della minzione), polidipsia (eccessiva sete), perdita di peso non spiegata e talvolta polifagia (aumento dell'appetito) e visione offuscata [2]. Inoltre, nei casi più gravi di iperglicemia non controllata, si può sviluppare una condizione chiamata chetoacidosi diabetica (*Diabetic Ketoacidosis*, DKA), un grave squilibrio metabolico. In questa condizione, il corpo produce quantità eccessive di chetoni, derivati dall'abbattimento degli acidi grassi per produrre energia in assenza di un adeguato apporto di glucosio. La chetoacidosi diabetica è caratterizzata da sintomi quali nausea, vomito, dolore addominale e una respirazione profonda e veloce, nota come respirazione di Kussmaul. Questa condizione può diventare rapidamente pericolosa, richiedendo un intervento medico immediato per evitare gravi complicanze, inclusa la potenziale fatalità se non trattata tempestivamente [23].

Secondo la Federazione Internazionale del Diabete [3], l'8,8% della popolazione adulta nel mondo ha il diabete. Di tutti gli individui con diabete, solo il 10-15% ha il T1DM; il diabete mellito di tipo 2 (*Type 2 Diabetes Mellitus*, T2DM) è la forma più comune. Tuttavia, il T1DM è la forma più diffusa di diabete nei bambini (sotto i 15 anni), e attualmente più di 500.000 bambini vivono con questa condizione a livello globale.

La patogenesi del T1DM, schematizzata nella Figura 1, è stata ipotizzata come un processo continuo che si può suddividere in fasi legate alla comparsa degli autoanticorpi e alla progressiva distruzione delle cellule β , che porta alla disfunzione glucidica e infine ai sintomi associati all'iperglicemia. Ciò che ancora deve essere definito è l'eziologia dell'autoimmunità mirata alle cellule β , la quale probabilmente coinvolge una combinazione di fattori ambientali e genetici che innescano o permettono la risposta autoimmune contro le cellule β . Questo evento spesso si verifica anni prima dello sviluppo finale della disfunzione glucidica e dei sintomi.

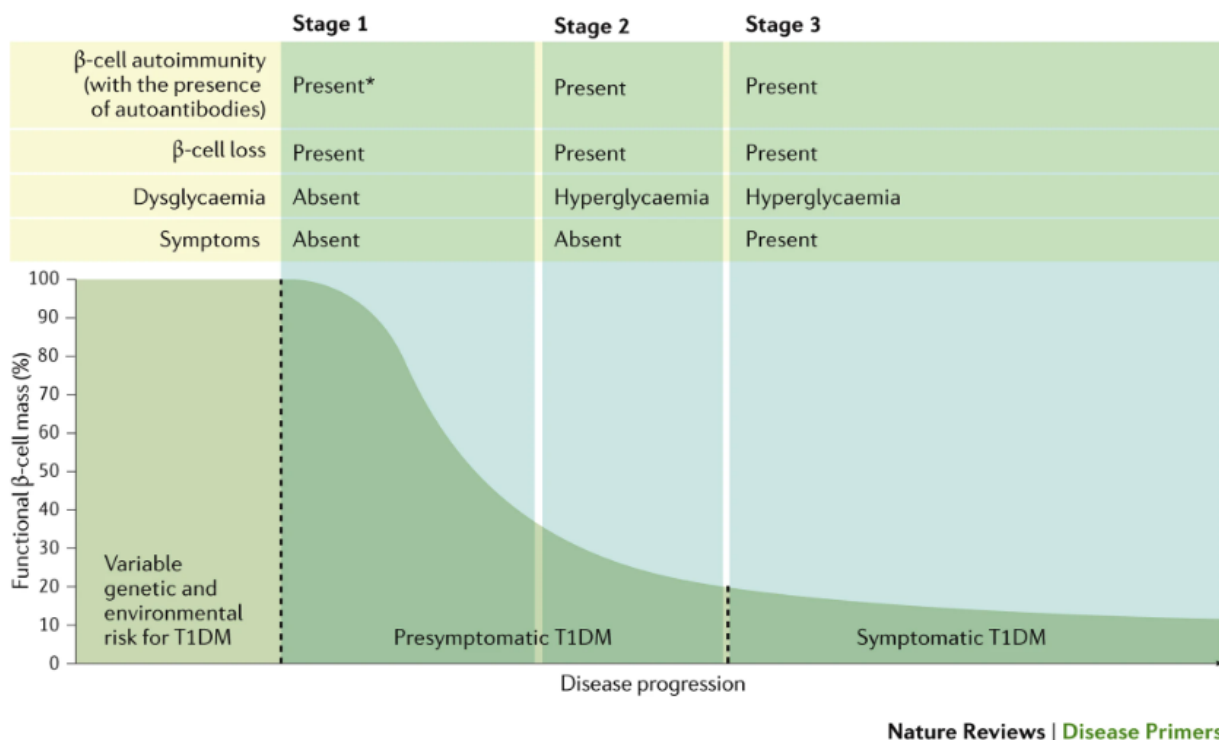


Figura 1 Il T1DM può essere suddiviso in tre fasi: la fase 1 è caratterizzata dalla presenza di autoanticorpi e dall'assenza di alterazione della glicemia; la fase 2 è caratterizzata dalla presenza sia di autoanticorpi che di alterazione della glicemia; e i sintomi compaiono solo nella fase 3, che corrisponde al T1DM sintomatico [1].

La procedura diagnostica per il T1DM coinvolge diversi passaggi essenziali. Oltre l'analisi dei sintomi tipici del diabete, elencati in precedenza, la conferma della diagnosi richiede l'esecuzione di test di laboratorio specifici. Tra questi, il test di glicemia svolge un ruolo chiave, consentendo di valutare i livelli di glucosio nel sangue. In aggiunta, la ricerca di autoanticorpi contro le cellule β pancreatiche, come gli autoanticorpi anti-insulina e anti-GAD65, può confermare la natura autoimmune del T1DM. Ulteriori valutazioni includono la misurazione dei livelli di C-peptide, che fornisce informazioni sulla produzione di insulina endogena [1]. L'integrazione di questi test consente una diagnosi precisa del diabete di tipo 1, consentendo così l'avvio di un trattamento

mirato e personalizzato per migliorare la gestione della malattia e prevenire complicanze a lungo termine.

1.2 Gestione e monitoraggio continuo del glucosio

I pazienti con T1DM richiedono terapia insulinica sostitutiva per tutta la vita a partire dal momento della diagnosi. La maggior parte dei pazienti deve seguire un regime di iniezioni multiple giornaliere (*Multiple Daily Injections*, MDI), attraverso l'utilizzo di specifiche penne per la somministrazione di boli di insulina e di insulina basale (Figura 3.b) [4]. L'insulina basale è essenziale per mantenere livelli stabili di glucosio nel sangue durante i periodi di digiuno, come durante la notte o tra i pasti. Questo tipo di insulina viene rilasciato lentamente nel corpo per contrastare la produzione di glucosio dal fegato e per mantenere l'equilibrio del glucosio nel sangue nel corso della giornata, contribuendo così a una prevenzione dell'iperglicemia. I boli di insulina contengono insulina ad azione rapida, e devono essere somministrati dal paziente prima dei pasti in modo da coprire le variazioni glicemiche postprandiali, oppure ogni volta che si desidera andare a correggere valori elevati di glicemia. Le dosi esatte di insulina variano a seconda dei casi: le dosi prandiali sono condizionate principalmente dalla quantità di carboidrati assunti; la dose di insulina basale dipende dal peso corporeo del paziente e dalla sua sensibilità all'insulina [8]. I pazienti T1D necessitano di entrambe le tipologie di insulina come riportato in Figura 2.

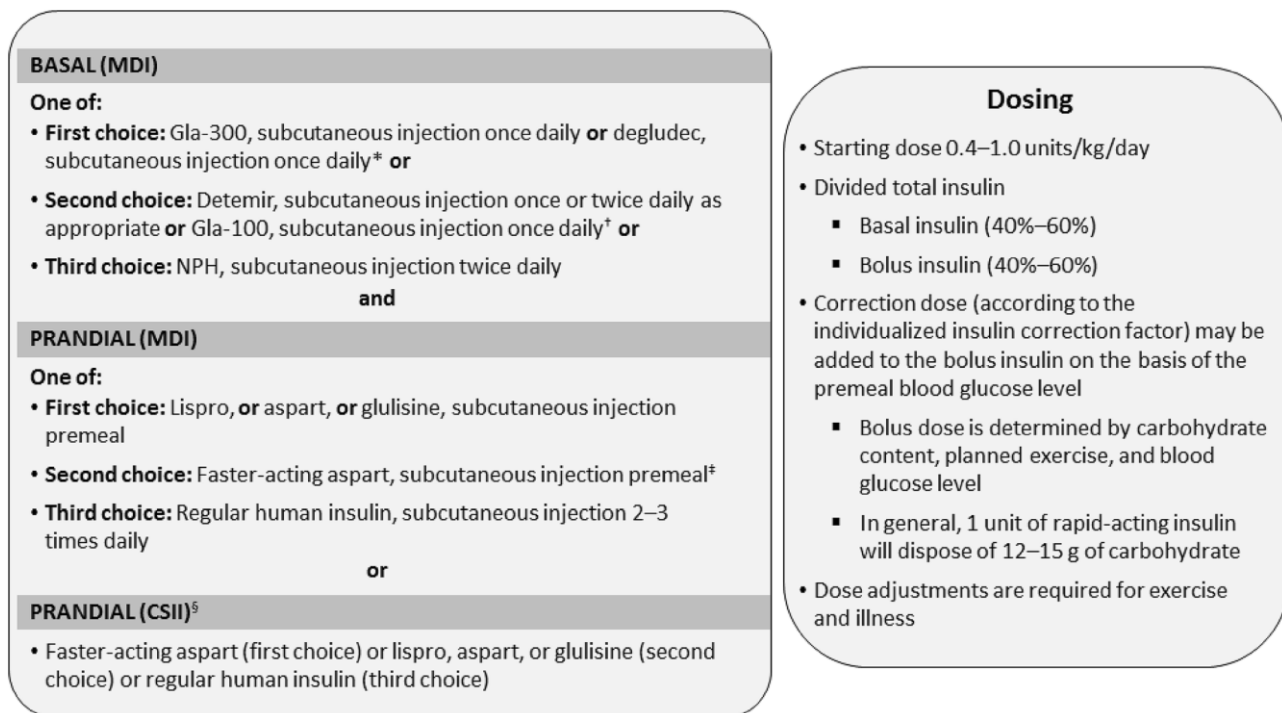


Figura 2 Esempio di trattamento per il diabete di tipo 1 suggerito per adulti non in gravidanza [4].

La misurazione frequente della glicemia e l'istruzione del paziente sull'adeguamento del dosaggio dell'insulina in base al conteggio dei carboidrati, alla glicemia pre-prandiale e all'attività fisica prevista sono fondamentali.

Secondo le linee guida del NICE (National Institute for Health and Care Excellence), si raccomanda un obiettivo di emoglobina glicata HbA1c (o emoglobina A1c) di circa il 6.5% (48 mmol/mol) per la maggior parte dei pazienti con T1DM. Tuttavia, si suggerisce di prendere in considerazione le attività quotidiane, le aspirazioni personali, la probabilità di complicazioni, le comorbilità, l'occupazione e la storia di ipoglicemia quando si stabiliscono obiettivi individualizzati [5].

L'ipoglicemia è una condizione caratterizzata dalla diminuzione eccessiva dei livelli di glucosio nel sangue. Per il suo trattamento, è essenziale intervenire tempestivamente per aumentare la glicemia. Ciò può essere ottenuto mediante il consumo rapido di carboidrati ad assorbimento veloce, oppure, nei casi più gravi, può essere richiesta l'amministrazione di glucagone, un ormone che agisce per aumentare i livelli di glucosio nel sangue.

Con l'avvento dei sistemi di monitoraggio continuo della glicemia (*Continuous Glucose Monitors*, CGM) si può ottenere una valutazione continua dei livelli di glucosio nel sangue, senza la necessità di prelievi di sangue tradizionali come il sistema di auto-monitoraggio della glicemia (*Self-Monitoring of Blood Glucose*, SMBG). L'SMBG, infatti, prevede il prelievo di una piccola goccia di sangue da parte del paziente da inserire in un glucometro, fornendo un'informazione relativa ai livelli di glucosio in quell'istante, ma non offre una visione continua dei trend glicemici nel corso

del tempo. D'altra parte, il CGM (Figura 3.a) offre un monitoraggio continuo dei livelli di glucosio nel fluido interstiziale, tramite un sensore sottocutaneo. Questo sensore misura il glucosio ad intervalli regolari, trasmettendo i dati ad un dispositivo portatile e perciò fornisce una visione in tempo reale dei livelli di glicemia e dei suoi trend nel corso del giorno. Inoltre, molti sistemi CGM offrono allarmi per segnalare ipoglicemie o iperglicemie imminenti, migliorando ulteriormente la sicurezza e la gestione della malattia. Il tempo trascorso all'interno della fascia di concentrazione della glicemia di 70-180 mg/dL (3,9-10 mmol/L) potrebbe diventare un ulteriore indicatore del controllo glicemico complessivo, sottolineando l'importanza della riduzione della variabilità della glicemia oltre ai valori medi della glicemia.

