



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
CORSO DI LAUREA MAGISTRALE IN BIOINGEGNERIA**

**Metodi per la caratterizzazione e la predizione delle ipoglicemie
postprandiali in individui sottoposti a bypass gastrico
da dati di monitoraggio in continua del glucosio**

Relatore: Prof. Andrea Facchinetti

Laureanda: Cristina Giacomini

Correlatore: Ph.D. Francesco Prendin

ANNO ACCADEMICO 2022 - 2023

Data di laurea 27 novembre 2023

Abstract

L'obesità, in costante aumento a livello globale, comporta gravi rischi per la salute, spingendo numerosi individui a considerare interventi di chirurgia bariatrica, come il bypass gastrico o la gastrectomia verticale, per cercare una soluzione al problema e ottenere una rapida perdita di peso. Nonostante la dimostrata efficacia di tali procedure nel controllo a lungo termine del peso, emergono frequentemente complicanze post-operatorie, tra cui episodi di ipoglicemia, che possono interessare fino al 30% della popolazione post-bariatrica. Considerando che questi episodi sono spesso associati all'assunzione di pasti, l'obiettivo di questa tesi è sviluppare un algoritmo di caratterizzazione e predizione delle ipoglicemie post-prandiali. Tale algoritmo è basato non solo su feature estratte da dati di monitoraggio continuo del glucosio (CGM) ma anche su informazioni riguardanti i pasti, in modo tale da fornire una comprensione più approfondita delle dinamiche glicemiche della popolazione post-bariatrica e migliorare la gestione delle ipoglicemie post-prandiali.

Il lavoro svolto nella tesi comprende lo sviluppo, l'applicazione e la valutazione di diversi modelli di regressione logistica utilizzando un dataset contenente dati CGM provenienti da pazienti che hanno subito un intervento di bypass gastrico, fornito dall'ospedale universitario di Berna. I risultati evidenziano la capacità di prevedere in tempo reale possibili eventi ipoglicemici sfruttando i dati CGM del paziente e le informazioni sui pasti consumati. Questo approccio offre la possibilità di implementare azioni preventive e correttive, mirate a ridurre il rischio di episodi ipoglicemici.

Nella prima parte del lavoro si esplora il contesto dell'obesità, approfondendo la chirurgia bariatrica insieme alle relative procedure e complicanze. Ci si focalizza poi sulle ipoglicemie post-bariatriche e sul monitoraggio in continua della glicemia. Nel secondo capitolo viene presentato il dataset, partendo dalla descrizione del clinical trial e del dataset stesso, per poi procedere con un'analisi preliminare. Nel terzo capitolo si illustrano le feature che verranno considerate nelle analisi successive e viene creato il framework dedicato alla caratterizzazione e predizione delle ipoglicemie. Nel capitolo 4, dopo aver introdotto la regressione logistica, ci si concentra sull'analisi di correlazione, indentificando le feature maggiormente correlate all'evento ipoglicemico. Il capitolo 5 si occupa dello sviluppo e della valutazione dell'algoritmo predittivo. Infine, nel capitolo 6, dedicato alle conclusioni, vengono riassunti i risultati ottenuti, evidenziando le limitazioni e suggerendo i possibili sviluppi futuri.

Indice

| | |
|--|-----------|
| 1 Introduzione | 1 |
| 1.1 L'obesità..... | 1 |
| 1.2 La chirurgia bariatrica..... | 2 |
| 1.2.1 Bypass gastrico alla Roux-en-Y..... | 2 |
| 1.3 L'ipoglicemia post-bariatrica | 3 |
| 1.3.1 Terapie attuali..... | 4 |
| 1.4 I sistemi di monitoraggio in continua del glucosio | 5 |
| 1.5 Predizione delle ipoglicemie post-bariatriche | 5 |
| 1.6 Obiettivo della tesi..... | 6 |
| 2 Dataset | 8 |
| 2.1 Descrizione del clinical trial..... | 8 |
| 2.2 Descrizione del dataset..... | 9 |
| 2.3 Analisi preliminare del dataset | 10 |
| 2.4 Discussione dei risultati | 12 |
| 3 Estrazione delle feature e definizione del framework di predizione | 14 |
| 3.1 Estrazione di feature da dati CGM | 14 |
| 3.1.1 Feature di descrizione statistica..... | 14 |
| 3.1.2 Feature basate sulle escursioni glicemiche..... | 15 |
| 3.1.3 Feature di rischio | 17 |
| 3.1.4 Feature short time..... | 18 |
| 3.1.5 Feature di descrizione locale | 19 |
| 3.1.6 Feature temporali..... | 19 |
| 3.2 Creazione del framework per la caratterizzazione e predizione delle ipoglicemie post-bariatriche..... | 20 |
| 3.2.1 Definizione di periodo post-prandiale..... | 20 |
| 3.2.2 Finestre post-prandiali per il calcolo delle feature..... | 22 |
| 3.2.3 Preprocessing del dataset | 23 |
| 4 Caratterizzazione delle ipoglicemie | 27 |
| 4.1 Regressione logistica | 27 |
| 4.1.2 Analisi e valutazione del modello logistico..... | 29 |
| 4.2 Analisi delle ipoglicemie con modello logistico a singola variabile..... | 30 |
| 4.3 Modello completo | 34 |
| 4.4 Modello ridotto..... | 36 |
| 5 Predizione di ipoglicemie post-prandiali | 40 |
| 5.1 Framework di predizione futura delle ipoglicemie | 40 |
| 5.2 Analisi modello completo e del modello con selezione delle variabili applicati all' intero dataset | 43 |
| 5.2.1 Modello Completo..... | 43 |
| 5.2.2 Modello ridotto..... | 45 |
| 5.2.3 Performance sull'intero dataset | 47 |

| | |
|--|-----------|
| 5.3 Analisi del modello completo e del modello con selezione delle variabili applicati a training e test set | 48 |
| 5.3.1 Modello completo | 48 |
| 5.3.2 Modello ridotto | 49 |
| 5.3.3 Performance sul test set | 51 |
| 5.4 Bilanciamento delle classi: analisi modello completo e del modello con selezione delle variabili applicati all' intero dataset | 52 |
| 5.4.1 Modello Completo | 52 |
| 5.4.2 Modello Ridotto | 53 |
| 5.4.3 Performance sull'intero dataset | 54 |
| 5.5 Bilanciamento delle classi: analisi del modello completo e del modello con selezione delle variabili applicati a training e test set | 55 |
| 5.5.1 Modello Completo | 55 |
| 5.5.2 Modello Ridotto | 57 |
| 5.5.3 Performance sul test set | 58 |
| 5.6 Risultati principali | 59 |
| 6 Conclusioni | 61 |
| Appendice A | 63 |
| A.1 Distribuzione delle variabili nelle due classi | 63 |
| A.2 Curve ROC dei modelli univariati per l'analisi di correlazione | 67 |
| A.3 Curve ROC dei modelli univariati per le feature locali | 72 |
| Bibliografia | 76 |

Capitolo 1

Introduzione

1.1 L'obesità

L'obesità è una malattia metabolica di origine multifattoriale caratterizzata da un eccessivo accumulo di grasso corporeo in misura tale da compromettere lo stato di salute. Secondo l'Organizzazione Mondiale della Sanità (OMS), l'obesità rappresenta uno dei principali problemi di salute pubblica nel mondo; in Europa e negli Stati Uniti, rispettivamente, circa il 60% e il 74% della popolazione è sovrappeso o obesa [1] [2]. È noto che all'origine degli aumenti di peso vi siano cattive abitudini alimentari e stili di vita sedentari, ma influiscono anche fattori genetici, ambientali, sociali e culturali.

Per classificare il grado di sovrappeso delle persone si utilizza l'indice di massa corporea (IMC), ottenuto dividendo il peso (in Kg) per il quadrato dell'altezza (in m). Le persone considerate in sovrappeso hanno un IMC compreso tra 25 Kg/m^2 e 30 Kg/m^2 , mentre quelle considerate obese hanno un IMC uguale o maggiore di 30 Kg/m^2 .

Le malattie e disturbi correlati all'obesità includono malattie metaboliche (diabete mellito di tipo 2, dislipidemia), neoplasie, malattie cardiovascolari (ipertensione arteriosa, cardiopatia, embolia polmonare, ictus), apnee notturne, asma, steatoepatite non alcolica, colecistopatia, osteoartrite. In aggiunta, le persone che soffrono di obesità sono spesso affette da svariati disturbi psichiatrici come depressione, immagine corporea compromessa e bassa autostima che deteriorano la qualità della vita.

In base al grado di severità dell'obesità, esistono tre tipi di trattamento: modifica dello stile di vita, terapia farmacologica e chirurgia bariatrica. Il trattamento dell'obesità basato sulla modificazione dello stile di vita implica un insieme di strategie che includono una dieta mirata alla perdita di peso, un cambiamento delle abitudini comportamentali ed esercizio fisico. A coloro che non sono riusciti a perdere peso attraverso la sola modifica dello stile di vita, possono essere prescritti farmaci antiobesità. Tra i farmaci più utilizzati si menzionano: l'orlistat, che riduce l'introito calorico ostacolando la digestione dei grassi, legandosi alla lipasi gastrica e pancreatica; il lorcaserin, agisce sul sistema nervoso centrale, riducendo l'appetito mediante la stimolazione dei recettori della serotonina nei centri dell'ipotalamo coinvolti nella regolazione della sazietà e dall'appetito; infine, la sibutramina agisce riducendo l'appetito. Inoltre, è importante notare che la depressione può contribuire all'aumento di peso, quindi, il trattamento con antidepressivi potrebbe essere considerato come parte del piano terapeutico [3].

1.2 La chirurgia bariatrica

La chirurgia bariatrica, ovvero il trattamento chirurgico dell'obesità, viene effettuata solo dopo che le altre soluzioni adottate per il controllo del peso sono fallite. Non ha scopi estetici e il suo principale obiettivo è determinare una perdita di peso, mantenuta a lungo termine e di conseguenza prevenire e/o migliorare le comorbidità associate all'obesità, permettendo un allungamento dell'aspettativa di vita. I possibili candidati all'intervento chirurgico sono i soggetti con un IMC superiore a 40 Kg/m² o maggiore di 35 Kg/m² associato ad una o più patologie concomitanti. Dal momento che l'intervento comporta un drastico cambiamento delle funzioni digestive, l'assistito ha bisogno di counseling sanitario prima e dopo l'intervento. Il follow up a lungo termine del paziente ha l'obiettivo non solo di monitorare i risultati in termini di peso, qualità della vita e di controllo delle eventuali patologie associate, ma anche di prevenire le possibili complicanze post-intervento.

La chirurgia bariatrica non solo porta a una significativa perdita di peso, la cui entità dipende dal tipo specifico di intervento e dalla risposta individuale, ma comporta anche notevoli miglioramenti nelle patologie concomitanti, tra cui il diabete mellito di tipo 2, ipertensione e apnea notturna. Visti i positivi risultati ottenuti sugli adulti, oggi l'intervento è esteso anche agli adolescenti [3]. Attualmente gli interventi chirurgici vengono effettuati prevalentemente con tecnica laparoscopica, anche se, in alcuni casi può essere necessaria la tradizionale laparotomia. Essi si suddividono in:

- interventi restrittivi: agiscono riducendo la capacità dello stomaco e quindi inducono sazietà precoce. Le tecniche più frequenti includono il bendaggio gastrico regolabile e la *sleeve gastrectomy*;
- interventi malassorbitivi: intervengono sull'intestino in modo da limitare l'assorbimento delle sostanze nutritive ingerite. La tecnica malassorbitiva classica è la diversione biliopancreatica;
- interventi misti: associano un'azione restrittiva ad una malassorbitiva. La tecnica principale è il bypass gastrico [4].