Il CGM, insieme a regimi intensivi di insulina, può aiutare i pazienti al controllo dell'emoglobina glicata HbA1c ed è utile per quei pazienti che soffrono di episodi di ipoglicemia privi di sintomi e/o frequenti.

In alternativa al regime MDI, è possibile ottenere una somministrazione continua di insulina sottocutanea (*Continuous Subcutaneous Insulin Infusion, CSII*), attraverso specifici dispositivi indossabili (anche nominati "pompe insuliniche"), che forniscono insulina basale in modo continuo e una quantità di insulina rapida personalizzabile attraverso un piccolo catetere sottocutaneo (Figura 3.c).



Figura 3 a) Sensore CGM [24]. b) Penne MDI [25]. c) Esempio di pompa insulinica [26].

La CSII sembra essere un metodo più fisiologico per erogare insulina rispetto MDI, poiché la somministrazione della basale può essere adattata per diversi intervalli temporali in base alle variazioni circadiane della sensibilità all'insulina. Questo può migliorare il controllo glicemico riducendo il rischio di ipoglicemia. Gli studi controllati randomizzati comparativi tra MDI e CSII hanno coinvolto un numero limitato di pazienti su breve periodo, ma le meta-analisi e i dati dei registri suggeriscono un lieve vantaggio per CSII per quanto riguarda i livelli di HbA1c e le frequenze di ipoglicemia grave [7]. Inoltre, rispetto a MDI, a CSII è associata a una migliore qualità della vita e maggiore soddisfazione del paziente [6].

1.3 Digital Twin

In generale, per “Digital Twin” (DT) si intende una rappresentazione virtuale che funge da controparte digitale di un oggetto fisico o di un processo [10].

Nel contesto della gestione del diabete di tipo 1, una sfida cruciale risiede, oltre che nel mantenere un controllo glicemico ottimale attraverso un monitoraggio glicemico costante e un trattamento adeguato, anche nello sviluppo e valutazione di nuove terapie. La creazione di nuove terapie è un compito critico per motivi di sicurezza dei pazienti, e può essere costoso e dispendioso in tempo. La valutazione clinica richiede inoltre di prendere in considerazione un campione sufficientemente numeroso per dimostrarne l'effettiva efficacia.

In questo contesto i trial clinici in silico (in-silico clinical trials, ISCT) hanno svolto un ruolo cruciale negli ultimi decenni, in quanto permettono di simulare vasti coorti virtuali che rappresentano la popolazione affetta da T1D su cui si possono testare nuove metodologie in modo sicuro ed economicamente vantaggioso. Inoltre, gli ISCT consentono di esplorare l'impatto dei nuovi trattamenti eseguendo numerosi test sullo stesso soggetto virtuale mantenendo le stesse condizioni identiche, il che è impossibile replicare nei trial clinici.

È in questo contesto che emerge il concetto innovativo del “Digital Twin”. Questa tecnologia offre la possibilità di creare una replica virtuale del paziente, la quale non solo descrive accuratamente i suoi processi metabolici, ma è in grado anche di simulare la risposta del paziente alle diverse terapie disponibili. Il Digital Twin del diabete di tipo 1 può essere utilizzato dai medici per considerare e valutare scenari di terapia alternative, testando ad esempio diversi regimi dietetici, livelli di attività fisica e dosaggi insulinici [27]. Questo approccio apre la strada a una gestione più personalizzata e precisa del diabete, consentendo ai pazienti di prendere decisioni informate e ai professionisti della salute di fornire trattamenti mirati.

È basata su questo concetto la metodologia di simulazione ReplayBG [11], sviluppata da Cappon et al., che consente di “riprodurre” lo scenario dietro ai dati già raccolti simulando la concentrazione di glucosio ottenuta in risposta a terapie alternative con insulina/carboidrati e valutarne l’efficacia. Questa tecnologia si articola in due fasi, come schematizzate in Figura 4: nella prima fase, viene individuato un modello personalizzato delle dinamiche glucosio-insulina, utilizzando come dati in input dati relativi all’insulina, ai carboidrati e al monitoraggio continuo della glicemia; nella seconda fase, tale modello viene impiegato per simulare la concentrazione di glucosio che si sarebbe ottenuta “riproducendo” la stessa porzione di dati attraverso l’adozione di una terapia diversa. ReplayBG elabora predizioni sulla glicemia, la cui accuratezza è strettamente legata alla qualità dei dati inseriti, come le informazioni relative ai pasti e all’insulina. Se, ad esempio, le quantità di carboidrati dei pasti sono inserite erroneamente, ciò influenzerà direttamente le stime dei parametri del modello e, di conseguenza, le previsioni sulla glicemia quando si riproducono scenari con input variati. Per risolvere questa problematica, è necessario integrare dei moduli dedicati al rilevamento e alla correzione di potenziali errori o difetti nei dati, come fase preliminare all’identificazione del modello.

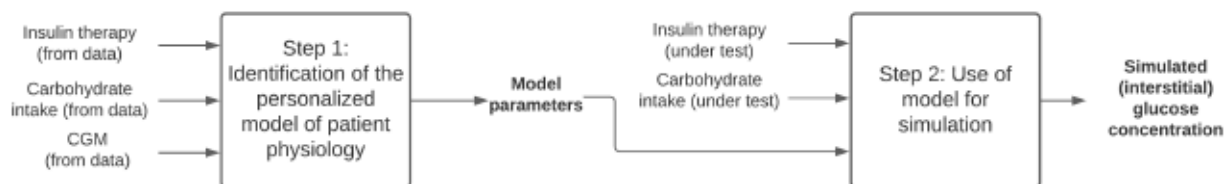


Figura 4 Overview dei due step di ReplayBG [11].

1.4 Scopi e obiettivi della tesi

La presente ricerca si concentra su una delle sfide cruciali nell’ambito della gestione del diabete: l’incertezza dei dati relativi all’assunzione di pasti da parte dei pazienti diabetici, in particolare dei pazienti in età adolescenziale. Questa incertezza, sia nella tempistica che nella quantità di carboidrati ingeriti, rende complesso il processo di inserimento di tali informazioni nei modelli di simulazione, come il sistema ReplayBG. In particolare, si osserva che soprattutto in pazienti adolescenti con pompa insulinica, oltre che a riscontrare difficoltà nell’effettuare stime precise dei pasti consumati, non sempre annunciano al sistema il pasto: la presenza di pasti non annunciati con

una frequenza di due a settimana ha un impatto significativo nell'efficacia della terapia come riportato in [28]. Nel caso invece di pazienti che seguono un regime di iniezioni multiple di insulina, i pasti assunti non vengono registrati affatto. Pertanto, senza un algoritmo dedicato per la rilevazione dei pasti, ReplayBG risulterebbe di limitato utilizzo nei pazienti in regime di MDI. L'obiettivo principale della tesi è quindi sviluppare un algoritmo basato su tecniche euristiche in grado di identificare retrospettivamente i pasti a partire dai dati ottenuti attraverso il monitoraggio continuo della glicemia e l'assunzione di insulina. Tale algoritmo consentirebbe di ottenere dati più affidabili e accurati, fondamentali per il corretto funzionamento di tecnologie come ReplayBG. Si presta particolare attenzione alla fase di rilevamento dei pasti, con l'intento di fornire un metodo affidabile e automatizzato per identificare i momenti in cui il paziente assume del cibo. Il raggiungimento di questo obiettivo rappresenta un passo cruciale verso l'ottimizzazione della gestione del diabete, consentendo ai clinici di prendere decisioni informate e personalizzate per il trattamento dei pazienti diabetici.

Capitolo 2

Metodi di letteratura per l'identificazione automatica dei pasti

2.1 Approcci alla Rilevazione dei pasti nello stato dell'arte

In letteratura, l'approccio retrospettivo all'identificazione automatica dei pasti dai dati dei pazienti affetti da diabete di tipo 1 è poco trattato, poiché attualmente vi è maggiore enfasi nello sviluppo di algoritmi per il rilevamento e l'analisi dei pasti da applicare ai pancreas artificiali (AP), che sono orientati verso un'applicazione in tempo reale degli algoritmi. Di seguito verranno riportati gli approcci più citati e più rilevanti presenti in letteratura. La maggior parte di queste procedure richiede come input campioni di glucosio provenienti da dispositivi CGM e informazioni sui boli di insulina e sull'insulina basale. Per semplicità, gli approcci presentati in questo capitolo vengono suddivisi in approcci euristici, approcci basati sul machine learning e approcci basati su sistemi di controllo.

2.1.1 Approcci euristici

Gli approcci euristici sono approcci non-esatti che possono operare sulla base delle informazioni disponibili o di conoscenze a priori e non mirano a garantire una soluzione ottimale. Tra questi, troviamo sistemi basati su regole decisionali, ovvero su soglie di derivate del glucosio (ROC del glucosio), sul controllo dell'area sotto la curva (AUC), per annunciare quando un pasto è rilevato. Uno dei lavori principali basati su questo approccio è quello Dassau et al [12]., schematizzato in Figura 5 il quale mostra l'utilizzo del calcolo del rate of change (ROC) del glucosio in varie modalità, sia sui dati grezzi che sulla stima del glucosio derivata dal filtro di Kalman, che viene poi confrontata con i valori di soglia adattati individualmente per rilevare i pasti.

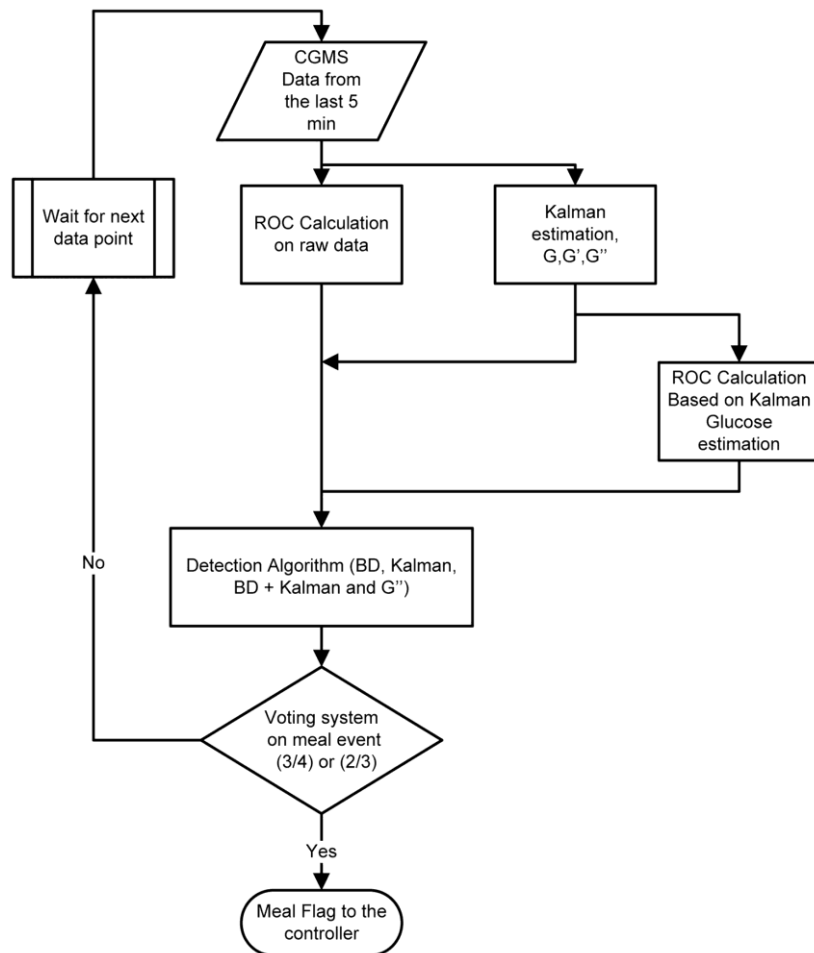


Figura 5 Schema algoritmo meal detection di Dassau et al. [12]

Un altro approccio euristico, leggermente più complesso di quello basato su regole decisionali, è quello basato sulla logica fuzzy. La logica fuzzy è un metodo che permette di gestire l'incertezza assegnando gradi di verità ai predicati anziché valori binari di vero o falso. Questo modello offre una rappresentazione più flessibile della realtà, adattandosi meglio a situazioni ambigue e complesse. Samadi et al. [13] hanno sviluppato un algoritmo per il rilevamento e la stima dei pasti integrando tre tecniche: filtro wavelet, rappresentazione qualitativa e logica fuzzy. Per il rilevamento dei pasti, hanno impiegato esclusivamente le letture CGM, mentre per stimare la quantità del pasto hanno utilizzato sia il segnale CGM che i dati sull'insulina. Il segnale CGM viene trasformato in una rappresentazione qualitativa fuzzy. Una variabile derivata dalla rappresentazione qualitativa fuzzy determina i periodi di segnalazione attiva e inattiva del pasto. Durante il periodo di segnalazione attiva del pasto, un sistema fuzzy stima incrementalmente il contenuto di carboidrati del pasto [13].

2.1.2 Approcci basati sul machine learning

Il machine learning rappresenta un sottoinsieme dell'intelligenza artificiale che sfrutta principi statistici, matematici, di programmazione e di ottimizzazione al fine di predire e classificare i dati, consentendo una migliore comprensione delle relazioni presenti all'interno di un dataset.

Nell'ambito del diabete è possibile trovare approcci di reti neurali ricorrenti (RNN) e loro sottotipi come reti LSTM (Long-Short Term Memory). Ne è un esempio il lavoro di Athanasiou et al. [14], in cui hanno sviluppato e valutato un approccio personalizzato per la rilevazione delle perturbazioni dei pasti nei pazienti con diabete mellito di tipo 1. Si sfrutta l'intrinseca capacità delle reti neurali LSTM di gestire in modo efficiente dati sequenziali all'interno di una strategia di apprendimento d'insieme (Ensemble Learning) per lo sviluppo di diverse versioni di modelli d'insieme. I modelli ricevono come input sequenze di misurazioni del monitoraggio continuo del glucosio della durata di 120 minuti e le classificano come positive o negative per l'inizio di un pasto ingerito [14]. Altri studi basati sul deep learning sono documentati in letteratura; si rimanda alla bibliografia per ulteriori dettagli [15].

Tra gli algoritmi di classificazione binaria troviamo metodi basati sull'ensemble learning [16] per problemi di classificazione o regressione, combinati con gli alberi decisionali. Nel contesto del diabete, esiste un procedimento che sfrutta modelli come il random forest e boosted tree per individuare come determinate caratteristiche, quali glucosio, frequenza cardiaca, attività fisica, temperatura corporea, temperatura della pelle e frequenza respiratoria, possano essere impiegate per creare uno strumento automatizzato per il rilevamento dei pasti in individui metabolicamente sani [17].

2.1.3 Approcci basati su sistemi di controllo

I sistemi di controllo sono utilizzati per misurare e controllare segnali o variabili comuni ad un processo fisico. La variabile viene modellata tramite equazioni nel dominio del tempo (equazioni differenziali) o nel dominio della frequenza (funzioni di trasferimento). Nell'ambito del diabete, può essere utile utilizzare sistemi di controllo che si basano sull'osservazione delle "perturbazioni" (DOBS, Disturbance Observer Based Control) del sistema: infatti, l'assunzione di un pasto può essere vista come la maggior perturbazione che altera i livelli di glucosio; in questo senso il DOBS cerca di stimare questa perturbazione e di "rifiutare" l'assunzione del pasto [18]. Altri approcci vedono l'utilizzo del filtro di Kalman unscented (UKF), utilizzato per la previsione basata sul modello minimo di Bergman [19] e per stimare parametri personalizzati. Il rilevamento dei pasti è

determinato dal valore previsto dell'interferenza del pasto e dalla differenza tra la concentrazione di glucosio nel sangue prevista dall'UKF e la concentrazione effettiva di glucosio nel sangue [20].

2.2 Considerazioni e limitazioni delle applicazioni attuali

Gli approcci descritti sopra sono stati sviluppati in primo luogo per applicazioni online e sono stati testati principalmente con dati in silico (UVA/Padova T1DM, OhioT1DM, Pima dataset [18]) piuttosto che con dati reali. Sono state presentate numerose proposte e approcci che coprono una vasta gamma di metodi. Le tecniche di rilevamento e stima dei pasti spesso si basano su regole decisionali piuttosto che su procedure più complesse: le regole decisionali risultano più praticabili per una facile implementazione, in quanto utilizzano una logica semplice e richiedono minori risorse computazionali rispetto agli approcci più avanzati. Pertanto, questa tesi si è concentrata sulla progettazione di un algoritmo offline basato su regole decisionali per il rilevamento automatico dei pasti, utilizzando dati provenienti da monitoraggio continuo del glucosio e somministrazioni di insulina raccolti su pazienti reali.

Capitolo 3

Dataset e preprocessing

3.1 Descrizione del dataset

Il presente studio utilizza un sottoinsieme della coorte di studio T1DEXI [21], composto da 78 pazienti di nazionalità americana, con un'età compresa tra i 12 e i 17 anni. La distribuzione di genere all'interno del campione comprende 34 femmine e 44 maschi.

Il dataset comprende una serie di variabili demografiche e cliniche, tra cui età, altezza, peso, data di nascita, sesso, paese di residenza e nazionalità. Inoltre, sono inclusi dati relativi alla gestione del diabete, come il livello di emoglobina glicata (HbA1c), la modalità di somministrazione dell'insulina (Pompa Insulinica in closed-loop o MDI), il dispositivo di somministrazione dell'insulina e i tipi di insulina utilizzati (basale e rapida).

Ulteriori informazioni riguardano le abitudini alimentari, con dati sul numero medio di pasti principali e spuntini consumati al giorno.

Per ogni paziente, il dataset include tracce relative a uno o più giorni, che forniscono dettagli temporali e metabolici. Questi dati includono:

- il tempo (data e ora di ogni campionamento, con un tempo di campionamento ogni 5 minuti)
- i livelli di glucosio nel sangue [mg/dL]
- la dose basale di insulina somministrata [U/min]
- le dosi dei boli di insulina [U/min]
- informazioni relative ai pasti [grammi]

Come riassunto in Figura 6 in totale abbiamo dati relativi a 78 soggetti affetti da diabete di tipo 1, per un totale di 134 tracce relative a diverse giornate. Dei 78 soggetti, 54 utilizzano la pompa insulinica (con un totale di 88 tracce) mentre 24 seguono il regime di MDI, rappresentato da 46 tracce in totale.

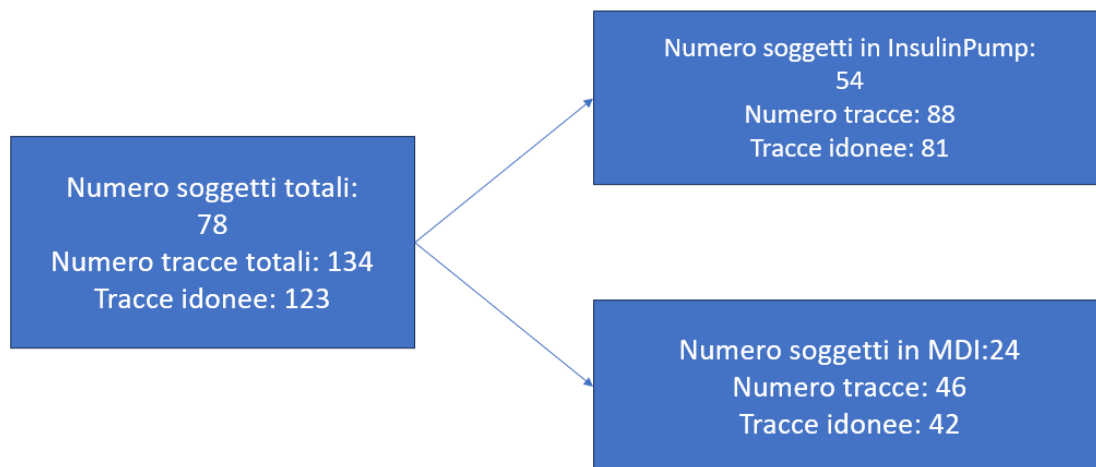


Figura 6 Schema riassuntivo dataset.

Tuttavia, per garantire l'affidabilità dei dati, sono state escluse dallo studio le tracce che presentavano una percentuale di valori mancanti (NaN) nel segnale CGM maggiore o uguale al 15%. Inoltre, sono state escluse le tracce che comprendevano un unico pasto reale, al fine di evitare una rappresentazione distorta dei pattern alimentari e metabolici.

Di conseguenza, sono state escluse un totale di 11 tracce dallo studio. Il numero di tracce totali ammesse allo studio è quindi pari a 123, di cui 81 relative a soggetti che utilizzano la pompa insulinica e 42 relative a soggetti che seguono il regime di MDI.

3.2 Pre-processing dei dati

Prima di procedere con l'analisi, come accennato sopra, è stato condotto un controllo sulla percentuale di dati mancanti nel segnale CGM e sul numero di pasti effettivamente registrati dai soggetti partecipanti allo studio. Sono state escluse dall'analisi le tracce che presentavano più del 15% di valori mancanti nel segnale CGM e quelle che riportavano meno di un pasto registrato. Successivamente, i pasti registrati entro un intervallo temporale di un'ora sono stati sommati, e il valore risultante è stato attribuito a un punto temporale intermedio tra i pasti adiacenti.

È stato osservato che alcuni soggetti presentavano un pattern di somministrazione di insulina caratterizzato da numerosi piccoli boli consecutivi anziché un singolo bolo di ampiezza maggiore,

noti come ‘boli ad onda quadra’. Per uniformità nell’analisi, i boli ad onda quadra sono stati sommati e collocati temporalmente al momento del primo bolo effettuato.

In aggiunta, qualora il numero totale di boli somministrati durante il corso di una giornata superi una soglia predeterminata (10 boli o più), viene calcolata la mediana dell’ampiezza di tali boli. I boli con ampiezza inferiore al valore mediano non vengono quindi considerati nelle analisi successive. Questa procedura è stata introdotta al fine di eliminare boli anomali o eccezionalmente bassi che potrebbero essere associati ad azioni correttive più che a boli prandiali, e quindi non funzionali all’identificazione dell’assunzione di carboidrati.

Successivamente, il segnale CGM è stato filtrato con un filtro a media mobile di ordine 5 per ridurre il rumore nei dati CGM e fornire una rappresentazione più accurata delle variazioni del glucosio nel tempo. Il codice Matlab utilizzato è il seguente:

```
% Moving Average filtering

windowSize = 5;
b = (1/windowSize)*ones(1,windowSize);
a = 1;
fCGM=filtfilt(b,a,CGM(~isnan(CGM)));
N=length(fCGM);
```

Il filtro a media mobile utilizzato può essere definito nel seguente modo:

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_1 + b_2 z^{-1} + \dots + b_{Nb} z^{-Nb}}{1}$$

Dove $b_i = \frac{1}{N_w}$ per $i=1,2, \dots, N_w$ e N_w è la dimensione della finestra del filtro, fissata a 5. Nb

rappresenta il numero di coefficienti del numeratore del filtro, che corrisponde alla dimensione della finestra del filtro. Il filtro a media mobile è applicato ai dati CGM, $CGM(t)$, attraverso la convoluzione con i coefficienti b_i :

$$CGM_f(t) = \sum_{i=1}^N b_i \cdot CGM(t - i + 1)$$

Dove $CGM_f(t)$ rappresenta i dati CGM filtrati ed N è la lunghezza dei dati originali. Questa convoluzione viene eseguita utilizzando la funzione ‘**filtfilt()**’ di MATLAB, che applica il filtro sia in avanti che all’indietro, per eliminare qualsiasi ritardo temporale e garantire un filtraggio a fase nulla.

Questa parte di lavoro svolge un ruolo fondamentale nell'elaborazione dei dati CGM, fornendo una rappresentazione più pulita e accurata delle variazioni del glucosio nel tempo, che è essenziale per le analisi successive e per il funzionamento dell'algoritmo.

In ultima analisi, il dataset viene suddiviso in un training set ed un test set mediante l'impiego della tecnica della Hold-Out Cross Validation. Questo metodo mira a mitigare l'overfitting e a consentire l'ottimizzazione dei parametri in base alle caratteristiche specifiche del dataset in esame.