1.2.1 Bypass gastrico alla Roux-en-Y

Uno degli interventi più comuni di chirurgia bariatrica è il bypass gastrico alla Roux-en-Y (RYGB), che consiste nella creazione di una piccola tasca gastrica (20-30ml) nella porzione superiore dello stomaco che viene anastomizzata con l'intestino tenue mediante un'ansa digiunale a forma di Y. Il tratto di intestino che rimane così escluso dal transito degli alimenti, chiamato "tratto biliopancreatico", viene a sua volta collegato con il "tratto alimentare" a formare il "tratto comune".

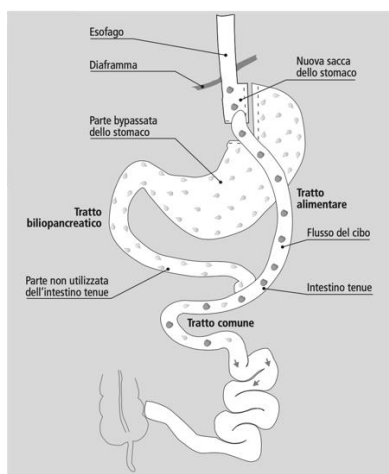


Fig.1.2.1-1 Schema del bypass gastrico alla Roux-en-Y

Con questo intervento si ottiene in media una riduzione del 70% del peso in eccesso nei primi 12-24 mesi, con ampia variabilità da caso a caso, soprattutto nel mantenimento a lungo termine. Nel meccanismo di perdita di peso influiscono la riduzione della quantità di cibo introdotto, per cui si raggiunge un precoce senso di sazietà e il bypass dello stomaco distale e del duodeno che comporta un ridotto assorbimento delle sostanze nutritive.

Tra le più note complicanze che si possono verificare dopo l'intervento vi sono: anemia da carenza di ferro e/o Vitamina B12 e/o acido folico, osteoporosi da carenza di calcio, ulcera anastomotica, intolleranza all'assunzione di alcuni cibi ed ernia interna [4] [5].

1.3 L'ipoglicemia post-bariatrica

Un'importante complicanza evidenziatasi di recente è l'insorgenza di eventi ipoglicemici, per lo più conseguenti all'assunzione di cibo, che avvengono a distanza di almeno un anno dall'intervento [6]. Per ipoglicemia si intende un valore glicemico minore di 70 mg/dl, che viene definita severa se inferiore a 54 mg/dl, per almeno 15 minuti [7].

L'ipoglicemia post-bariatrica (PBH) si verifica tipicamente da una a tre ore dopo i pasti, con aumento della gravità dopo l'assunzione di carboidrati ad alto indice glicemico. La frequenza con cui si verificano gli eventi PBH varia da soggetto a soggetto e recenti studi hanno dimostrato essere una condizione che si verifica approssimativamente nel 30% dei pazienti che si sono sottoposti all'intervento chirurgico [8] [9]. Considerando inoltre che molti pazienti risultano essere asintomatici, ovvero sperimentano episodi ipoglicemici senza manifestare sintomi evidenti, si stima una incidenza di ipoglicemie ancora più elevata [10].

Nonostante le cause dell'insorgenza di ipoglicemia siano ancora in parte sconosciute, la comparsa di eventi ipoglicemici è chiaramente legata all'intervento di bypass gastrico che ha comportato

un'alterazione nell'assorbimento di nutrienti e della risposta ormonale. Infatti, dopo l'assunzione, il cibo transita rapidamente dalla tasca gastrica all'intestino, causando un aumento repentino del glucosio nel sangue (entro 30 min) a livelli piuttosto elevati (oltre i 200 mg/dl). In risposta a ciò, vi è una secrezione eccessiva di incretine, tra cui il GLP-1 (Glucagon-Like Peptide 1), che agendo sulle cellule β del pancreas aumentano la secrezione di insulina causando una rapida diminuzione del livello di glucosio, con un nadir a distanza di poche ore dal pasto [11].

La sintomatologia varia da individuo a individuo e a seconda della gravità dell'evento; tipicamente i sintomi sono classificati come autonomici (tremore, ansia, palpitazioni e sudorazione) e neuroglicopenici (vertigini, confusione, difficoltà a concentrarsi e visione alterata). Nel caso di ipoglicemia severa, possono verificarsi convulsioni, perdita di coscienza, cadute, incidenti automobilistici e persino la morte. La disabilità associata e la compromissione della qualità della vita possono essere parecchio gravi e la condizione tende a perdurare nel tempo. Pertanto, la gestione efficace dell'ipoglicemia è di vitale importanza per prevenire tali conseguenze [12]. Nel caso di una crisi di ipoglicemia di grado lieve-moderato, viene consigliata l'assunzione di carboidrati semplici come pastiglie o gel di glucosio per riportare la glicemia nel target range.

1.3.1 Terapie attuali

In assenza di una terapia farmacologica approvata, il trattamento principale di questa patologia consiste in una modifica della dieta. Viene raccomandato di assumere carboidrati a basso indice glicemico, di evitare carboidrati a rapido assorbimento, di prediligere proteine e grassi salutari, di evitare alcol e bevande durante i pasti e di pianificare pasti regolari durante il giorno.

Oltre ad una dieta calibrata, potrebbe essere necessario anche un trattamento medico aggiuntivo, volto a ridurre i picchi di glucosio e lo stimolo per la secrezione di insulina. Terapie farmacologiche di base prevedono l'utilizzo di uno o più farmaci come l'acarbosio, il diazossido e l'ocreatide. In particolare, l'acarbosio ritarda e limita l'assorbimento di glucosio inibendo l' α -glucosidasi; il diazossido riduce la secrezione di insulina mediante l'inibizione dei canali potassio ATP-sensibili delle cellule β ; infine, l'ocreatide è in grado di diminuire la secrezione di insulina e di GLP-1 legandosi con i sottotipi 2 e 5 dei recettori di somatostatina.

Nel caso in cui, sia la dieta che la terapia farmacologica falliscano, vengono presi in considerazione possibili interventi chirurgici, come l'utilizzo di un tubo gastrostomico per immettere cibo nella regione dello stomaco bypassata o un completo ripristino delle condizioni originarie mediante la rimozione del RYGB [13].

1.4 I sistemi di monitoraggio in continua del glucosio

Il monitoraggio in continua del glucosio (CGM), oggigiorno utilizzato ampiamente per il controllo della glicemia in pazienti diabetici, può essere uno strumento utile anche per il trattamento delle ipoglicemie post-bariatriche.

I sensori CGM hanno lo scopo di misurare la concentrazione di glucosio nel fluido interstiziale, tipicamente ogni 5 min, fornendo quindi l'andamento della curva glicemica in tempo reale per diversi giorni. Sono costituiti da un sensore elettrochimico ad ago sottocutaneo, collegato ad un trasmettitore wireless in grado di trasmettere la misura al ricevitore, che può essere un comune smartphone [14].



Fig. 1.4-1 Dexcom G6: un sensore per il monitoraggio in continua del glucosio

Il dispositivo CGM non solo fornisce la concentrazione del glucosio nel sangue dell'individuo in real-time, ma anche trends, velocità di variazione della glicemia e allarmi al paziente, utili per la rilevazione di eventi ipoglicemici e quindi per il trattamento precoce e la prevenzione di essi. La disponibilità di dati fornita dal sensore è un fattore chiave che consente di sviluppare metodologie per migliorare la gestione e l'insorgenza delle ipoglicemie postprandiali.

1.5 Predizione delle ipoglicemie post-bariatriche

L'utilizzo dei dati CGM per la predizione della glicemia futura è un campo di ricerca relativamente maturo nell'ambito del diabete di tipo 1 (*Type 1 Diabetes – T1D*), con svariati contributi di letteratura che esplorano l'utilizzo di numerosi algoritmi e diverse tipologie di input. Al contrario, questo argomento è stato solo parzialmente investigato nell'ambito dei soggetti post-bariatrici, data la relativa novità di questa patologia.

Il primo contributo nel campo della rilevazione e predizione dell'ipoglicemia post-bariatrica è stato dato dal *clinical study* condotto da Laguna Sanz et al. [15]. Lo studio ha sviluppato un algoritmo

predittivo per l'ipoglicemia legato ad un sistema di somministrazione di glucagone che si attiva in caso di allarme ipoglicemico. In particolare, l'algoritmo prevede l'azione contemporanea di due blocchi di rilevazione di ipoglicemie per la generazione degli allarmi: un blocco LGP, che rileva imminenti episodi di ipoglicemia; ed un blocco PBH, che si attiva quando viene consumato un pasto e prevede l'insorgenza di ipoglicemia fino a 30 minuti prima che si verifichi. La combinazione di questi allarmi PBH e LGP consente quindi di prevenire eventuali cali del livello di glucosio. Tuttavia, il lavoro presenta alcune limitazioni, tra cui la dimensione ridotta del *clinical trial* ed il fatto che il protocollo veniva modificato ad ogni fase, in modo da testare le iterazioni dell'algoritmo per la rilevazione di ipoglicemie e per aumentare le dosi di glucagone. Inoltre, lo studio è stato condotto e validato in un ambiente altamente controllato, di conseguenza i risultati sono difficilmente applicabili nella vita quotidiana.

Un recente lavoro di Prendin et al. [16] (collaborazione tra il Dipartimento di Ingegneria dell'Informazione dell'Università di Padova e l'Ospedale Universitario di Berna) ha valutato la possibilità di utilizzare dati CGM e modelli *black box* per predire eventi PBH. In particolare, sono stati proposti tre diversi algoritmi predittivi: un modello autoregressivo (AR1), un modello autoregressivo integrato a media mobile (ARIMA) e una rete neurale *feed-forward*. Tra questi algoritmi, impiegati solitamente per i pazienti affetti da T1D, il modello ARIMA ha presentato la migliore performance di predizione di eventi ipoglicemici. Anche in questo studio sono emerse alcune limitazioni come: la necessità di eseguire una pre-elaborazione sui dati CGM a causa del rumore nel segnale, l'utilizzo di un test set composto solamente da 8 soggetti, l'assenza di informazioni riportate dai soggetti riguardo le azioni correttive assunte dai pazienti.

1.6 Obiettivo della tesi

Lo scopo di questa tesi magistrale è quello di caratterizzare gli eventi di ipoglicemia postprandiale e di superare le attuali limitazioni dei contributi presenti nello stato dell'arte in merito alla gestione delle ipoglicemie in pazienti che hanno subito un intervento di bypass gastrico. Partendo dalla caratterizzazione degli eventi PBH e dall'estrazione di feature da dati CGM che abbiano un potere predittivo, l'obiettivo finale è sviluppare un algoritmo, basato su modelli di regressione logistica, per la predizione in real-time degli eventi PBH, che sfrutti le feature estratte dall'analisi di dati CGM in una finestra temporale antecedente l'episodio stesso.

In seguito ad una fase iniziale di pre-elaborazione del dataset, si procederà all'individuazione e alla selezione dei periodi post-prandiali dai dati CGM, da cui verranno estratte le caratteristiche più rilevanti ai fini dell'analisi. A ciascun periodo verrà associata la classe 1, nel caso in cui il pasto conduca a un evento ipoglicemico, o la classe 0 in caso contrario.

Al fine di identificare le caratteristiche più significative e con maggiore potere predittivo per gli eventi ipoglicemici, sarà eseguita un'analisi di correlazione. Infine, verrà sviluppato e successivamente convalidato un modello logistico finalizzato alla previsione delle ipoglicemie. Inizialmente verrà utilizzato l'intero dataset per l'addestramento del modello, e in una fase successiva il dataset verrà suddiviso in training e test set, con un'ulteriore operazione di bilanciamento delle classi. Attraverso l'impiego di questo modello, che sfrutta i dati CGM e le informazioni sui pasti fornite dai pazienti, sarà possibile prevedere in tempo reale e, quindi, prevenire eventuali episodi di ipoglicemia connessi all'assunzione del pasto.

Capitolo 2

Dataset

2.1 Descrizione del clinical trial

Per la caratterizzazione e lo sviluppo dell' algoritmo di predizione delle ipoglicemie, sono stati utilizzati i dati CGM raccolti da uno studio clinico (ID: NCT05212207) condotto dalla Professoressa Lia Bally presso il Dipartimento di Diabetologia, Endocrinologia, Clinica Nutrizionale e Metabolismo dell'Ospedale Universitario di Berna (Inselspital), in Svizzera. Le serie CGM raccolte si riferiscono a pazienti adulti che hanno subito un intervento di bypass gastrico alla Roux-en-Y e sperimentano eventi ipoglicemici (glucosio <70 mg/dL) a seguito di pasti caratterizzati da un elevato indice glicemico. Dallo studio clinico sono stati esclusi i pazienti: con l'impossibilità di fornire il consenso informato, in gravidanza o in allattamento, controindicati o non in grado di partecipare allo studio (a causa di problemi linguistici, disturbi psicologici ecc.), sotto cura di farmaci che interferiscono con la glicemia. I soggetti reclutati, oltre ad utilizzare il sensore CGM che acquisiva il valore glicemico ogni 5 min, indossavano anche uno smartwatch e annotavano informazioni riguardanti i pasti, il sonno e l'attività fisica tramite la piattaforma IMPACT (Integrated Mobile Platform for Automated Clinical Trials), sviluppata dal team di ricerca di Bioingegneria del DEI [17]. IMPACT consente di raccogliere e strutturare facilmente i dati tramite un app mobile personalizzabile. La piattaforma offre anche un'interfaccia web di telemonitoraggio per consentire ai medici di monitorare lo stato dei pazienti in tempo reale durante i *clinical trials*.

Il dataset è costituito dalle informazioni presenti in *Tab 2.1-1* raccolte per ogni soggetto.

| Variabili | Descrizione |
|------------------|---|
| Time | Data e ora (in formato <i>datetime</i>) |
| glucose | Valori glicemici registrati dal sensore CGM [mg/dl] |
| rate | Tasso di variazione (<i>Rate of change</i>) ottenuto dal sensore CGM [mg/dL/min] |
| smbg | Automonitoraggio glicemico (<i>Self-monitoring of blood glucose</i>) [mg/dl] |
| meal_type | Tipo di pasto (0: alcol, 1: trattamento di ipoglicemia, 2: pasto con quantità di carboidrati < 30g, 3: pasto con quantità di carboidrati > 30g) |
| meal_note | Informazioni sul pasto riportate dal paziente |
| meal | Riporta le informazioni di <i>meal_type</i> tramite un codice |

| | |
|----------------------------|---|
| symptom | Tipo di sintomo (ad es. gastrointestinale, neuroglicopenico, vasomotore) |
| symptom_intensity | Intensità dei sintomi |
| symptom_note | Informazioni sui sintomi |
| mealFlagged | Contiene le stesse informazioni riportate dalla variabile <i>meal</i> + i pasti segnati dai clinici |
| steps_count | Numero di passi |
| sensor_state | Stato del sensore CGM |
| sensor_command_sent | Comando inviato dal sensore CGM (Warmup, InCalibration) |
| activity_intensity | Intensità dell'attività fisica (varia da 0 a 9) |
| activity_type | Tipo di attività (ad es. housework) |
| sleep | Intensità del sonno (varia da 1 a 3) |
| drug | Nome e quantità dei farmaci assunti |
| patient_note | Testo libero riportato dal soggetto |

Tab. 2.1-1 Descrizione delle variabili raccolte dallo studio clinico.

2.2 Descrizione del dataset

Il dataset originario è formato da 50 soggetti, di cui per ciascun soggetto sono riportate alcune informazioni personali (come genere, età, peso e altezza), la traccia CGM e le altre informazioni riguardanti i pasti, il sonno e l'attività fisica (Tab 2.1-1).

| | Rapporto | Media (SD) | Mediana (25° - 75° percentile) |
|--|-----------------|-------------------|---|
| Età | / | 50.32 (12.27) | / |
| Genere (maschile : femminile) | 11:39 | / | / |
| Giorni di monitoraggio | / | 49.07 (11.99) | 49.95 (49.48 - 52.17) |

Tab. 2.2-1 Caratteristiche dei soggetti reclutati.

L'intero lavoro di tesi è stato svolto utilizzando il software MATLAB; in particolare, per l'analisi delle tracce CGM, è stato utilizzato il pacchetto AGATA (Automated Glucose dATa Analysis), sviluppato da Cappon et al.[18]. Questo strumento consente di elaborare, analizzare e visualizzare agevolmente serie di dati CGM. Il pacchetto offre una serie di funzioni utili per eseguire il pre-processing dei dati, calcolare metriche relative al controllo glicemico, rilevare eventi avversi (iper/ipoglicemici), fornendo la possibilità di visualizzare i risultati. Inoltre, AGATA permette

anche di calcolare metriche di errore di previsione tra i dati originali e i profili glicemici predetti, eseguire un'analisi automatica di un insieme di dati e comparare due tracce differenti.

Di seguito, viene riportata una porzione di una traccia CGM visualizzata utilizzando il pacchetto AGATA.

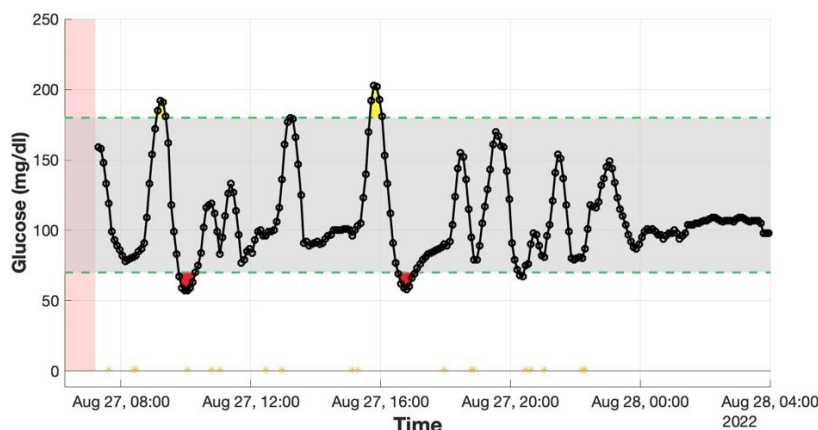


Fig. 2.2-2 Esempio di traccia CGM (in nero), con eventi ipoglicemici (in rosso) e iperglicemici (in giallo), periodi con missing values (banda rossa), il range euglicemico (linee tratteggiate verdi) e i momenti dei pasti (asterischi).

Il segnale CGM è rappresentato in figura dalla linea nera, gli eventi ipoglicemici (glucosio <70 mg/dl) e iperglicemici (glucosio >180 mg/dl) sono evidenziati rispettivamente mediante aree di colore rosso e giallo, i periodi con assenza di dati sono rappresentati con bande rosse e le due linee tratteggiate verdi indicano il range glicemico di normalità (70-180 mg/dl). Sono stati inseriti nell'immagine anche degli asterischi gialli per rappresentare i momenti dei pasti.

2.3 Analisi preliminare del dataset

Come analisi preliminare sono state calcolate alcune metriche, accompagnate dai relativi boxplot, su tutte le serie CGM in modo da evidenziare alcune caratteristiche ed anomalie delle tracce.

Le caratteristiche prese in considerazione comprendono:

- indici di variabilità glicemica: media, deviazione standard e range glicemico.
- variabili temporali: forniscono la percentuale di tempo trascorso in specifici intervalli glicemici. Comprendono il tempo in ipoglicemia (glucosio <70 mg/dl), ipoglicemia L1 (54-70 mg/dl) e L2 (<54 mg/dl); in iperglicemia (glucosio >180 mg/dl), iperglicemia L1 (181-250 mg/dl) e L2 (>250 mg/dl); e in euglicemia (70-180 mg/dl).
- eventi di ipoglicemia e iperglicemia: comprendono l'identificazione temporale e la durata media degli eventi ipoglicemici (L1 e L2) e iperglicemici (L1 e L2).
- eventi avversi ai pasti: includono la percentuale di eventi correlati ai pasti e viceversa. Ad esempio, un evento di ipoglicemia viene considerato causato dal pasto se si verifica entro tre ore

dopo l'assunzione. Nota: i pasti considerati in questa analisi sono quelli segnati nella categoria 'mealFlagged' o in 'meal'.

- indici di variabilità post-prandiale: media del glucosio post e pre-prandiale. La media post-prandiale viene calcolata facendo la media dei valori glicemici registrati a tre ore dal pasto; mentre la media pre-prandiale viene calcolata mediando i valori nell'ora precedente il pasto.

Si riportano in *Tab 2.3-1* i valori di mediana, venticinquesimo e settantacinquesimo percentile delle variabili considerate.

| | Mediana | 25° percentile | 75° percentile |
|--|----------------|-----------------------|-----------------------|
| Media | 106.21 | 102.85 | 116.85 |
| SD | 32.39 | 28.29 | 38.85 |
| Range | 260.00 | 225.00 | 312.50 |
| Time in hyperglycemia | 4.13 | 2.45 | 7.44 |
| Time in L1 hyperglycemia | 3.78 | 2.38 | 6.57 |
| Time in L2 hyperglycemia | 0.27 | 0.06 | 1.32 |
| Time in hypoglycemia | 4.30 | 2.66 | 7.62 |
| Time in L1 hypoglycemia | 3.33 | 2.12 | 5.69 |
| Time in L2 hypoglycemia | 0.97 | 0.46 | 1.44 |
| Time in Target | 89.91 | 86.92 | 93.07 |
| Durata media eventi iperglicemici non severi | 29.93 | 26.88 | 32.81 |
| Durata media eventi iperglicemici severi | 25.41 | 21.25 | 30.75 |
| Durata media eventi ipoglicemici non severi | 29.69 | 27.37 | 37.50 |
| Durata media eventi ipoglicemici severi | 28.30 | 23.82 | 34.31 |
| %eventi ipoglicemici non severi legati ai pasti mealFlagged | 64.87 | 57.47 | 75.00 |
| %pasti mealFlagged legati a eventi ipoglicemici non severi | 13.77 | 10.73 | 20.44 |
| %eventi ipoglicemici severi legati ai pasti mealFlagged | 70.73 | 56.86 | 87.63 |

| | | | |
|--|--------|--------|--------|
| %pasti <i>mealFlagged</i> legati a eventi ipoglicemici severi | 6.01 | 2.46 | 10.05 |
| %pasti <i>mealFlagged</i> legati a eventi ipoglicemici | 19.35 | 12.17 | 30.60 |
| Media glu pre-pasto <i>mealFlagged</i> | 102.78 | 98.49 | 112.04 |
| Media glu post-pasto <i>mealFlagged</i> | 106.83 | 104.08 | 115.85 |
| %eventi ipoglicemici non severi legati ai pasti <i>meal</i> | 43.78 | 33.33 | 57.14 |
| %pasti <i>meal</i> legati a eventi ipoglicemici non severi | 12.66 | 8.84 | 18.88 |
| %eventi ipoglicemici severi legati ai pasti <i>meal</i> | 52.79 | 40.00 | 66.67 |
| %pasti <i>meal</i> legati a eventi ipoglicemici severi | 5.15 | 2.60 | 9.73 |
| %pasti <i>meal</i> legati a eventi ipoglicemici | 18.38 | 11.51 | 27.81 |
| Media glu pre-pasto <i>meal</i> | 103.73 | 98.55 | 113.45 |
| Media glu post-pasto <i>meal</i> | 110.28 | 104.19 | 118.79 |

Tab 2.3-1 Valori di mediana, 25° e 75° percentile delle variabili selezionate.

Dall'osservazione delle serie glicemiche e dai risultati ottenuti emerge che, in generale, i soggetti presentano un numero limitato di eventi sia iperglicemici che ipoglicemici, e in accordo con la letteratura, le dinamiche della glicemia risultano essere rapide [16].