Capitolo 4

Algoritmo di identificazione dei pasti

4.1 Architettura dell'algoritmo

L'algoritmo di identificazione dei pasti formulato in questo lavoro di tesi si basa sull'analisi dei boli di insulina, sfruttando le informazioni derivanti da essi. L'ipotesi alla base è che la presenza di boli di insulina durante il corso di una giornata possa indicare due situazioni principali: il loro correlarsi con l'assunzione di pasti o la necessità di correggere livelli elevati di glicemia. Pertanto, la struttura generale dell'algoritmo è impostata sull'analisi delle porzioni di tempo in cui sono presenti i boli e sull'analisi delle porzioni di tempo in cui non sono presenti boli. L'algoritmo, schematizzato in Figura 7, si suddivide in quattro blocchi:

- 1) Attivazione della meal flag in corrispondenza dei boli di insulina
- 2) Distinzione tra i boli prandiali e boli correttivi
- 3) Analisi degli eventi ipoglicemici
- 4) Analisi dei picchi del segnale CGM e della sua derivata prima

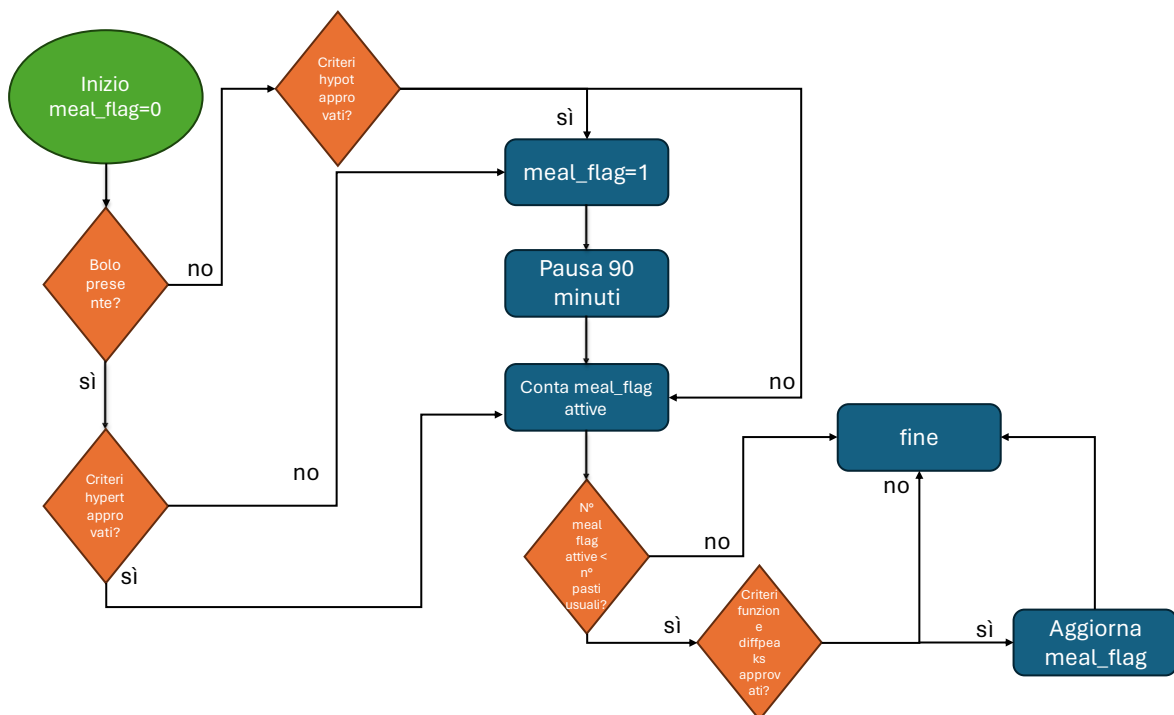


Figura 7 Schematizzazione dell'algoritmo di rilevamento dei pasti formulato.

4.1.1 Attivazione della meal flag in corrispondenza dei boli di insulina

In questa porzione di algoritmo viene inizializzato un vettore di flag dei pasti, dove ogni elemento rappresenta un punto nel tempo e viene inizializzato a zero. Successivamente, per ciascun bolo di insulina presente nei dati, viene individuata una finestra temporale di 60 minuti intorno ad esso. All'interno di questa finestra, viene individuato il bolo con il valore più alto e il relativo indice viene utilizzato per attivare il flag del pasto corrispondente. Dopo questa fase, viene eseguito un processo di “pausa” per eliminare eventuali sovrapposizioni tra pasti consecutivi dove, per 90 minuti dopo ogni pasto rilevato, i flag dei pasti successivi vengono disattivati. In Figura 8 viene mostrato il risultato di questo primo blocco.

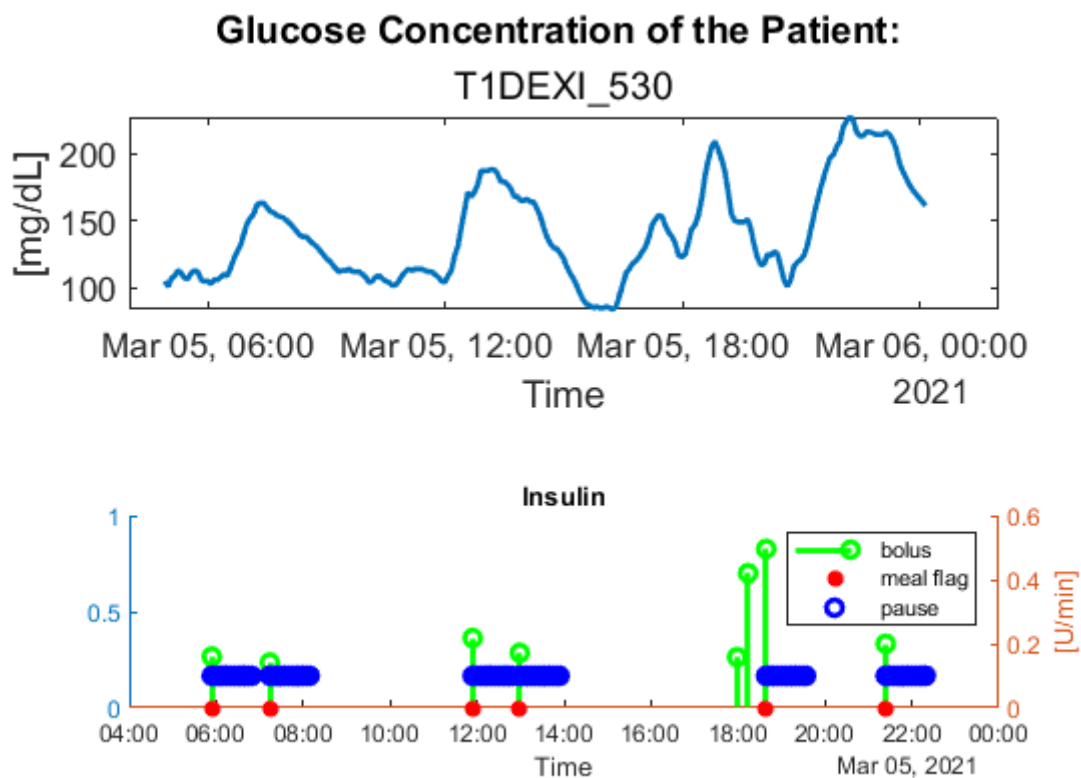


Figura 8 Esempio funzione del primo blocco basato sui boli di insulina. La meal flag (punto rosso) viene attivata in corrispondenza dei boli con valore maggiore tra boli vicini. La linea blu rappresenta la pausa di 90 minuti, attivata ogni volta che viene rilevato un pasto.

4.1.2 Distinzione tra i boli prandiali e boli correttivi

Questo blocco viene attivato solo se il numero di dosi di insulina somministrate durante il corso di una giornata supera il numero normale di pasti consumati dal paziente. Per iniziare, vengono

calcolati i percentili delle dosi di insulina, fornendo una panoramica delle quantità di insulina somministrate durante la giornata.

Considerando che i percentili più elevati delle dosi di insulina sono probabilmente associati ai pasti principali, mentre quelli più bassi potrebbero essere associati a correzioni della glicemia, viene valutata sia la quantità di insulina somministrata che il livello di glucosio nel sangue associato a quella dose. Se la glicemia è uguale o superiore a 250 mg/dL al momento dell'assunzione di una dose di insulina che si colloca al 25° percentile o inferiormente rispetto alle dimensioni delle dosi, allora quella dose viene considerata come bolo correttivo e la meal flag viene disattivata. La Figura 9 e la Figura 10 mostrano come le meal flag che erano state attivate dalla sezione precedente vengano disattivate se i boli corrispondenti vengono identificati come boli correttivi.

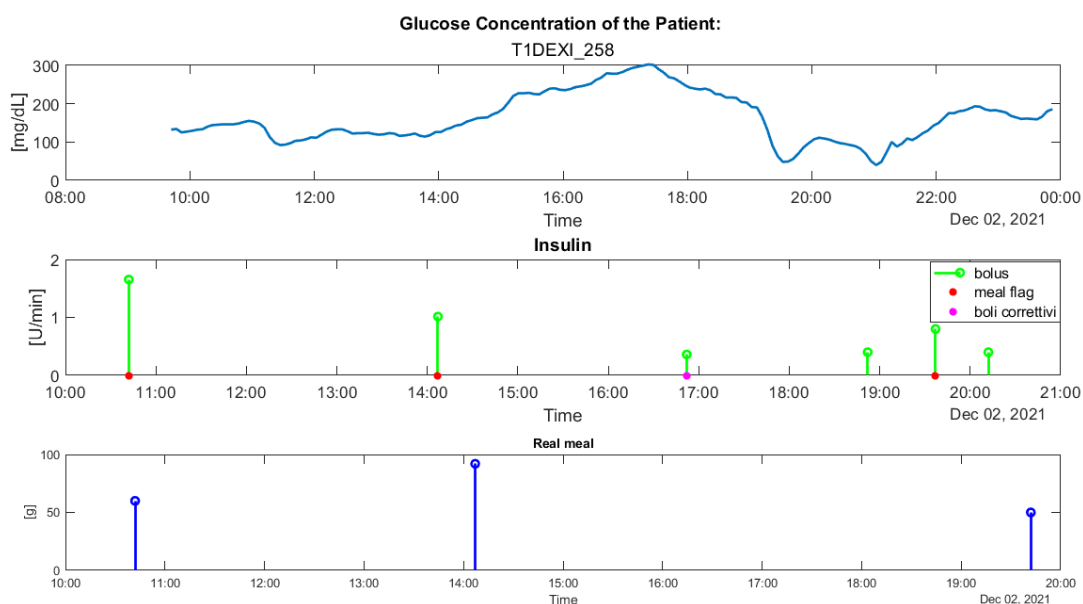


Figura 9 Esempio di funzionamento del secondo blocco dell'algorithm. Il punto viola rappresenta una meal flag che era stata attivata nel primo blocco, successivamente disattivata in quanto rispecchiava le condizioni di un bolo correttivo.

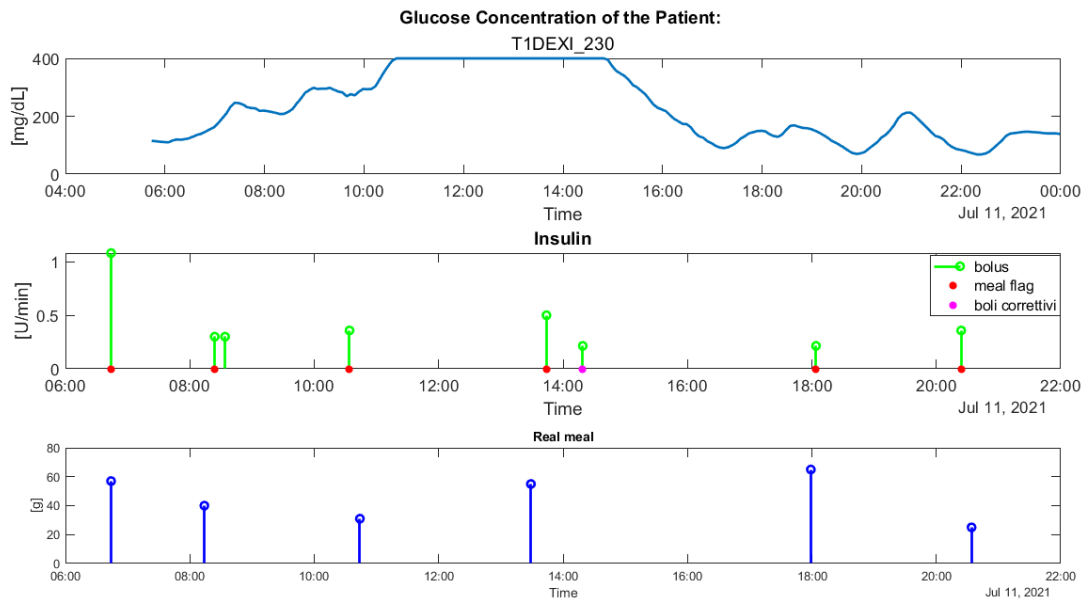


Figura 10 Un altro esempio di disattivazione della meal flag in corrispondenza di un bolo identificato come bolo correttivo.