2.4 Discussione dei risultati

Dall'analisi dei valori di media, SD e range glicemico si deduce che il valore medio di glicemia è nella norma, ma la deviazione standard elevata e il range ampio indicano un'alta variabilità nei livelli glicemici ed una differenza significativa tra i valori massimi e minimi di glicemia.

Osservando i valori delle variabili temporali, si nota che per la maggior parte del tempo la glicemia si trova nel range ottimale (70-180 mg/dl), con episodi sporadici, ma duraturi di ipoglicemia e iperglicemia.

Focalizzandosi sulle ipoglicemie, emerge che la maggioranza dei pasti non comporta eventi ipoglicemici, ma la maggior parte delle ipoglicemie è causata dall'assunzione del pasto. Infatti, la percentuale di ipoglicemie causate dai pasti risulta abbastanza elevata (67% per *mealFlagged* e 48% per *meal*), a differenza della percentuale di pasti che causano eventi di ipoglicemia (19% per *mealFlagged* e 18% per *meal*).

Nota: andando a confrontare il numero di ipoglicemie legate ai pasti con il numero di pasti che causano ipoglicemie, è possibile ottenere valori differenti, poiché singoli eventi ipoglicemici possono essere causati da più pasti ravvicinati.

Infine, dall'analisi delle distribuzioni relative alle variabili 'durata media eventi ipoglicemici', 'giorni di monitoraggio' e 'percentuale di missing values' emergono alcune anomalie presenti nelle tracce CGM. Tra le più rilevanti, si menzionano:

- la rilevazione di falsi eventi ipoglicemici, principalmente durante la notte, dati da artefatti di compressione;
- la presenza di gap all'interno della serie glicemica causati da errori di connessione Bluetooth tra il sensore CGM, il trasmettitore e IMPACT, che avvengono in concomitanza con cambiamenti repentini della concentrazione di glucosio.

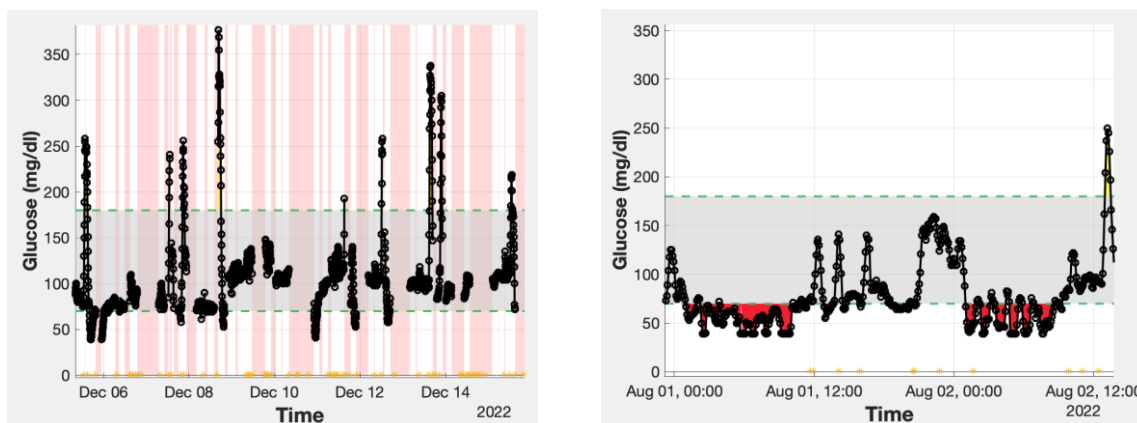


Fig. 2.4.1 Esempi di tracce CGM, raffiguranti le anomalie più frequenti: periodi di dati mancanti (a sinistra) e falsi eventi ipoglicemici causati da artefatti di compressione (a destra).

Per risolvere queste problematiche sono stati esclusi dall'analisi successiva: il soggetto 33 poiché presentava un numero insufficiente di campioni CGM e i soggetti che presentavano una percentuale di valori mancanti all'interno della traccia CGM superiore al 60% (soggetto 6, 11, 20 e 21).

Di conseguenza, il dataset si è ridotto a 45 soggetti. Nel caso del soggetto 34, è stata considerata solo la parte finale del tracciato, in quanto la prima parte presentava una percentuale elevata di *missing values*. Per gestire i periodi di *missing values* con durata inferiore o uguale a 20 minuti, è stata applicata un'interpolazione lineare.

In riferimento alle ipoglicemie notturne (23pm-6am), quelle legate all'assunzione dei pasti sono state lasciate nel dataset, mentre per quelle non causate dai pasti, i valori glicemici corrispondenti sono stati sostituiti con 'NaN'.

Capitolo 3

Estrazione delle feature e definizione del framework di predizione

Questo capitolo è composto da due macro-sezioni: nella prima verranno illustrate le variabili impiegate per la caratterizzazione e lo sviluppo dell' algoritmo di predizione delle ipoglicemie post-bariatriche; nella seconda si presenteranno le azioni di processing per estrarre le feature dalla serie temporale CGM e creare un framework di classificazione. Come menzionato in precedenza, per l'estrazione delle variabili è stato utilizzato il pacchetto MATLAB *AGATA*.

3.1 Estrazione di feature da dati CGM

3.1.1 Feature di descrizione statistica

Le feature di descrizione statistica sono variabili ottenute da serie di dati che forniscono informazioni statistiche sulla distribuzione e la variabilità dell'insieme. Vengono ampiamente impiegate nell'analisi di dati per riassumere, valutare ed estrarre informazioni rilevanti.

- **Mean**: esprime il livello medio del glucosio e si calcola come media dei valori CGM.

$$m = \frac{1}{N} \sum_{k=1}^N BG(k)$$

- **Median**: dato un insieme di dati CGM disposti in ordine crescente, la mediana è quel valore che occupa la posizione centrale dell'insieme, separando l'insieme in due parti uguali.

$$me = \begin{cases} BG\left(\frac{N+1}{2}\right) & \text{se } N \text{ è dispari} \\ \frac{BG\left(\frac{N}{2}\right) + BG\left(\frac{N}{2} + 1\right)}{2} & \text{se } N \text{ è pari} \end{cases}$$

- **Standard Deviation (SD):** stima la variabilità dell'insieme di dati CGM. Viene calcolato come:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (BG(k) - m)^2}$$

- **Coefficient of Variation (CV):** è definito in percentuale come rapporto tra la deviazione standard e la media della serie CGM:

$$CV_{\%} = \frac{\sigma}{m} * 100$$

- **Range:** è definito come la differenza tra il valore massimo e minimo della serie CGM.

$$\Delta BG = \max [BG(k)] - \min[BG(k)]$$

- **J-index:** tiene conto sia della media e che della variabilità glicemica e viene calcolato come:

$$J = 0.001 * (m + \sigma)^2$$

con la concentrazione di BG in mg/dL [19].

3.1.2 Feature basate sulle escursioni glicemiche

Le feature basate sulle escursioni glicemiche sono misure utilizzate per caratterizzare le fluttuazioni della glicemia, fornendo informazioni sull'entità e la frequenza delle variazioni.

- **MAGE index (*Mean Amplitude of Glycemic Excursion*):** calcolato mediando il valore di tutte le escursioni glicemiche significative, solitamente maggiori di una certa soglia (ad esempio la SD) [20].

Esiste però un calcolo alternativo ed automatico per il MAGE preferibile a quello classico nel caso di segnali CGM [21]:

- Per ciascuna traccia CGM si identificano gli estremi locali chiamati *turning points*.
- Viene fatta una pulizia dell'insieme dei *turning points*, eliminando i punti che non sono significativamente diversi dai loro vicini e quelli che fanno parte di una sequenza monotona di tre punti.

- Per il calcolo delle escursioni glicemiche si sottrae a ciascun elemento dei *turning points* il suo precedente.
- Vengono calcolati quindi gli indici $MAGE+$ e $MAGE-$ rispettivamente come media delle escursioni positive e negative.

$$MAGE+ = \frac{1}{N_{turn}^+} \sum_{\Delta BG_d(k) > 0} \Delta BG_d(k)$$

$$MAGE- = \frac{1}{N_{turn}^-} \sum_{\Delta BG_d(k) < 0} |\Delta BG_d(k)|$$

- Il valore finale di $MAGE$ è dato dalla media degli indici $MAGE+$ e $MAGE-$:

$$MAGE = \frac{MAGE+ + MAGE-}{2}$$

- **EF index (*Excursion Frequency*)**: fornisce informazioni complementari a quelle dell'indice $MAGE$ individuando il numero di escursioni glicemiche significative (>75 mg/dl) nel profilo CGM. Viene calcolato come [22]:

$$EF = \sum_{|\Delta BG(k)| > 75} 1$$

Nell'analisi sono anche stati considerati i seguenti indici di escursione glicemica:

- **gluPos**: La somma degli incrementi positivi del profilo CGM nella finestra;
- **gluNeg**: La somma degli incrementi negativi del profilo CGM nella finestra;
- **maxDiff**: La massima variazione tra un campione e quello successivo, considerando il segno;
- **minDiff**: La minima variazione tra un campione e quello successivo, considerando il segno;
- **gluMax**: Il valore massimo di glucosio della serie CGM nella finestra;
- **last local max**: l'ultimo massimo locale presente nella finestra;
- **prominence**: La differenza tra l'ultimo massimo locale del profilo e il minimo locale che lo precede.

3.1.3 Feature di rischio

Le feature di rischio sono utilizzate per quantificare e valutare il rischio di eventi ipoglicemici e iperglicemici in relazione all'andamento e alle variazioni dei livelli di glucosio nel sangue.

- **BGRI (Blood Glucose Risk Index)**: si focalizza sia sulla variabilità glicemica che sul rischio di ipo/iperglicemia [23].

- Per prima cosa, ogni misura $BG(k)$ viene trasformata utilizzando la formula:

$$f(BG(k)) = \gamma * [\ln(BG(k))^\alpha - \beta]$$

Con $\alpha = 1.084$, $\beta = 5.381$ e $\gamma = 1.509$ se $BG(k)$ è espresso in mg/dl.

- Si calcola la funzione di rischio tramite la formula:

$$r(BG(k)) = 10 * f(BG(k))^2$$

separando due casi:

$$r_l(BG(k)) = r(BG(k)) \quad \text{se } f(BG(k)) < 0$$

$$r_h(BG(k)) = r(BG(k)) \quad \text{se } f(BG(k)) > 0$$

- Vengono quindi calcolati LBGI e HBGI come media dei rischi $r_l(BG)$ e $r_h(BG)$, rispettivamente:

$$LBGI = \frac{1}{N} \sum_{k=1}^N r_l(BG(k))$$

$$HBGI = \frac{1}{N} \sum_{k=1}^N r_h(BG(k))$$

- La somma di LBGI e HBGI costituisce il *Blood Glucose Risk Index*:

$$BGRI = LBGI + HBGI$$

- **ADRR (Average Daily Risk Range)**: si ottiene utilizzando i rischi calcolati precedentemente, sommando i valori massimi dei rischi $r_l(BG)$ e $r_h(BG)$ [24]:

$$ADRR = \max r_l(BG(k)) + \max r_h(BG(k))$$

- **Dinamic Risk (DR)**: è un indice strettamente legato al rischio statico che considera anche il tasso di variazione (*Rate Of Change*), consentendo di modulare l'entità del rischio di ipoglicemia e iperglicemia a seconda dell'andamento glicemico [25].

Definendo il rischio statico come:

$$SR(BG(k)) = r_h(BG(k)) - r_l(BG(k))$$

Il rischio dinamico (DR) risulta:

$$DR\left(BG(k), \frac{dBG(k)}{dt}\right) = \begin{cases} SR(BG(k)) * e^{+\mu \frac{dr}{dt}} & \text{se } SR(BG(k)) > 0 \\ SR(BG(k)) * e^{-\mu \frac{dr}{dt}} & \text{se } SR(BG(k)) < 0 \end{cases}$$

$$\text{con } \frac{dr}{dt} = \left\{ 10\gamma^2 * [(\ln(BG(k)))^{2\alpha-1} - \beta(\ln(BG(k)))^{\alpha-1}] * 2\alpha \frac{1}{BG(k)} \right\} * \frac{dBG(k)}{dt}$$

e μ è un parametro da definire, che determina quanto il rischio statico SR viene amplificato o smorzato. Nel caso di pazienti PBH, a causa della diversa distribuzione dei valori glicemici rispetto ai soggetti diabetici di tipo 1, sono stati utilizzati altri valori per i parametri alfa, beta e gamma: $\alpha=2.5077$, $\beta=49.9276$ e $\gamma=0.0923$.