4.1.3 Analisi degli eventi ipoglicemici

L'obiettivo dell'algoritmo non si limita all'identificazione dei pasti principali, ma si estende anche al rilevamento di eventuali spuntini consumati dal paziente senza l'accompagnamento di una dose di insulina. Questo scenario può verificarsi durante episodi di ipoglicemia, quando il consumo limitato di carboidrati potrebbe non richiedere la somministrazione di insulina ma avviene per contrastare l'evento ipoglicemico, per cui risulta impossibile identificare l'assunzione di carboidrati basandosi unicamente sui boli di insulina somministrati. Di conseguenza, sono state sviluppate regole decisionali per individuare tali situazioni e consentire una valutazione più completa dei comportamenti alimentari del paziente.

L'analisi degli eventi ipoglicemici come pasti avviene nelle porzioni temporali in cui non è presente una meal flag attiva nell'arco di 90 minuti prima e dopo l'evento ipoglicemico. Gli eventi ipoglicemici vengono individuati evidenziando i minimi del segnale CGM inferiori a 75 mg/dL: una meal flag viene attivata se la derivata prima del segnale CGM successiva a tali minimi è positiva per i successivi 60 minuti. Nell'esempio in Figura 11, si può notare che la meal flag si attiva in

corrispondenza di un evento ipoglicemico successivo ad un pasto, probabilmente causato da un'eccessiva dose di insulina effettuata durante il pasto.

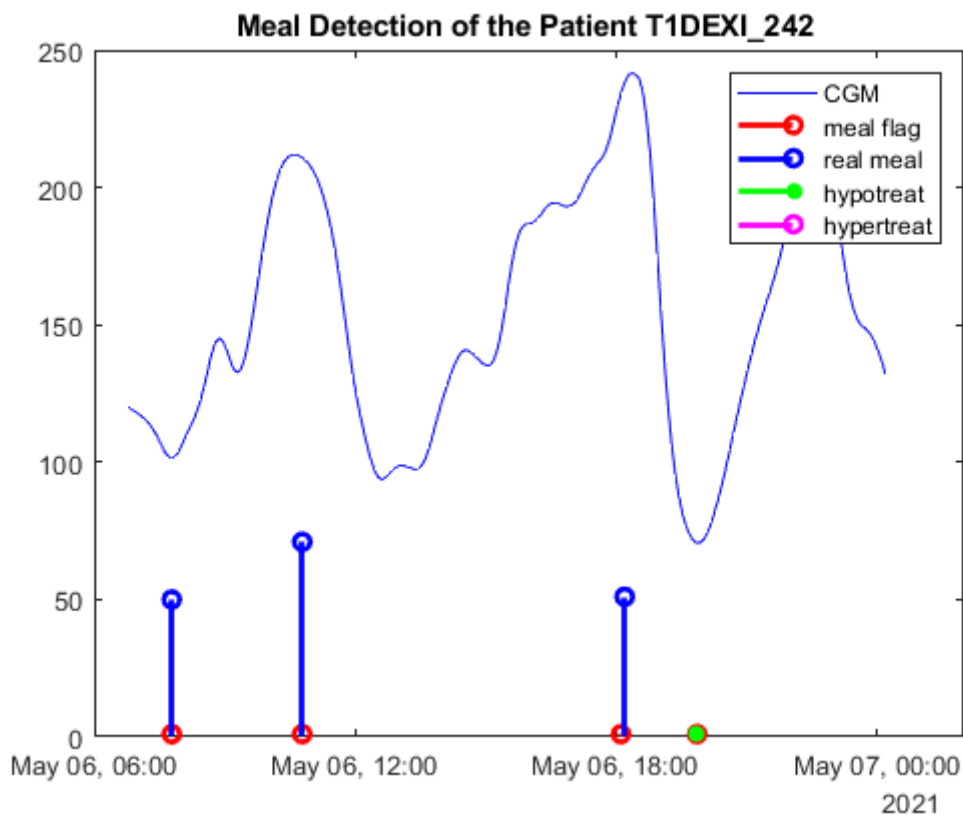


Figura 11 In verde la meal flag attivata dal blocco che analizza gli eventi ipoglicemici.

4.1.4 Analisi dei picchi del segnale CGM e della sua derivata prima

Nell'ultimo blocco di analisi, si è tenuto conto della natura dei dati reali utilizzati, che appartengono a pazienti in età adolescenziale. Si è considerato che tali pazienti tendono a deviare dal regime insulinico standard, talvolta omettendo del tutto l'assunzione di insulina o somministrandola in orari non corrispondenti ai pasti [22]. Pertanto, è stato introdotto un ulteriore blocco che si concentra sull'osservazione dei picchi nel segnale CGM e della sua derivata, al fine di individuare i pasti che potrebbero non essere stati seguiti dalla somministrazione di insulina. Questo approccio mira a fornire una maggiore sensibilità nell'identificazione dei momenti in cui il paziente ha consumato cibo senza assumere insulina. Questo blocco, tuttavia, viene attivato solo se il numero di pasti rilevati dai blocchi precedenti è inferiore al numero di pasti usuali effettuati dal paziente, per limitare il numero di falsi positivi. Come si può osservare dalla Figura 12, i punti in nero sono i pasti rilevati dall'analisi del segnale CGM, mentre i punti in rosso sono quelli rilevati dalle regole

decisionali dei blocchi precedenti. Questi ultimi hanno la precedenza rispetto ai primi, pertanto le meal flag che si otterranno in output sono i punti neri e i punti rossi cerchiati di verde.

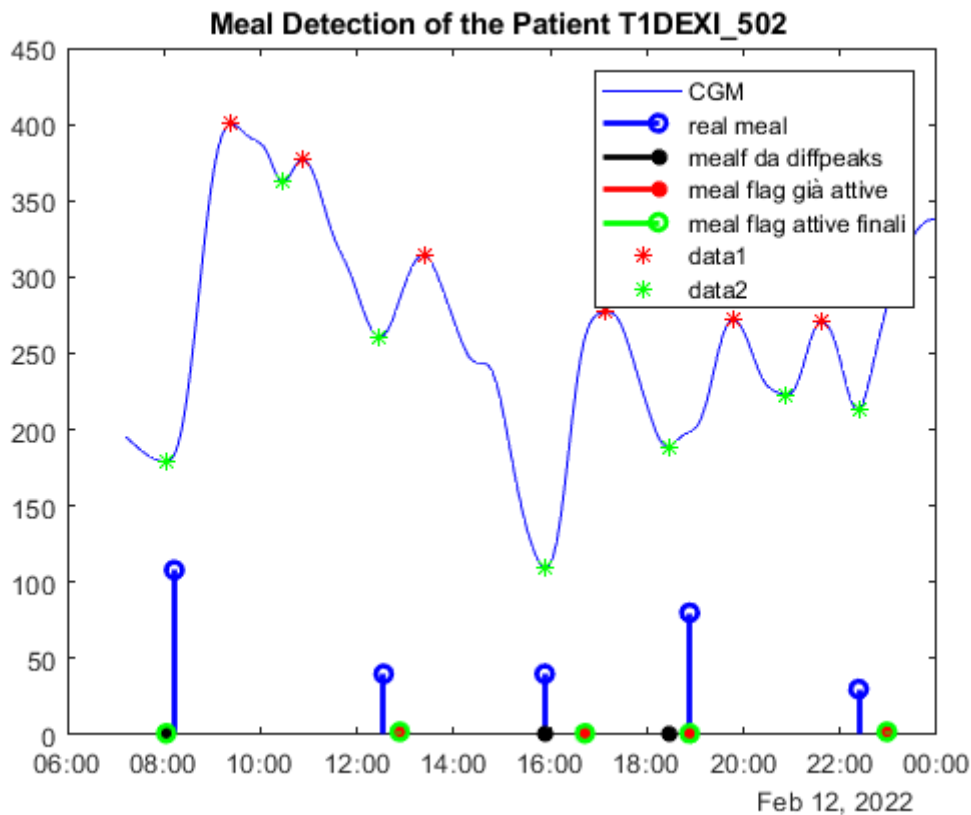


Figura 12 Esempio di meal detection attraverso la sezione dell'analisi del segnale CGM.

4.2 Dettagli implementativi dell'algoritmo

I parametri della prima sezione dell'algoritmo di rilevamento sono:

- τ , tempo in minuti corrispondente alla finestra temporale che si vuole analizzare, qui impostato a 60 minuti.
- $w = \frac{\tau}{T_s}$, numero di campioni in analisi, con T_s il tempo di campionamento, in questo caso di 5 minuti.
- *pause*, qui tarato a 90 minuti.

I parametri considerati per la discriminazione tra boli prandiali e boli correttivi sono:

- $prc25_{bolus}$, corrispondente al 25° percentile del valore dei boli. Viene calcolato per ogni nuova traccia in ingresso.

- θ_{hyper} , valore di glicemia [mg/dL] oltre il quale il bolo è considerato un bolo correttivo, qui imposto a 250 mg/dL.

Per quanto riguarda il blocco relativo all'analisi degli eventi ipoglicemici, i parametri fissati sono i seguenti:

- $\tau_{dG>0}$ corrispondente al tempo desiderato di positività della derivata, in questo lavoro impostato a 60 minuti.
- θ_{hypo} , valore di glicemia [mg/dL] sotto il quale si identificano gli eventi ipoglicemici, qui imposto a 75 mg/dL.

Per ottenere un'approssimazione della derivata prima, si è ricorso al metodo delle differenze finite applicando la seguente formula:

$$dCGM_f(i) = CGM_f(i + 1) - CGM_f(i)$$

per $i=1, \dots, N-1$, con N corrispondente al numero di campioni del segnale CGM filtrato (CGM_f);

L'attivazione di una meal flag data dall'analisi del segnale CGM avviene tramite il controllo di due condizioni: la differenza Δ_G tra punti di minimo e punti di massimo adiacenti nel segnale CGM e il controllo della positività della derivata prima¹ successiva ai punti di minimo per un tempo predefinito. Il tutto avviene attraverso una funzione definita “*diffpeaks*” che prende come input:

- il segnale CGM filtrato, G
- θ_{diff} , ovvero la soglia di Δ_G da superare per la detection.
- $\tau_{dG>0}$ corrispondente al tempo desiderato di positività della derivata, qui impostato a 60 minuti.
- $\theta_{perc_dG>0}$ corrisponde alla soglia di percentuale di positività della derivata da superare per la detection.

Forniti i dati in input, la funzione calcola:

$$[p_{max}, l_{max}] = findpeaks(G);$$

$$[p_{min}, l_{min}] = findpeaks(-G);$$

Con $k = 1, \dots, \min(n_{min}, n_{max})$, con n_{min} e n_{max} il numero di punti di minimo ($l(k)_{min}$) e di massimo ($l(k)_{max}$) di G , rispettivamente, si prosegue col calcolo:

$$\Delta G = G(l_{max(k)}) - G(l_{min(k)});$$

$$dGi = G_{i+1} - Gi;$$

¹ Calcolata come differenza prima come riportato precedentemente.

per $i=1, \dots, N-1$, con N corrispondente al numero di campioni del segnale CGM;

$range = [l(k)_{min} : l(k)_{min} + \tau_{dG>0}]$

If $\Delta G \geq \theta_{diff}$ && $perc_dG(range) \geq \theta_{perc_dG>0}$

$meal_flag(l(k)_{min}) = 1;$

end

La funzione restituisce in output il vettore *meal_flag*.

È stato aggiunto anche un controllo secondo cui, se il primo punto rilevato del segnale CGM è un punto di massimo, allora si attiva una meal flag a metà tra il primo dato campionato $G(1)$ e il primo punto di massimo e $l(1)_{max}$.

I risultati ottenuti dall'analisi tramite *diffpeaks* vengono confrontati e integrati con le meal flag precedentemente attivate. Vengono quindi considerate solo le sezioni temporali prive di qualsiasi meal flag già attiva.

Capitolo 5

Risultati e discussione

5.1 Metriche di valutazione

Nel contesto di questo lavoro, i veri positivi (*True Positive*, TP), i falsi negativi (*False Negative*, FN), i falsi positivi (*False Positive*, FP) e i veri negativi (*True Negative*, TN) vengono definiti nel seguente modo:

- Sono considerati TP i pasti rilevati correttamente all'interno dell'arco temporale di un'ora dalla loro assunzione.
- Sono considerati FN i pasti non rilevati all'interno dell'arco temporale di un'ora dalla loro assunzione.
- Sono considerati FP i pasti rilevati in modo errato.
- I TN non vengono considerati, in quanto non si hanno gli strumenti per verificare che un pasto venga correttamente non rilevato.