3.1.4 Feature short time

Le feature short-time vengono calcolate sulla porzione finale del tracciato, in modo da quantificare il trend glicemico.

- **Glucose ROC (*Glucose Rate Of Change*)**: fornisce una stima della derivata della serie BG. Al tempo t, è definita come la differenza tra il campione BG al tempo t ed il campione al tempo t-15, diviso per l'intervallo temporale tra i due campioni (15 minuti); pertanto, la ROC relativa al campione k-esimo è pari a:

$$ROC(k) = \frac{BG(k) - BG(k-3)}{15}$$

Nell'analisi è stato considerato l'ultimo valore di ROC (**lastROC**), ovvero la differenza tra l'ultimo campione ed il campione a 15 minuti prima, il tutto diviso 15; e la media della ROC (**meanROC**).

- **Differenza a X minuti:** è definita come la differenza tra l'ultimo campione della finestra ed il campione registrato X minuti prima. Sono state considerate le differenze a 10, 20 e 30 minuti (**diff10**, **diff20** e **diff30**).

3.1.5 Feature di descrizione locale

Le feature di descrizione locale comprendono una serie di misure specifiche per caratterizzare la glicemia in relazione ai pasti e ai sintomi.

- **Meal2Peak Diff:** la differenza tra il valore di glucosio al picco e quello al pasto, tenendo conto solo delle differenze positive;
- **Meal2Peak Time:** distanza temporale tra picco e pasto;
- **peakAngle:** angolo al picco;
- **mealGlu:** valore di glucosio al pasto;
- **lastGlu:** ultimo valore di glucosio della serie nella finestra;
- **mealTime:** ora a cui avviene il pasto;
- **numMeals:** numero di pasti presenti nella finestra;
- **typeMeal:** tipo di pasto, legato alla quantità di carboidrati (1: carbs>30 g / 0: carbs<30g);
- **symptoms:** sintomi presenti nella finestra (1: sono presenti sintomi / 0: non sono presenti).
-

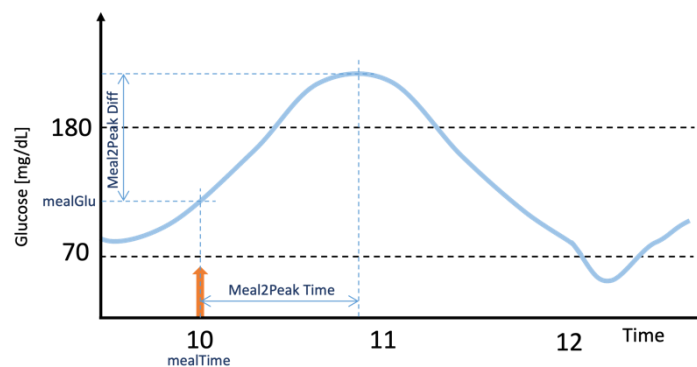


Fig. 3.1.5-1 Esempio di traccia CGM, con evidenziate le feature: Meal2Peak Diff, Meal2Peak Time, mealGlu e mealTime.

3.1.6 Feature temporali

Le feature temporali sono misure utilizzate per quantificare la durata in cui il segnale CGM si trova in determinati intervalli glicemici.

- **TAR130:** percentuale di campioni CGM con concentrazione di glucosio superiore a 130 mg/dl;
- **TAR180:** percentuale di campioni CGM con concentrazione di glucosio superiore a 180 mg/dl.

3.2 Creazione del framework per la caratterizzazione e predizione delle ipoglicemie post-bariatriche

L'estrazione delle feature illustrate nel paragrafo precedente non avviene sull'intera serie glicemica, ma solamente su opportune finestre temporali che definiscono il periodo post-prandiale.

Il primo passo riguarda la segmentazione della serie glicemica in periodi post-prandiali, grazie all'informazione del pasto riportata dai soggetti. Inoltre, dato che non tutti i pasti causano eventi ipoglicemici è stata introdotta una variabile aggiuntiva, denominata 'Y', che associa il valore 1 se il pasto comporta un evento ipoglicemico e 0 nel caso contrario. Questa variabile 'Y' definisce le due classi del problema e sarà usata come variabile *target* durante il training dei modelli di descrizione dell'ipoglicemia.

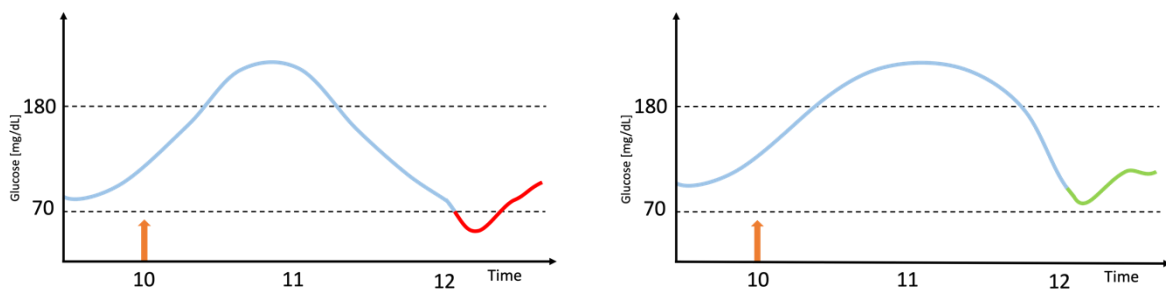


Fig. 3.2-1 Esempi di tracce CGM: caso $Y=1$ (a sinistra) e caso $Y=0$ (a destra), con momento del pasto messo in evidenza (freccia arancione).

3.2.1 Definizione di periodo post-prandiale

La definizione di periodo post-prandiale in cui si verifica un episodio di ipoglicemia ($Y=1$) è semplice da determinare: l'inizio è fissato a 30 minuti prima del pasto e la fine al nadir dell'evento ipoglicemico, che rappresenta il valore minimo della glicemia raggiunto dal soggetto durante l'ipoglicemia e indica l'istante temporale in cui il soggetto ha intrapreso un'azione correttiva nei confronti dell'evento avverso (Fig. 3.2.1-1).

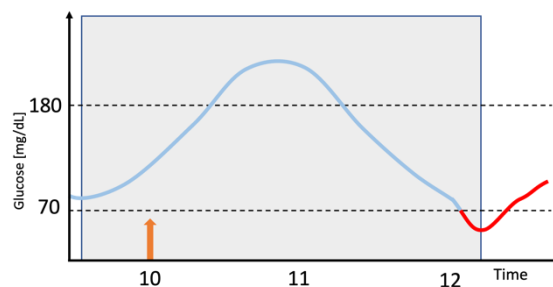


Fig. 3.2.1-1 Esempio di traccia CGM, caso $Y=1$ con periodo post-prandiale in evidenza (finestra grigia).

La definizione di periodo post-prandiale nel caso in cui il pasto non causa un evento di ipoglicemia ($Y=0$) richiede di identificare il punto di minimo, caratterizzato da una variazione della concavità della traccia, all'interno di una finestra temporale di 3 ore dopo il pasto (Fig. 3.2.1-2). Il periodo post-prandiale inizia quindi 30 minuti prima del pasto e termina in coincidenza del punto di minimo identificato (Fig. 3.2.1-3).

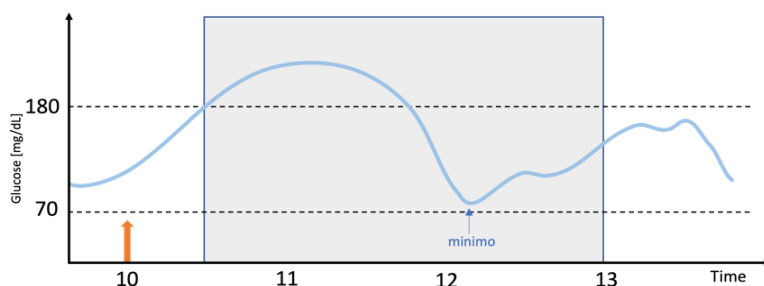


Fig. 3.2.1-2 Esempio di traccia CGM, caso $Y=0$ con finestra da 30 min post-pasto a 3 ore post-pasto. In evidenza il punto di minimo nella finestra.

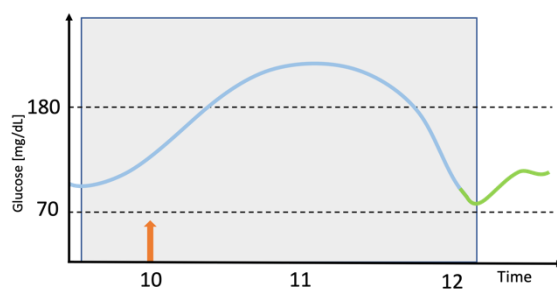


Fig. 3.2.1-3 Esempio di traccia CGM, caso $Y=0$ con periodo post-prandiale in evidenza (finestra grigia).

Bisogna, inoltre, considerare che possono verificarsi altri pasti ad una distanza temporale inferiore alle 3 ore utilizzate in fase di definizione del periodo post-prandiale. In questo caso, si procede partendo dal momento del pasto e analizzando la traccia CGM da lì fino a tre ore dopo:

- se è presente soltanto il pasto in questione, la finestra finale viene costruita su quel pasto;
- se sono presenti altri pasti entro un'ora dal pasto precedente, la finestra finale viene costruita sul primo pasto;
- se si verifica un secondo pasto a distanza maggiore di 1 ora (e minore di 3 ore) dal primo, la finestra finale viene costruita sul secondo pasto;
- nel caso in cui siano presenti altri pasti a distanza maggiore di 1 ora (e minore di 3 ore) dal precedente, la finestra viene costruita basandosi sull'ultimo pasto. (Fig. 3.2.1-4)

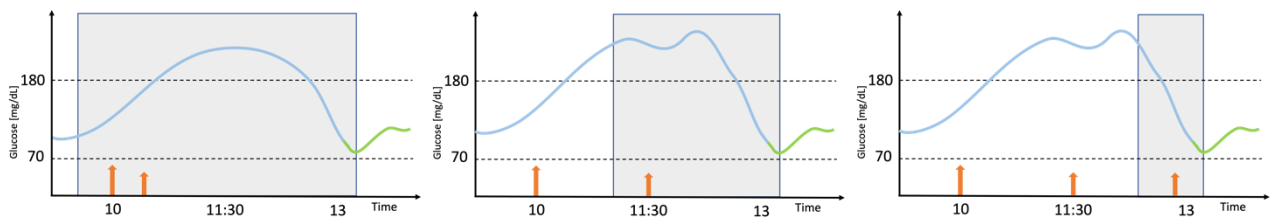


Fig. 3.2.1-4 Esempi di tracce CGM rappresentanti tre diverse costruzioni della finestra: il primo caso in cui è presente un secondo pasto a distanza < 1 ora dal precedente (a sinistra); il caso in cui è presente un secondo pasto a distanza > 1 ora e < 3 ore dal precedente (al centro); e l'ultimo caso in cui sono presenti più pasti a distanza > 1 ora e < 3 ore da quello precedente (a destra).

3.2.2 Finestre post-prandiali per il calcolo delle feature

Una volta estratti i periodi post-prandiali a cui sono associati i valori di Y , bisogna definire i periodi su cui calcolare le feature descritte nella prima parte di questo capitolo. Calcolare le diverse metriche sull'intera finestra non sarebbe corretto, poiché per costruzione, questi periodi contengono già l'informazione relativa alla presenza o meno dell'ipoglicemia.