Per andare a valutare le performance dell'algorithmo sono state utilizzate le seguenti metriche:

1) La sensibilità, che indica la capacità di un modello di classificazione di identificare correttamente tutti i casi positivi. È calcolata come il rapporto tra i veri positivi e la somma dei veri positivi e dei falsi negativi. Un alto valore di sensibilità indica una buona capacità del modello di identificare tutti i casi positivi.

$$sensibilità = \frac{TP}{TP + FN} * 100$$

2) la precisione, che indica la capacità del modello nel classificare correttamente i casi positivi tra tutti i casi classificati come positivi. È calcolata come il rapporto tra i veri positivi e la somma dei veri positivi e dei falsi positivi.

$$precisione = \frac{TP}{TP + FP} * 100$$

3) L’F1-score, che viene calcolato come la media armonica di precisione e sensibilità e fornisce un’indicazione complessiva della performance del modello ed è calcolato come segue:

$$F1score = \frac{2 * precisione * sensibilità}{precisione + sensibilità}$$

Per presentare i risultati finali, verrà calcolato il valor medio e il valore mediano delle tre metriche menzionate (sensibilità, precisione, F1-score) calcolate su ogni singola traccia. Verrà inoltre calcolata la performance di un “super-soggetto”, in cui si ipotizza che le tracce analizzate appartengano tutte ad un unico soggetto.

5.2 Analisi dei risultati ottenuti

Si sono analizzate le performance dell’algoritmo dividendo il dataset secondo tre condizioni:

- 1) con solo dati relativi ai soggetti in MDI (24 soggetti, per un totale di 42 tracce)
- 2) con solo dati relativi ai soggetti in pompa insulinica (54 soggetti, per un totale di 81 tracce)
- 3) unendo il dataset con i soggetti in MDI ed in pompa insulinica (78 soggetti, per un totale di 123 tracce)

Per ognuna delle tre situazioni verranno confrontate le performance dell’algoritmo, aggiungendo sequenzialmente ogni sezione di analisi, dove:

- Sezione 1: attivazione meal flag tramite la logica dei boli
- Sezione 2: distinzione tra boli prandiali e boli correttivi
- Sezione 3: analisi degli eventi ipoglicemici
- Sezione 4: analisi del segnale CGM

5.2.1 Caso solo MDI

I risultati ottenuti utilizzando solo l'attivazione della logica dei boli sono riportati in Tabella 1. Con questa condizione, si ottiene un'elevata precisione (91.59% considerando il valore medio della precisione ottenuto nel dataset) e una sensibilità promettente (pari a 79.56% considerando in modo analogo il caso medio). La differenza tra valori medi e mediani di quasi il 10% è causata da una distribuzione asimmetrica delle performance, come riportato in Figura 13.

Tabella 1 Performance dell'algoritmo dopo l'attivazione della Sezione 1 – caso soggetti in MDI

Sez 1	Sensibilità	F1-score	Precisione
Media	79.56%	83.72%	91.59%
Mediana	90%	88.89%	100%
Supersoggetto	76.5%	83.28%	91.37%

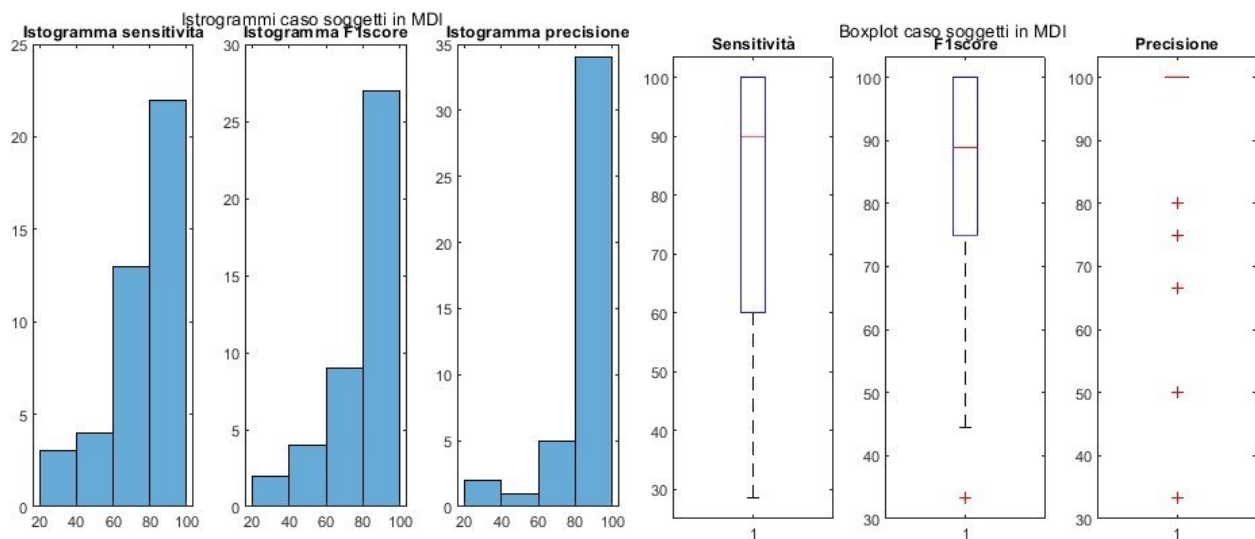


Figura 13 Rispettivamente istogramma e boxplot delle performance dell'algoritmo dopo l'attivazione della Sezione 1 per i soggetti in MDI.

Come riportato in Tabella 2, si può notare come, per i soggetti in MDI, l'aggiunta della logica per la distinzione dei boli correttivi non modifica le performance dell'algoritmo.

Tabella 2 Performance dell' algoritmo dopo l' attivazione della Sezione 1 e 2 – caso soggetti in MDI

Sez 1 + Sez 2	Sensitività	F1-score	Precisione
Media	79.56%	83.72%	91.59%
Mediana	90%	88.89%	100%
Supersoggetto	76.5%	83.28%	91.37%

La Tabella 3 riporta i risultati ottenuti con l' aggiunta della sezione degli eventi ipoglicemici: si può notare un lieve aumento delle performance del super-soggetto, mentre le altre sono rimaste pressoché invariate.

Tabella 3 Performance dell' algoritmo dopo l' attivazione della Sezione 1, 2 e 3 – caso soggetti in MDI

Sez 1 + Sez 2 + Sez 3	Sensitività	F1-score	Precisione
Media	80.04%	83.72%	91%
Mediana	90%	88.89%	100%
Supersoggetto	77.11%	83.39%	90.78%

Per l' analisi dell' aggiunta della sezione 4, è stato fatto un tuning dei parametri θ_{diff} e $\theta_{perc_dG>0}$, in modo da massimizzare il valore dell' F1-score medio. Il tuning è stato effettuato su un training set, composto dall' 80% del dataset iniziale (34 tracce), e la validazione è stata effettuata su un test set composto dal restante 20% dei dati (8 tracce).

La procedura di tuning dei parametri è stata effettuata in tre step:

1. Si sono definite due griglie di valori per testare rispettivamente per θ_{diff} e $\theta_{perc_dG>0}$. In particolare valori di $\theta_{diff} \in [10: 10: 350]$ mg/dL e di $\theta_{perc_dG>0} \in [0: 10: 100]\%$.
2. Per ogni valore di θ_{diff} e $\theta_{perc_dG>0}$ è stato calcolato $F1 - score(\theta_{diff}; \theta_{perc_dG>0})$
3. Si è scelto come θ_{diff} e $\theta_{perc_dG>0}$ i valori che massimizzano $F1 - score$.

La heatmap in Figura 14 mostra come varia l' F1-score con diversi valori dei parametri. Si può notare come, per piccoli valori della soglia θ_{diff} (10-60 mg/dL), la variazione della percentuale della positività della derivata va ad influenzare maggiormente il valore dell' F1-score medio; per valori di θ_{diff} da 70 a 100 mg/dL, la derivata influisce sempre meno, fino a non influire per niente nei risultati per valori di θ_{diff} sopra i 110 mg/dL.

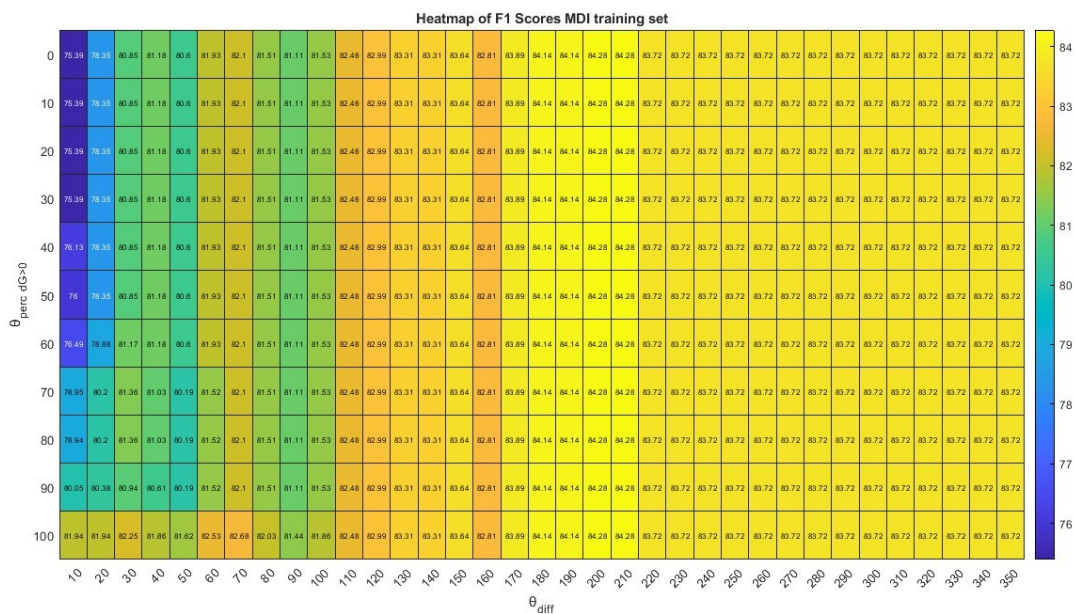


Figura 14 Heatmap dell'F1-score per il caso dei soli soggetti in MDI

Per una massimizzazione dell'F1-score per i soggetti in MDI, si è stabilito l'utilizzo di una soglia $\theta_{diff} = 200$ mg/dL, senza rendere necessario il controllo sulla derivata. Di seguito vengono riportati i risultati sul dataset completo e sul test set:

Tabella 4 Performance dell'algorithmo sul test set dopo il tuning dei parametri – caso soggetti in MDI

Test set con $\theta_{diff} = 200$ mg/dL	Sensitività	F1-score	Precisione
Media	84.58%	88.03%	93.75%
Mediana	80%	87.30%	100%
Supersoggetto	83.33%	87.72%	92.60%

Tabella 5 Performance dell'algorithmo sul dataset completo dopo il tuning dei parametri – caso soggetti in MDI

Dataset completo con $\theta_{diff} = 200$ mg/dL	Sensitività	F1-score	Precisione
Media	81.9%	85%	91,39%
Mediana	100%	88.89%	100%
Supersoggetto	78.91%	84.5%	90.97%

Si può notare come l'integrazione della sezione 4 migliora le performance generali dell'algoritmo per i soggetti in MDI.

5.2.2 Caso solo soggetti in pompa insulinica

Analogamente al caso dei pazienti in terapia MDI, i risultati ottenuti per ogni step dell'algoritmo sono stati riportati nelle seguenti tabelle.

Le performance ottenute applicando solo il primo modulo sono riportate in Tabella 6.

Tabella 6 Performance dell'algoritmo dopo l'attivazione della Sezione 1 – caso soggetti in pompa

Sez 1	Sensitività	F1-score	Precisione
Media	86.25%	81.20%	79.90%
Mediana	100%	85.71%	80%
Supersoggetto	85.03%	81.73%	78.67%

Rispetto al caso dei soggetti in MDI, i soggetti in pompa presentano una sensitività maggiore (pari a 86.25% nel caso medio) ma una precisione inferiore. Anche qui si può notare l'elevata differenza tra il valore medio della sensitività e il valore mediano, indicando un'asimmetria nelle distribuzioni delle performance, com'è possibile osservare dalla Figura 15.

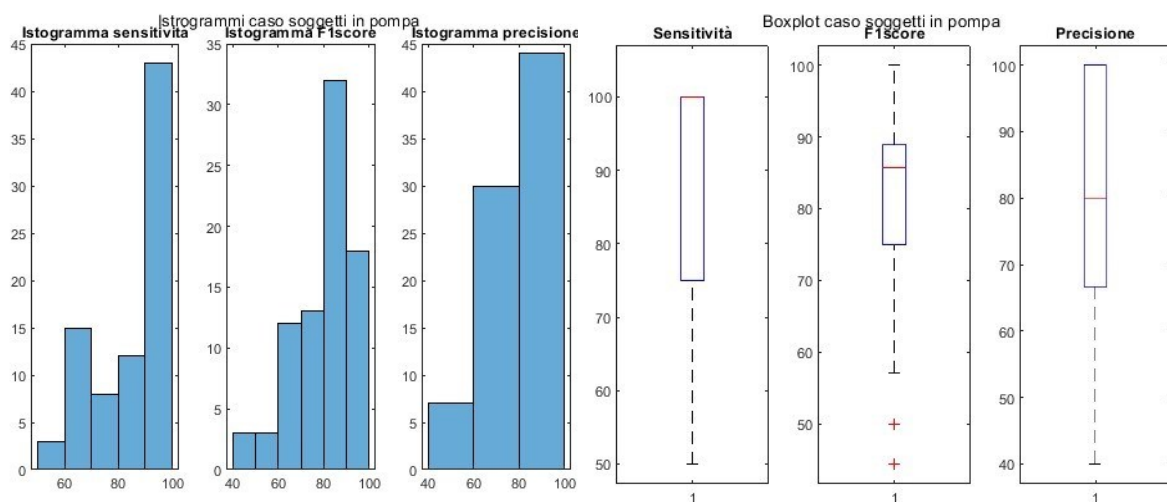


Figura 15 Rispettivamente istogramma e boxplot delle performance dell'algoritmo dopo l'attivazione della Sezione 1 per i soggetti in pompa.