Per quanto riguarda i periodi post-prandiali con $Y=1$, le feature verranno calcolate dall'inizio del periodo (i.e., 30 min prima del pasto) fino all'istante temporale precedente all'ipoglicemia. Per i periodi con $Y=0$, le feature verranno estratte dall'inizio del periodo (i.e., 30 min prima del pasto) fino a k campioni prima del valor minimo di glucosio (Fig. 3.2.2-2). Il valore di k è calcolato come la mediana di tutte le distanze temporali tra l'inizio di ipoglicemia e il nadir nel caso $Y=1$ (Fig. 3.2.2-1).

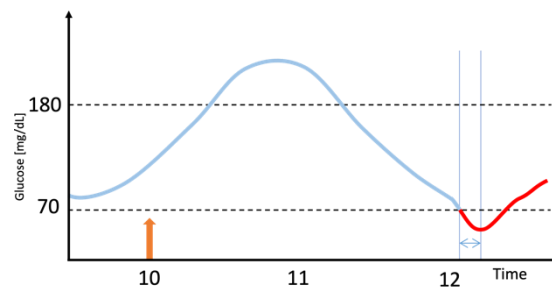


Fig. 3.2.2-1 Esempio di traccia CGM, caso $Y=1$ con distanza tra nadir e inizio dell'evento ipoglicemico in evidenza.

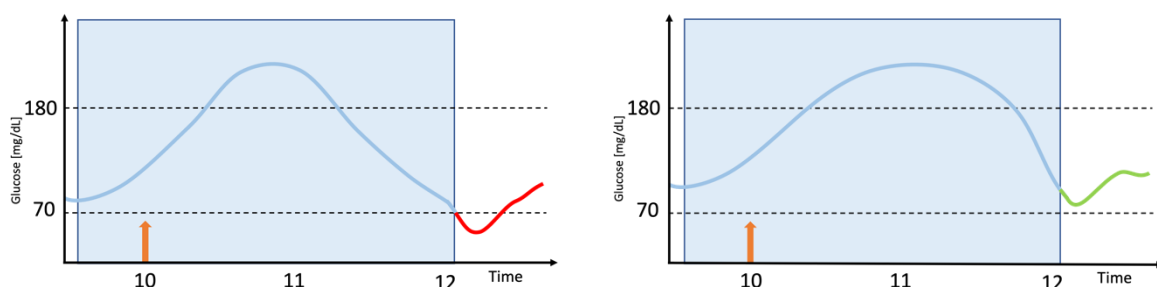


Fig. 3.2.2-2 Esempi di tracce CGM, caso $Y=1$ (a sinistra) e caso $Y=0$ (a destra) con finestra finale che parte da 30 min pre-pasto e finisce all'inizio del nadir.

3.2.3 Preprocessing del dataset

Analizzando i boxplot delle distribuzioni delle variabili e identificando gli outliers, si è deciso di escludere dalle analisi successive le finestre che corrispondono a periodi post-prandiali con dinamiche non “fisiologiche” e che soddisfano uno o più dei seguenti criteri:

- presenza di campioni in ipoglicemia nell'intervallo di 30 minuti precedenti al pasto;
- media glicemica inferiore a 70 mg/dl;
- CV inferiore al 10%;
- > di 7 campioni CGM mancanti ('NaN');
- differenza picco-pasto inferiore a 5 mg/dl.

In questa maniera sono stati esclusi i periodi non idonei all'analisi predittiva, ovvero i periodi che iniziano in ipoglicemia, quelli con un elevato numero di campioni al di sotto della soglia ipoglicemica, i profili piatti o quelli che contenevano molti dati mancanti (Fig. 3.2.3-1).

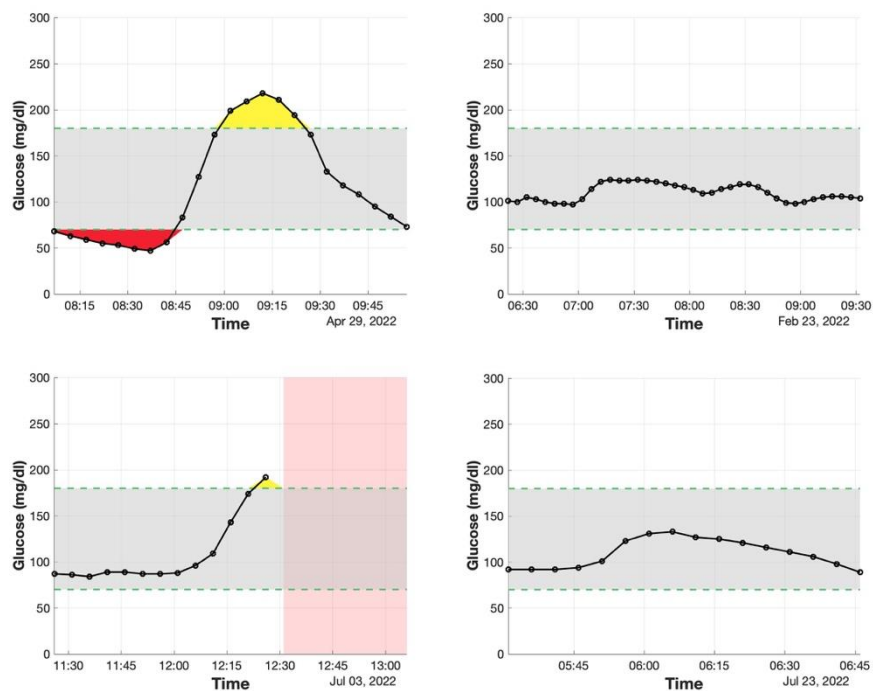


Fig. 3.2.3-1 Esempi di tracce CGM rimosse dal dataset

Il dataset finale è formato da un totale di 5341 periodi CGM, di cui il 27,9% appartiene alla classe 1 e il 72,1% alla classe 0. Per ciascun periodo, sono state estratte 38 diverse feature.

Successivamente, sono state confrontate le distribuzioni delle variabili relative alle due classi utilizzando i boxplot (i valori di mediana e IQR sono riportati di seguito, mentre per i boxplot si rimanda all'appendice). Per la variabile 'mealTime' è stato creato un istogramma, mentre per le

variabili categoriche, sono state riportate le percentuali corrispondenti. In aggiunta, dato che le distribuzioni delle feature per le 2 classi seguono una distribuzione normale, è stato eseguito il T-test, e riportato il valore del p-value associato, per valutare se le due distribuzioni differiscano in modo significativo.

| | Class 1 | Class 0 | p-value (T-test) |
|----------------|-----------------------------|-----------------------------|-------------------------|
| mean | 120.39 (106.02 - 136.18) | 121.28 (109.23 - 136.54) | 4,51 E-04 |
| median | 108 (97 - 123.5) | 113 (102 - 127.5) | 3,27 E-12 |
| std | 34.53 (25.82 - 45.95) | 26.35 (18.4 - 37.17) | 2,35 E-58 |
| CV | 28.82 (23.27 - 34.93) | 21.93 (16.16 - 28.3) | 1,66 E-134 |
| range | 110 (84 - 142.75) | 82 (59 - 113) | 3,11 E-83 |
| diff10 | -27 (-44 - -14) | -10 (-22 - -3) | 3,70 E-174 |
| diff20 | 54 (-82 - -29) | -18 (-41 - -5) | 1,97 E-204 |
| diff30 | -69 (-98 - -37) | -24 (-53 - -6) | 6,04 E-194 |
| MAGE | 89.5 (64.5 - 123.81) | 64.5 (42 - 96.5) | 1,16 E-60 |
| EF | 2 (1 - 2) | 1 (0 - 2) | 1,42 E-85 |
| lastROC | -2.8 (-4.33 - -1.4) | -0.93 (-2.13 - -0.27) | 7,94 E-195 |
| meanROC | -0.05 (-0.16 - 0.06) | 0.03 (-0.09 - 0.18) | 1,92 E-29 |
| J-index | 23.84 (17.89 - 32.77) | 21.85 (16.83 - 30.08) | 3,69 E-05 |
| LBGI | 1.04 (0.58 - 1.77) | 0.44 (0.13 - 0.95) | 1,42 E-154 |
| HBGI | 1.72 (0.64 - 3.67) | 1.27 (0.41 - 3.08) | 2,63 E-04 |
| BGRI | 3.32 (2.29 - 4.88) | 2.13 (1.26 - 3.57) | 2,62 E-40 |
| ADRR | 14.94 (10.32 - 22.27) | 8.67 (5.35 - 14.35) | 2,07 E-122 |
| minDR | -24.36 (-28.98 - -19.4) | -6.15 (-11.43 - -2.66) | 0,00 |
| maxDR | 28.67 (9.92 - 61.41) | 18.81 (4.94 - 47.51) | 1,08 E-09 |
| meanDR | 0.95 (-3.09 - 7.31) | 1.41 (-1.14 - 6.38) | 2,58 E-04 |
| maxDiff | 26 (20 - 35) | 18 (12 - 27) | 1,03 E-107 |

| | | | |
|-----------------------|--------------------------|---------------------------|-----------|
| minDiff | -24 (-30 - -19) | -21 (-27 - -16) | 1,35 E-26 |
| gluMax | 183 (157 - 216) | 171 (146.5 - 204) | 2,10 E-12 |
| last local max | 169 (134 - 205) | 149 (119 - 185) | 2,84 E-21 |
| prominence | 69 (23 - 109) | 39 (9 - 78) | 1,06 E-46 |
| gluPos | 1617 (1225.25 - 2028) | 1455.5 (1075.5 - 1898) | 5,14 E-11 |
| gluNeg | 1617 (1217.25 - 2106) | 1676.5 (1175.5 - 2231) | 1,70 E-01 |
| Meal2Peak Diff | 85 (60 - 116) | 67 (45 - 96) | 2,50 E-33 |
| Meal2Peak Time | 45 (35 - 70) | 40 (30 - 55) | 2,74 E-27 |
| peakAngle | 11.13 (7.61 - 18.27) | 13.25 (8.93 - 20.32) | 6,07 E-02 |
| mealGlu | 93 (84 - 105) | 98 (89 - 110) | 2,74 E-09 |
| lastGlu | 74 (71 - 79) | 98 (88 - 111) | 0,00 |
| TAR130 | 34.38 (21.74 - 45.16) | 33.33 (18.52 - 47.37) | 7,96 E-01 |
| TAR180 | 3.85 (0 - 22.22) | 0 (0 - 17.65) | 1,62 E-07 |

Tab. 3.2.3-2 Valori di mediana, 25° e 75° percentile per ciascuna variabile, distinguendo le due classi; e valore di p-value associato al T-test.

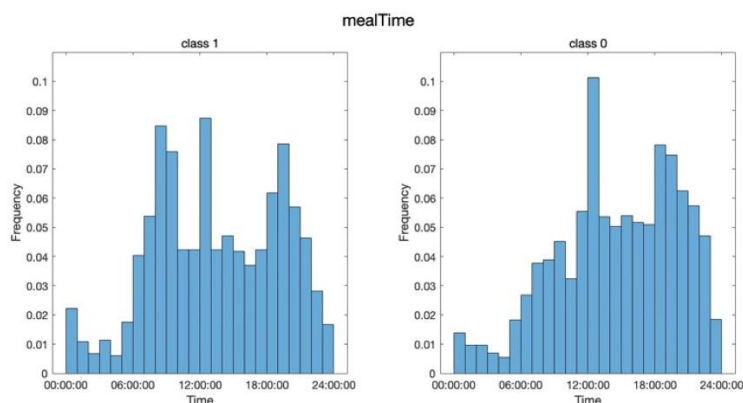


Fig. 3.2.3-3 Istogramma della variabile 'mealTime' per la classe 1 e 0.

| | Classe 1 | Classe 0 |
|-----------------------|-----------------|-----------------|
| % typeMeal = 0 | 18.54% | 20.66% |
| % typeMeal = 1 | 30.83% | 26.66% |

Tab. 3.2.3-4 Percentuale di pasti con carboidrati < 30 g (typeMeal =0) e > 30 g (typeMeal =1), per la classe 1 e 0.

| | Classe 1 | Classe 0 |
|--------------------------|-----------------|-----------------|
| %sintomi presenti | 8.46% | 5.97% |

Tab. 3.2.3-5 Percentuale di sintomi presenti nelle finestre per la classe 1 e 0.