Tabella 7 Performance dell' algoritmo dopo l'attivazione della Sezione 1 e 2 – caso soggetti in pompa

Sez 1 + Sez 2	Sensitività	F1-score	Precisione
Media	86.25%	82.49%	82.26%
Mediana	100%	85.71%	80%
Supersoggetto	85.03%	82.8%	80.68%

Successivamente si è applicato anche il secondo modulo e i risultati sono stati riportati in Tabella 7. Si può notare come l'aggiunta della sezione sui boli correttivi aumenti il livello di precisione dell'algoritmo, andando a diminuire il numero di falsi positivi trovati dalla Sezione 1. Resta invece invariata la sensitività.

Tabella 8 Performance dell' algoritmo dopo l'attivazione della Sezione 1, 2 e 3 – caso soggetti in pompa

Sez 1 + Sez 2 + Sez 3	Sensitività	F1-score	Precisione
Media	86.25%	81.63%	80.75%
Mediana	100%	85.71%	80%
Supersoggetto	85.03%	81.96%	79.11%

L'aggiunta del blocco con l'analisi degli eventi ipoglicemici non migliora le performance dell'algoritmo, come riportato in Tabella 8. Tuttavia, questa analisi può essere utile nell'ottica di rilevare eventuali pasti non annunciati, anche se non va a migliorare numericamente le performance generali.

Anche in questo caso è stata fatta l'analisi dell'aggiunta della Sezione 4, facendo il tuning dei parametri θ_{diff} e $\theta_{perc_dG>0}$, in modo da massimizzare il valore dell'F1-score medio, utilizzando training set (65 tracce) e test set (16 tracce).

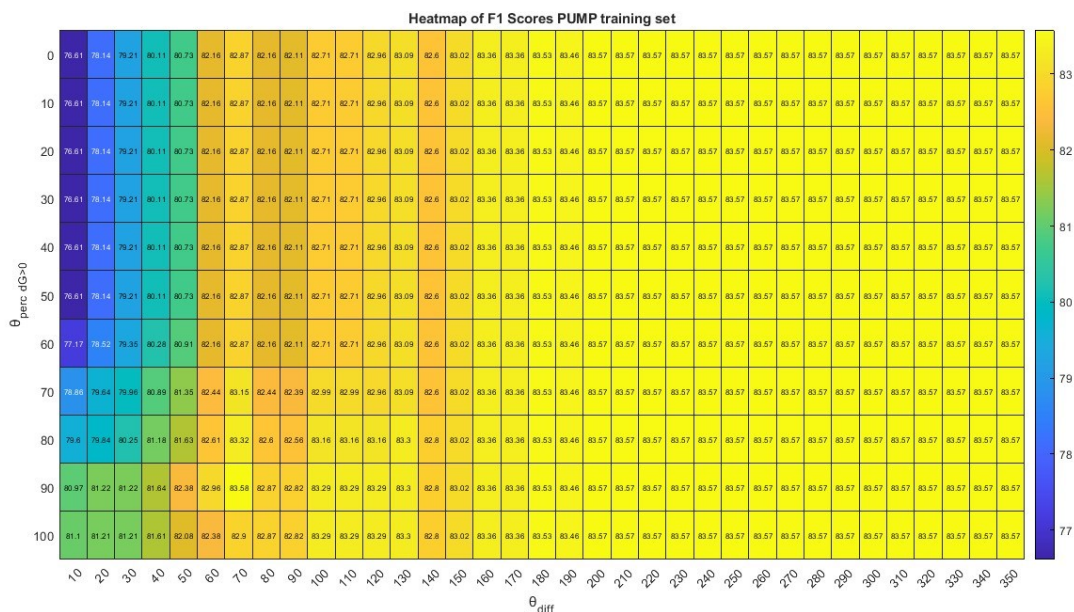


Figura 16 Heatmap dell'F1-score per il caso dei soggetti in pompa.

Come nel caso dei soggetti in MDI, la heatmap in Figura 16 mostra come, per piccoli valori della soglia θ_{diff} la variazione della percentuale della positività della derivata va ad influenzare maggiormente il valore dell'F1-score medio, andando ad influire sempre meno per valori di θ_{diff} sempre maggiori. In questo caso però otteniamo un F1-score massimizzato (83.58%) per valori di θ_{diff} pari a 70 mg/dl e di $\theta_{perc_dG>0}$ pari al 90%. Tuttavia, si ottengono valori di F1-score molto simili (83.57%) utilizzando valori di $\theta_{diff} \geq 200$ mg/dl, indipendentemente da $\theta_{perc_dG>0}$. Le performance sul test set e sul dataset completo terminato il tuning sono riportati in Tabella 9 e in Tabella 10:

Tabella 9 Performance dell'algorithmo sul test set dopo il tuning dei parametri – caso soggetti in pompa

Test set con $\theta_{diff} =$ 70 mg/dl e $\theta_{perc_dG>0} = 90\%$	Sensitività	F1-score	Precisione
Media	92.6%	79.07%	71.46%
Mediana	100%	85.71%	75%
Supersoggetto	91.52%	79.41%	70.13%

Tabella 10 Performance dell’algoritmo sul dataset completo dopo il tuning dei parametri – caso soggetti in pompa

Dataset completo con $\theta_{diff} = 70 \text{ mg/dl e}$ $\theta_{perc_dG>0} = 90\%$	Sensitività	F1-score	Precisione
Media	89.75%	82.20%	78.45%
Mediana	100%	85.71%	75%
Supersoggetto	88.62%	82.56%	77.28%

Si può notare come l’aggiunta della Sezione 4 tenda ad aumentare la sensitività dell’algoritmo, mentre ne risente la precisione, a causa dell’aumento di falsi positivi.

5.2.3 Caso dataset completo (pompa + MDI)

Di seguito vengono riportati i risultati relativi al dataset in cui sono presenti sia le tracce relative ai soggetti in pompa insulinica sia quelle relative ai soggetti in MDI.

In particolare, nella Tabella 11 sono riportati i valori di sensitività, precisione e F1-score utilizzando solo la prima logica.

Tabella 11 Performance dell’algoritmo dopo l’attivazione della Sezione 1 – caso dataset completo (pompa + MDI)

Sez 1	Sensitività	F1-score	Precisione
Media	83.97%	82.06%	83.89%
Mediana	100%	85.71%	100%
Supersoggetto	82.2%	82.2%	82.2%

Successivamente è stata aggiunta l’attivazione della Sezione 2: come riportato in Tabella 12, questo nuovo set-up va ad aumentare il livello di precisione dell’algoritmo senza introdurre peggioramento in termini di sensitività.

Tabella 12 Performance dell’algoritmo dopo l’attivazione della Sezione 1 e 2 – caso dataset completo (pompa + MDI)

Sez 1 + Sez 2	Sensitività	F1-score	Precisione
Media	83.97%	82.91%	85.44%
Mediana	100%	85.71%	100%

Supersoggetto	82.2%	82.95%	83.71%
----------------------	-------	--------	--------

Come riportato in Tabella 13, l'integrazione della Sezione 3 va ad aumentare leggermente la sensitività, tuttavia l'F1-score e la precisione diminuiscono leggermente.

Tabella 13 Performance dell'algorithmo dopo l'attivazione della Sezione 1, 2 e 3 – caso dataset completo (pompa + MDI)

Sez 1 + Sez 2 + Sez 3	Sensitività	F1-score	Precisione
Media	84.13%	82.34%	84.25%
Mediana	100%	85.71%	100%
Supersoggetto	82.4%	82.4%	82.4%

Infine, analogamente ai casi precedentemente descritti, anche per il dataset completo, per l'attivazione della Sezione 4 è stato fatto un tuning dei parametri θ_{diff} e $\theta_{perc_dG>0}$, per massimizzare il valore dell'F1-score medio, utilizzando training set (99 tracce) e test set (24 tracce).

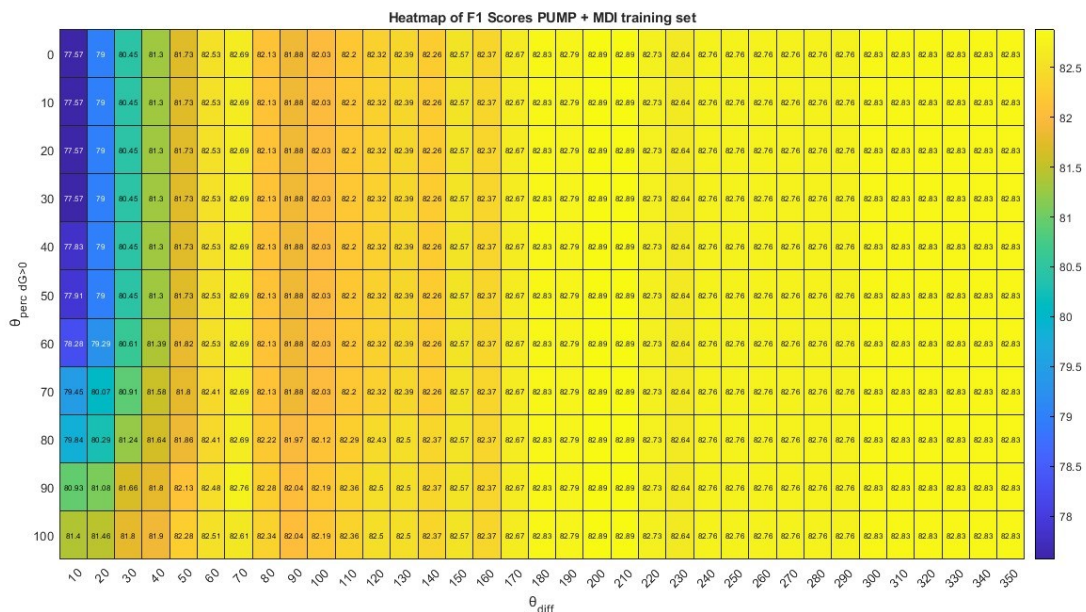


Figura 17 Heatmap dell'F1-score sul dataset completo (pump+mdi)

Come atteso, dalla Figura 17 emergono conclusioni analoghe a quelle ottenute per i casi precedenti. Si può notare un aumento dell'F1-score in corrispondenza di $\theta_{diff} = 70$ mg/dl e $\theta_{perc_dG>0} = 90\%$, mentre è massimo per valori di $\theta_{diff} \geq 200$ mg/dl ed indifferente ai valori di $\theta_{perc_dG>0}$.

Le performance sul test set e sul dataset completo con i parametri determinati dal tuning sono riportati rispettivamente nelle seguenti tabelle:

Tabella 14 Performance dell’algoritmo sul test set dopo il tuning dei parametri – caso dataset completo (pompa + MDI)

Test set con $\theta_{diff} =$ 200 mg/dL	Sensitività	F1-score	Precisione
Media	85.97%	84.18%	85.56%
Mediana	100%	88.89%	100%
Supersoggetto	83.67%	83.25%	82.83%

Tabella 15 Performance dell’algoritmo sul dataset completo dopo il tuning dei parametri – caso dataset completo (pompa + MDI)

Dataset completo con $\theta_{diff} =$ 200 mg/dL	Sensitività	F1-score	Precisione
Media	85.34%	82.88%	83.98%
Mediana	100%	85.71%	100%
Supersoggetto	83.6%	82.85%	82.12%

5.2.4 Considerazioni finali

Dall’analisi delle variazioni delle performance ottenute attivando sequenzialmente tutte e quattro le sezioni, si può notare come la Sezione 1 sia quella che ha più rilevanza sui risultati finali. La Sezione 2 aiuta a discriminare tra boli prandiali e boli correttivi, ma la sua efficacia è limitata, in quanto è difficile stabilire con esattezza delle regole che siano in grado di distinguere questi due casi, soprattutto trattandosi di dati reali di soggetti adolescenti.

La Sezione 3 in generale non contribuisce al miglioramento numerico delle performance, ma può essere utile per andare a rilevare eventuali pasti non annunciati o che non sono accompagnati da un bolo di insulina. Questi casi non sono stati considerati in questo lavoro in quanto richiederebbe l’identificazione di pasti non annunciati attraverso un labeling manuale possibile tramite ispezione visiva. È tuttavia documentato in letteratura che la presenza di pasti non annunciati e di boli prandiali non effettuati è frequente soprattutto nella popolazione adolescente, quindi, studi futuri verificheranno l’efficacia di queste soglie in presenza di questi eventi.

Per quanto riguarda la Sezione 4, il tuning dei parametri θ_{diff} e $\theta_{perc_dG>0}$ ha effettivamente portato ad un leggero aumento nelle performance dell' algoritmo, specialmente in termini di sensitività.

In seguito, si vuole riportare il confronto tra il numero effettivo di TP, FP e FN che si ottiene utilizzando i parametri definiti dal tuning e il numero che si ottiene forzando i due parametri a $\theta_{diff} = 70 \text{ mg/dl}$ e $\theta_{perc_dG>0} = 90\%$:

Tabella 16 Confronto performance con parametri forzati e parametri determinati dal tuning – caso MDI

MDI	$\theta_{diff} = 70 \text{ mg/dl}$ && $\theta_{perc_dG>0} = 90\%$	$\theta_{diff} = 200 \text{ mg/dl}$ && $\theta_{perc_dG>0} = 0$
TP	137	131
FP	26	13
FN	29	35

Tabella 17 Confronto performance con parametri forzati e parametri determinati dal tuning – caso pompa

Pompa	$\theta_{diff} = 70 \text{ mg/dl}$ && $\theta_{perc_dG>0} = 90\%$	$\theta_{diff} = 200 \text{ mg/dl}$ && $\theta_{perc_dG>0} = 0$
TP	296	287
FP	87	78
FN	38	47

Tabella 18 Confronto performance con parametri forzati e parametri determinati dal tuning – caso pompa + MDI

Pompa + MDI	$\theta_{diff} = 70 \text{ mg/dl}$ && $\theta_{perc_dG>0} = 90\%$	$\theta_{diff} = 200 \text{ mg/dl}$ && $\theta_{perc_dG>0} = 0$
TP	433	418
FP	113	91
FN	67	82

Com'è possibile notare, nonostante il valore di $\theta_{diff} = 200 \text{ mg/dl}$ massimizzi l'F1-score, il numero di TP trovati diminuisce (e di conseguenza aumenta il numero di FN) rispetto al caso con la forzatura dei parametri. Perciò, a seconda dell'applicazione dell'algoritmo, bisognerebbe andare a valutare se si preferisce ottenere una sensitività più alta (quindi aumentare il numero di TP a

discapito di qualche FP in più) oppure avere una precisione più alta (limitando il numero di FP a discapito di ottenere più FN).

Le performance sul dataset completo, nei tre casi, utilizzando i valori arbitrari $\theta_{diff} =$

70 mg e $\theta_{perc_dG>0} = 90\%$ sono riportate nelle tabelle seguenti:

Tabella 19 Performance dell' algoritmo completo per il caso dei soggetti in MDI

MDI	Sensitività	F1-score	Precisione
Media	81.9%	85%	91,39%
Mediana	100%	88.89%	100%
Supersoggetto	78.91%	84.5%	90.97%

Tabella 20 Performance dell' algoritmo completo per il caso dei soggetti in pompa

Pompa	Sensitività	F1-score	Precisione
Media	89.75%	82.20%	78.45%
Mediana	100%	85.71%	75%
Supersoggetto	88.62%	82.56%	77.28%

Tabella 21 Performance dell' algoritmo completo per il caso dei soggetti in MDI e dei soggetti in pompa

Pompa + MDI	Sensitività	F1-score	Precisione
Media	87.95%	82.76%	80.88%
Mediana	100%	85.71%	80%
Supersoggetto	86.6%	82.79%	79.30%

Conclusion

In questo percorso di tesi si è cercato di dare un contributo alla risoluzione di un problema relativo all'applicazione pratica di strumenti per il supporto alla decisione e l'individualizzazione della terapia e gestione del diabete di tipo 1, cioè sull'incertezza dei dati relativi all'assunzione di pasti. Infatti, questa incertezza ha dimostrato di complicare il processo di inserimento delle informazioni nei modelli di Digital Twin per la personalizzazione della terapia, come il tool ReplayBG, compromettendo l'efficacia nella valutazione delle terapie corrette. Attraverso lo sviluppo di un algoritmo basato su tecniche euristiche, questa tesi ha proposto una soluzione per identificare retrospettivamente i pasti in modo automatico a partire da dati relativi al segnale CGM e dai boli di insulina, migliorando così la precisione e l'affidabilità dei dati utilizzati da tecnologie come i Digital Twin.

L'algoritmo per l'identificazione di pasti sviluppato in questa tesi si è basato sull'analisi di dati reali, provenienti dal dataset T1DEXI, relativi a 123 giornate di 78 pazienti adolescenti affetti da diabete di tipo 1. L'analisi è stata effettuata prima separatamente per i pazienti in MDI e per i pazienti in pompa insulinica, e poi in modo congiunto. Per i pazienti in MDI l'algoritmo ha dimostrato un'elevata precisione (90.97%), ma una minore sensibilità (78.91%); per i pazienti in pompa insulinica si è registrata un'elevata sensibilità (88.62%), ma una precisione inferiore (77.28%); nell'analisi del caso congiunto si è ottenuta una sensibilità dell'83.6% ed una precisione dell'82.12%.

Tra le regole decisionali definite e le sezioni delineate, l'analisi dei dati derivanti dai boli di insulina ha dimostrato una particolare efficacia nell'individuare i pasti. L'integrazione dell'analisi del segnale CGM e della sua derivata ha ulteriormente migliorato la sensibilità complessiva dell'algoritmo. Sezioni come quella dedicata alla distinzione tra boli correttivi e boli prandiali, così come quella relativa all'analisi degli eventi ipoglicemici, hanno mostrato risultati promettenti, ma richiedono ulteriori approfondimenti per ottimizzare i valori delle soglie predefinite.

Le prossime fasi di sviluppo includono l'implementazione di una stima automatica del contenuto di carboidrati nei pasti e il miglioramento della precisione nel rilevamento temporale degli stessi.

Inoltre, per ulteriori sviluppi futuri, sarebbe opportuno espandere il dataset, aumentando il numero di partecipanti e potenzialmente includendo anche la popolazione pediatrica. Questo approccio potrebbe contribuire ad arricchire la metodologia e migliorare ulteriormente la sua efficacia e applicabilità, consentendo una comprensione più completa e approfondita del comportamento alimentare dei pazienti diabetici e delle loro esigenze specifiche, per una gestione sempre più semplice ed efficace della malattia.

Ringraziamenti

Desidero esprimere la mia sincera gratitudine al Prof. Andrea Facchinetti per la sua guida esperta e il suo costante sostegno durante il mio percorso di ricerca. Senza la sua saggezza e il suo incoraggiamento, questa tesi non sarebbe stata possibile.

Un ringraziamento speciale anche al Dott. Giacomo Cappon e alla Dott.ssa Elena Idi per il loro contributo prezioso e i loro preziosi consigli lungo il percorso. La loro competenza e il loro impegno hanno arricchito il mio lavoro in modo significativo.

Bibliografia

- [1] Katsarou, A., Gudbjörnsdóttir, S., Rawshani, A. et al. Type 1 diabetes mellitus. *Nat Rev Dis Primers* 3, 17016 (2017).
- [2] American Diabetes Association. "Diagnosis and classification of diabetes mellitus." *Diabetes care* 33. Supplement_1 (2010): S62-S69.
- [3] International Diabetes Federation. *IDF diabetes atlas*. IDF
- [4] Janež, Andrej, et al. "Insulin therapy in adults with type 1 diabetes mellitus: a narrative review." *Diabetes Therapy* 11 (2020): 387-409.
- [5] National Institute for Health and Care Excellence (NICE). *Type 1 diabetes in adults: diagnosis and management* London: National Institute for Health and Care Excellence (UK); 2015.
- [6] Misso ML, Egberts KJ, Page M, O'Connor D, Shaw J. Continuous subcutaneous insulin infusion (CSII) versus multiple insulin injections for type 1 diabetes mellitus. *Cochrane Database Syst Rev*. 2010;5(1): CD005103.
- [7] Yeh HC, Brown TT, Maruthur N, et al. Comparative effectiveness and safety of methods of insulin delivery and glucose monitoring for diabetes mellitus: a systematic review and meta-analysis. *Ann Intern Med*. 2012; 157:336–47.
- [8] American Diabetes Association. *Standards of medical care in diabetes—2019*. *Diabetes Care*. 2019;42: S1–193.
- [9] Chu Y, Li S, Tang J and Wu H (2023) The potential of the Medical Digital Twin in diabetes management: a review.
- [10] 8. Emmert-Streib F, Yli-Harja O. What is a digital twin? Experimental design for a data-centric machine learning perspective in health. *Int J Mol Sci*. (2022)
- [11] Cappon, Giacomo, et al. "Replaybg: A digital twin-based methodology to identify a personalized model from type 1 diabetes data and simulate glucose concentrations to assess alternative therapies." *IEEE Transactions on Biomedical Engineering* (2023)
- [12] Dassau, Eyal, et al. "Detection of a meal using continuous glucose monitoring: implications for an artificial β -cell." *Diabetes care* 31.2 (2008): 295-300.
- [13] Samadi, Sediqeh, et al. "Meal detection and carbohydrate estimation using continuous glucose sensor data." *IEEE journal of biomedical and health informatics* 21.3 (2017): 619-627.
- [14] M. Athanasiou, K. Zarkogianni, K. Karytsas and K. S. Nikita, "An LSTM-based Approach Towards Automated Meal Detection from Continuous Glucose Monitoring in Type 1 Diabetes

- Mellitus," 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE), Kragujevac, Serbia, 2021, pp. 1-5.
- [15] Daniels J, Herrero P, Georgiou P. A Deep Learning Framework for Automatic Meal Detection and Estimation in Artificial Pancreas Systems. *Sensors* 2022, 22(2):466.
- [16] Dietterich TG. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*. 2020, 1-15.
- [17] Palacios V, Woodbridge DMK, Fry JL. Machine Learning-based Meal Detection Using Continuous Glucose Monitoring on Healthy Participants: An Objective Measure of Participant Compliance to Protocol. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (pp. 7032-7035). IEEE. 2021.
- [18] Rodriguez, Edward, and Rodolfo Villamizar. "Artificial Pancreas: A Review of Meal Detection and Carbohydrates Counting Techniques." *The Review of Diabetic Studies: RDS* vol. 18,4 171–180. 31 Dec. 2022, doi:10.1900/RDS.2022.18.171
- [19] Bergman RN. Toward physiological understanding of glucose tolerance: minimal-model approach. *Diabetes* 1989, 38(12):1512-27.
- [20] 47. Xu J, Song R. Meal estimation based on UKF and postprandial glucose control for patients with type I postprandial glucose control for patients with type I diabetes. In *2021 China Automation Congress (CAC) 2021*, 6827-32.
- [21] Riddell, Michael C et al. "Examining the Acute Glycemic Effects of Different Types of Structured Exercise Sessions in Type 1 Diabetes in a Real-World Setting: The Type 1 Diabetes and Exercise Initiative (T1DEXI)." *Diabetes care* vol. 46,4 (2023): 704-713. doi:10.2337/dc22-1721
- [22] Wisting, Line, et al. "Eating patterns in adolescents with type 1 diabetes: associations with metabolic control, insulin omission, and eating disorder pathology." *Appetite* 114 (2017): 226-231.
- [23] Dhatariya, K.K., Glaser, N.S., Codner, E. et al. Diabetic ketoacidosis. *Nat Rev Dis Primers* 6, 40 (2020).
- [24] Redazione Roma. (2020). CGM, I sistemi di monitoraggio in continuo della glicemia. <https://www.nurse24.it/dossier/diabete/cgm-sistema-monitoraggio-continuo-glicemia.html>.
- [25] PHARMASTAR. (2023). Diabete: arrivano le penne intelligenti per registrare e monitorare la terapia con l'insulina. <https://www.pharmastar.it/news/diabete/diabete-arrivano-le-penne-intelligenti-per-registrare-e-monitorare-la-terapia-con-linsulina-40986>.
- [26] <https://www.inpatientdiabetes.org.uk/11-special-patient-groups>.
- [27] Pellizzari, Elisa, et al. "drCORRECT: An Algorithm for the Preventive Administration of Postprandial Corrective Insulin Boluses in Type 1 Diabetes Management." *Journal of Diabetes Science and Technology* (2023): 19322968231221768.

[28] Burdick, Jonathan, et al. "Missed insulin meal boluses and elevated hemoglobin A1c levels in children receiving insulin pump therapy." *Pediatrics* 113.3 (2004): e221-e224.