Dall'osservazione dei boxplot sono emerse le variabili che mostrano una maggiore capacità di discriminazione tra le due classi. Queste variabili risultano particolarmente informative e rilevanti per l'identificazione delle differenze tra i due gruppi. Le variabili individuate includono: CV, std, range, diff10, diff20, diff30, MAGE, lastROC, ADRR, minDR, maxDR, maxDiff, prominence, Meal2Peak Diff, lastGlu. Secondo il T-test, le distribuzioni che non presentano differenze significative tra le due classi ($p\text{-value} > 0.05$) riguardano le variabili gluNeg, peakAngle e TAR130.

Capitolo 4

Caratterizzazione delle ipoglicemie

In questo capitolo, si propone un'analisi delle feature estratte nel capitolo precedente con l'obiettivo di caratterizzare gli eventi ipoglicemici. A tal fine, si utilizzeranno modelli logistici a singola variabile per valutare l'impatto di ciascuna variabile sull'insorgenza di tali episodi. Verrà inoltre sviluppato un modello completo che includerà tutte le variabili, in modo da analizzare l'effetto simultaneo delle variabili sull'outcome. Infine, verrà identificato un modello "ridotto" mediante la procedura di selezione automatica delle variabili. In particolare, nella prima parte di questo capitolo viene descritta la regressione logistica, mentre nella seconda parte vengono presentati i risultati principali.

4.1 Regressione logistica

La regressione logistica è una tecnica statistica utilizzata per analizzare la relazione tra una variabile dipendente binaria e una o più variabili indipendenti, di natura continua o categorica. Questa tecnica è comunemente impiegata per scopi di classificazione, il cui obiettivo è prevedere a quale delle due categorie la variabile dipendente appartiene, in base ai valori delle variabili indipendenti.

Prima di applicare la regressione logistica è importante verificare alcune condizioni, tra cui l'indipendenza delle osservazioni, l'assenza di multicollinearità tra le variabili e che si abbia un numero sufficiente di osservazioni.

Il risultato della regressione logistica è espresso tramite una funzione logistica:

$$p = \frac{e^{\beta_0 + \sum_{j=1}^n \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^n \beta_j X_j}}$$

dove p è la probabilità che la variabile dipendente sia uguale a 1, $p(Y=1)$; X_j sono le variabili predittive e β_j rappresentano i coefficienti del modello.

Dalla precedente equazione si ottiene:

$$\frac{p}{1-p} = e^{\beta_0 + \sum_{j=1}^n \beta_j X_j}$$

La quantità $p/(1-p)$ viene chiamata *odds* ed è un numero compreso tra 0 e infinito. Valori di *odds* vicini a 0 e infinito indicano rispettivamente una bassa ed un'alta probabilità di appartenere alla classe 1.

$$odds = e^{\beta_0 + \sum_{j=1}^n \beta_j X_j} = (e^{\beta_0})(e^{\beta_1 X_1}) \dots (e^{\beta_n X_n})$$

L'*odds* risulta uguale a e^{β_0} quando tutte le covariate X_j sono nulle.

e^{β_j} rappresenta l'*odds ratio* relativo alla covariata j-esima (OR_j) e indica l'effetto della variabile X_j sull'*odds*. In particolare, se $OR_j > 1$ (quindi $\beta_j > 0$) la covariata è un fattore di rischio poiché implica un aumento della probabilità; se $OR_j < 1$ (quindi $\beta_j < 0$) la covariata corrisponde a un fattore di protezione dato che causa una diminuzione della probabilità; se $OR_j = 1$ la covariata è ininfluente. Calcolando il logaritmo naturale da entrambe le parti, si ottiene il *log-odds* o *logit*:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{j=1}^n \beta_j X_j + e$$

Con $\beta_0 + \sum_{j=1}^n \beta_j X_j = 0$, la probabilità p sarà uguale a 0.5; con un valore di *logit* elevato la probabilità tenderà a 1, altrimenti con un *logit* basso la probabilità sarà tendente a 0.

In un modello di regressione logistica aumentando di una unità la covariata j-esima X_j il *log odds* cambia della quantità β_j .

I coefficienti β_j sono sconosciuti e vengono stimati utilizzando il metodo della massima verosimiglianza (*Maximum Likelihood*). Si definisce la funzione di verosimiglianza come segue:

$$L(\beta|y) = \prod_{i=1}^m Pr(y_i|X_i, \beta)$$

Le stime ottimali dei coefficienti β_j sono quindi calcolate massimizzando la *likelihood function*:

$$\hat{\beta} = \operatorname{argmax} L(\beta|y)$$

Una volta ottenuti i coefficienti $\hat{\beta}_j$ è possibile calcolare la stima di probabilità:

$$\hat{p} = \frac{e^{\hat{\beta}_0 + \sum_{j=1}^n \hat{\beta}_j X_j}}{1 + e^{\hat{\beta}_0 + \sum_{j=1}^n \hat{\beta}_j X_j}}$$

Per effettuare la classificazione binaria, viene utilizzata una soglia (comunemente 0.5); se la probabilità p è superiore alla soglia allora viene predetta la classe 1, altrimenti la classe 0.

La regressione logistica è uno metodo impiegato non solo per valutare l'influenza di ciascuna variabile indipendente sull'outcome, ma anche per gestire situazioni in cui sono presenti variabili confondenti che influenzano sia le variabili indipendenti che la variabile dipendente. In queste circostanze il modello di regressione logistica viene utilizzato per controllare e tenere in considerazione gli effetti di queste variabili confondenti al fine di evitare conclusioni errate.

In questo lavoro il modello logistico verrà utilizzato in questa fase di caratterizzazione per valutare l'effetto delle variabili sull'insorgenza delle ipoglicemie e anche nella fase successiva di predizione degli eventi ipoglicemici [26].

4.1.2 Analisi e valutazione del modello logistico

Una volta stimato il modello logistico è possibile analizzare la significatività dei coefficienti tramite il *Wald test*. Il suo scopo è determinare se il coefficiente stimato è significativamente diverso da zero, attraverso il calcolo dello *Z-score*:

$$Z_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{SE_{\hat{\beta}_j}} \sim \mathcal{N}(0,1)$$

dove $\hat{\beta}_j$ rappresenta il coefficiente stimato per la variabile j -esima nel modello di regressione e $SE_{\hat{\beta}_j}$ è lo *Standard Error* ovvero la stima della deviazione standard del coefficiente $\hat{\beta}_j$.

Considerando come soglia di significatività $\alpha=0.05$, se lo *Z-score* è inferiore a -1.96 o superiore a 1.96, il coefficiente $\hat{\beta}_j$ risulta significativamente diverso da zero, indicando che la variabile associata è significativa nel modello.

Inoltre, è possibile sfruttare gli *Odds Ratio* per interpretare l'effetto delle feature sull'output, infatti:

- OR = 1: implica che le probabilità dell'evento sono le stesse per entrambe le classi del problema. In altre parole, la variabile predittiva non ha alcun effetto sulle probabilità dell'evento positivo.

- $OR > 1$: suggerisce che all'aumentare della variabile predittiva di una unità, le probabilità dell'evento positivo aumentano. Ciò indica un'associazione positiva o un aumento della probabilità dell'evento positivo.
- $OR < 1$: implica che all'aumentare della variabile predittiva di una unità, le probabilità dell'evento positivo diminuiscono. Ciò indica un'associazione negativa o una diminuzione della probabilità dell'evento positivo.

Per valutare le performance dei modelli logistici, si utilizzano le seguenti metriche:

- *Accuracy*: $(TP+TN)/(TP+TN+FP+FN)$, rappresenta la percentuale di previsioni corrette rispetto a tutte le predizioni effettuate dal modello;
- *Sensitivity*: $TP/(TP+FN)$, noto anche come *Recall* o *True Positive Ratio*, indica la percentuale di veri positivi rispetto all'insieme totale di casi positivi;
- *Specificity*: $TN/(TN+FP)$, indica la percentuale di veri negativi rispetto al totale dei casi negativi;
- *Precision*: $TP/(TP+FP)$, indica la percentuale di veri positivi rispetto al totale delle predizioni positive;

Per visualizzare come variano le performance al variare della soglia di probabilità, si utilizza la *ROC curve*, che fornisce una rappresentazione grafica delle prestazioni del modello mostrando il tasso di veri positivi sull'asse delle ordinate, rispetto al tasso di falsi positivi sulle ascisse al variare della soglia di classificazione.

L'AUC, strettamente correlato alla curva ROC, indica il grado di discriminazione tra le due classi. Maggiore è il valore di AUC, migliore è la capacità del modello di classificare correttamente. Un valore di AUC pari a 0.5 indica che il modello effettua classificazioni in maniera casuale.

4.2 Analisi delle ipoglicemie con modello logistico a singola variabile

Attraverso la creazione di una serie di modelli logistici univariati, ognuno basato su una variabile diversa, è stata condotta un'analisi di correlazione, in modo da identificare le feature maggiormente correlate all'output, cioè al verificarsi o meno dell'ipoglicemia post-prandiale. Per ogni modello vengono riportate le stime dei coefficienti β , gli *Standard Error* (SE), gli *Z-score* ottenuti con il *Wald test* e gli *Odds Ratio* (OR). Inoltre, per ogni modello è stata generata una curva ROC e calcolato il valore dell'AUC per valutare le prestazioni del modello; tali informazioni sono consultabili nell'appendice.

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | -0,3434 | 0,1753 | -1,9589 | 0,7094 |
| mean | -0,0049 | 0,0014 | -3,5012 | 0,9951 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | 0,2164 | 0,1698 | 1,2744 | 1,2416 |
| median | -0,0102 | 0,0015 | -6,9145 | 0,9899 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,9711 | 0,0756 | -26,0568 | 0,1393 |
| std | 0,0310 | 0,0020 | 15,2931 | 1,0315 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -3,2152 | 0,1093 | -29,4163 | 0,0401 |
| CV | 0,0874 | 0,0039 | 22,5497 | 1,0914 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -2,2526 | 0,0811 | -27,7595 | 0,1051 |
| range | 0,0127 | 0,0007 | 18,0072 | 1,0128 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,1975 | 0,0683 | -17,5445 | 0,3019 |
| J-index | 0,0095 | 0,0023 | 4,0928 | 1,0096 |

Tab.4.2-1 Risultati dei modelli univariati per feature di descrizione statistica.

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,8678 | 0,0517 | -36,1480 | 0,1545 |
| diff10 | -0,0437 | 0,0018 | -24,7882 | 0,9572 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -2,0117 | 0,0548 | -36,7290 | 0,1338 |
| diff20 | -0,0267 | 0,0010 | -26,5861 | 0,9737 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -2,0118 | 0,0557 | -36,1487 | 0,1337 |
| diff30 | -0,0214 | 0,0008 | -26,1004 | 0,9788 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,9528 | 0,0533 | -36,6455 | 0,1419 |
| lastROC | -0,4806 | 0,0184 | -26,0640 | 0,6184 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9569 | 0,0310 | -30,8287 | 0,3841 |
| meanROC | -1,1628 | 0,1052 | -11,0530 | 0,3126 |

Tab.4.2-2 Risultati dei modelli univariati per feature short time.

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,8410 | 0,0671 | -27,4473 | 0,1587 |
| MAGE | 0,0106 | 0,0007 | 15,6014 | 1,0107 |

| variable | beta | SE | Z-score | ODDS |
|----------|------|----|---------|------|
|----------|------|----|---------|------|

| | | | | |
|-----------|---------|--------|----------|--------|
| intercept | -1,7206 | 0,0556 | -30,9364 | 0,1790 |
| EF | 0,6213 | 0,0335 | 18,5565 | 1,8614 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -2,3012 | 0,0762 | -30,1919 | 0,1001 |
| maxDiff | 0,0560 | 0,0028 | 20,3629 | 1,0576 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,7694 | 0,0860 | -20,5827 | 0,1704 |
| minDiff | -0,0351 | 0,0034 | -10,4292 | 0,9655 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,7848 | 0,1249 | -14,2882 | 0,1678 |
| gluMax | 0,0045 | 0,0007 | 6,9594 | 1,0045 |

| variable | beta | SE | Z-score | ODDS |
|----------------|---------|--------|----------|--------|
| intercept | -1,8404 | 0,1019 | -18,0595 | 0,1588 |
| last local max | 0,0054 | 0,0006 | 9,3147 | 1,0055 |

| variable | beta | SE | Z-score | ODDS |
|------------|---------|--------|----------|--------|
| intercept | -1,4523 | 0,0494 | -29,3893 | 0,2340 |
| prominence | 0,0083 | 0,0006 | 13,8265 | 1,0084 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,4481 | 0,0834 | -17,3586 | 0,2350 |
| gluPos | 0,0003 | 0,0000 | 6,5126 | 1,0003 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,8523 | 0,0776 | -10,9880 | 0,4264 |
| gluNeg | -0,0001 | 0,0000 | -1,3715 | 0,9999 |

Tab.4.2-3 Risultati dei modelli univariati per feature basate su escursioni glicemiche.

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,8156 | 0,0510 | -35,6038 | 0,1627 |
| LBG1 | 0,9227 | 0,0408 | 22,6226 | 2,5161 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0412 | 0,0399 | -26,0944 | 0,3530 |
| HBGI | 0,0377 | 0,0104 | 3,6146 | 1,0384 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,4348 | 0,0508 | -28,2451 | 0,2382 |
| BGRI | 0,1440 | 0,0116 | 12,3618 | 1,1549 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,9912 | 0,0608 | -32,7560 | 0,1365 |
| ADRR | 0,0755 | 0,0036 | 20,8925 | 1,0784 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -4,8881 | 0,1273 | -38,3896 | 0,0075 |
| minDR | -0,2539 | 0,0069 | -36,7570 | 0,7758 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,1022 | 0,0402 | -27,3855 | 0,3322 |
| maxDR | 0,0040 | 0,0007 | 6,0107 | 1,0040 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9036 | 0,0328 | -27,5319 | 0,4051 |
| meanDR | -0,0131 | 0,0036 | -3,6403 | 0,9869 |

Tab.4.2-4 Risultati dei modelli univariati per feature di rischio.

| variable | beta | SE | Z-score | ODDS |
|----------------|---------|--------|----------|--------|
| intercept | -1,6452 | 0,0685 | -24,0335 | 0,1930 |
| Meal2Peak Diff | 0,0085 | 0,0007 | 11,6755 | 1,0085 |

| variable | beta | SE | Z-score | ODDS |
|----------------|---------|--------|----------|--------|
| intercept | -1,5158 | 0,0631 | -24,0361 | 0,2196 |
| Meal2Peak Time | 0,0110 | 0,0010 | 10,5781 | 1,0111 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,8665 | 0,0488 | -17,7654 | 0,4204 |
| peakAngle | -0,0042 | 0,0022 | -1,8737 | 0,9958 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | -0,1075 | 0,1456 | -0,7384 | 0,8981 |
| mealGlu | -0,0084 | 0,0014 | -5,8794 | 0,9917 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|---------------------|
| intercept | 15,9885 | 0,4862 | 32,8854 | 8,8x10 ⁶ |
| lastGlu | -0,1997 | 0,0060 | -33,3536 | 0,8190 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | -0,4287 | 0,0817 | -5,2444 | 0,6513 |
| mealTime | -0,0372 | 0,0055 | -6,7664 | 0,9634 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0559 | 0,0641 | -16,4753 | 0,3479 |
| numMeals | 0,0608 | 0,0323 | 1,8821 | 1,0627 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0592 | 0,0699 | -15,1632 | 0,3467 |
| typeMeal | 0,2539 | 0,0896 | 2,8325 | 1,2890 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9773 | 0,0318 | -30,7563 | 0,3763 |
| symptoms | 0,3755 | 0,1153 | 3,2571 | 1,4558 |

Tab.4.2-5 Risultati dei modelli univariati per feature locali.

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9632 | 0,0580 | -16,6154 | 0,3817 |
| TAR130 | 0,0004 | 0,0015 | 0,2590 | 1,0004 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0646 | 0,0382 | -27,8981 | 0,3449 |
| TAR180 | 0,0108 | 0,0021 | 5,2081 | 1,0108 |

Tab.4.2-6 Risultati dei modelli univariati per feature temporali.

Da questa analisi emerge che le caratteristiche con maggiore potere predittivo ($AUC > 0.7$) risultano essere: CV, diff10, diff20, diff30, lastROC, LBGI, ADRR, minDR, maxDiff e lastGlu. I corrispondenti OR risultano maggiori di 1 per le variabili CV, LBGI, ADRR, maxDiff e minori di 1 per le variabili diff10, diff20, diff30, lastROC, minDR e lastGlu; indicando rispettivamente che all'aumentare di queste caratteristiche vi è un aumento delle probabilità dell'evento positivo e una diminuzione delle probabilità.

Di conseguenza, le feature che sono state estratte nella parte finale della finestra come diff10, lastGlu ecc., (che sono temporalmente più vicine all'evento di ipoglicemia) risultano essere le più significative. La metrica di rischio LBGI è più significativa di HBGI, poiché LBGI è specificamente progettata per valutare il rischio di ipoglicemia; e ADRR ha dimostrato di essere altrettanto significativa in quanto correlata a LBGI. Anche la variabilità della glicemia e la velocità di cambiamento, rappresentate dal valore di CV e maxDiff, sembrano giocare un ruolo per la caratterizzazione dell'ipoglicemia. Infine, il valore di minDR si è rivelato anch'esso importante: un valore basso indica un rischio inferiore di ipoglicemia.

È interessante notare che metriche come maxGlu e Meal2Peak Diff i cui valori sono legati all'entità del picco e al pasto consumato, non risultano essere significative nel discriminare nella presenza o meno dell'evento ipoglicemico.

4.3 Modello completo

In seguito alla creazione di modelli univariati, si è proceduto a sviluppare un modello completo, inizialmente utilizzando l'intero dataset e poi solo sul training set. È stata rimossa dal dataset la variabile 'typeMeal' dato che conteneva una percentuale di dati mancanti superiore al 50%. Per ogni variabile del modello vengono riportati: la stima del coefficiente β , lo *Standard Error* (SE), l'intervallo di confidenza, lo *Z-score*, l'*Odds Ratio* e l'intervallo di confidenza dell'*Odds Ratio*.

Modello Completo - intero dataset

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int. | |
|-----------|---------|--------|-----------|---------|----------|--------|----------------|--------|
| intercept | -3,9030 | 0,1942 | -4,2835 | -3,5224 | -20,1020 | 0,0202 | 0,0138 | 0,0295 |
| mean | -0,9186 | 1,3484 | -3,5614 | 1,7242 | -0,6813 | 0,3991 | 0,0284 | 5,6082 |
| median | 0,1862 | 0,2795 | -0,3616 | 0,7340 | 0,6663 | 1,2047 | 0,6966 | 2,0835 |

| | | | | | | | | |
|-----------------------|---------|--------|---------|---------|----------|--------|--------|----------|
| std | -1,6983 | 2,1357 | -5,8842 | 2,4876 | -0,7952 | 0,1830 | 0,0028 | 12,0327 |
| CV | 1,0664 | 1,2506 | -1,3848 | 3,5176 | 0,8527 | 2,9049 | 0,2504 | 33,7044 |
| range | -0,6280 | 1,3194 | -3,2140 | 1,9580 | -0,4760 | 0,5336 | 0,0402 | 7,0848 |
| diff10 | -2,6641 | 0,3015 | -3,2551 | -2,0732 | -8,8361 | 0,0697 | 0,0386 | 0,1258 |
| diff20 | 0,7698 | 0,4864 | -0,1836 | 1,7232 | 1,5826 | 2,1594 | 0,8323 | 5,6026 |
| diff30 | -0,1841 | 0,2371 | -0,6489 | 0,2806 | -0,7766 | 0,8318 | 0,5226 | 1,3240 |
| MAGE | -0,4927 | 0,2152 | -0,9146 | -0,0708 | -2,2890 | 0,6110 | 0,4007 | 0,9316 |
| EF | -0,1176 | 0,1357 | -0,3836 | 0,1483 | -0,8669 | 0,8890 | 0,6814 | 1,1599 |
| lastROC | 0,3033 | 0,5285 | -0,7325 | 1,3391 | 0,5739 | 1,3543 | 0,4807 | 3,8155 |
| meanROC | 0,5350 | 0,1792 | 0,1838 | 0,8863 | 2,9853 | 1,7075 | 1,2017 | 2,4262 |
| J-index | 1,9726 | 2,5670 | -3,0587 | 7,0039 | 0,7684 | 7,1893 | 0,0469 | 1100,898 |
| LBDI | -0,2536 | 0,7588 | -1,7409 | 1,2336 | -0,3343 | 0,7760 | 0,1754 | 3,4336 |
| BGRI | 0,3803 | 1,4410 | -2,4440 | 3,2046 | 0,2639 | 1,4627 | 0,0868 | 24,6450 |
| ADRR | 0,9990 | 0,5026 | 0,0138 | 1,9842 | 1,9875 | 2,7156 | 1,0139 | 7,2731 |
| minDR | -2,3359 | 0,2789 | -2,8826 | -1,7892 | -8,3747 | 0,0967 | 0,0560 | 0,1671 |
| maxDR | -0,4052 | 0,6942 | -1,7658 | 0,9555 | -0,5837 | 0,6669 | 0,1710 | 2,5999 |
| meanDR | -0,7786 | 0,9099 | -2,5620 | 1,0048 | -0,8557 | 0,4590 | 0,0771 | 2,7313 |
| maxDiff | -0,0805 | 0,1210 | -0,3176 | 0,1566 | -0,6656 | 0,9226 | 0,7279 | 1,1695 |
| minDiff | 0,0561 | 0,1288 | -0,1963 | 0,3085 | 0,4357 | 1,0577 | 0,8218 | 1,3614 |
| gluMax | 0,6259 | 1,4315 | -2,1799 | 3,4318 | 0,4372 | 1,8700 | 0,1131 | 30,9310 |
| last local max | 0,2293 | 0,1820 | -0,1274 | 0,5861 | 1,2598 | 1,2577 | 0,8803 | 1,7969 |
| prominence | -0,0452 | 0,1397 | -0,3190 | 0,2286 | -0,3234 | 0,9558 | 0,7269 | 1,2568 |
| gluPos | 0,2206 | 0,1538 | -0,0809 | 0,5220 | 1,4341 | 1,2468 | 0,9223 | 1,6854 |
| gluNeg | 0,0402 | 0,1497 | -0,2531 | 0,3336 | 0,2689 | 1,0411 | 0,7764 | 1,3960 |
| Meal2Peak Diff | 0,0011 | 0,6329 | -1,2395 | 1,2417 | 0,0017 | 1,0011 | 0,2895 | 3,4613 |
| Meal2Peak Time | 0,1237 | 0,1277 | -0,1265 | 0,3740 | 0,9692 | 1,1317 | 0,8812 | 1,4535 |
| peakAngle | 0,0143 | 0,1103 | -0,2018 | 0,2305 | 0,1301 | 1,0145 | 0,8172 | 1,2593 |
| mealGlu | 0,1188 | 0,3824 | -0,6307 | 0,8682 | 0,3106 | 1,1261 | 0,5322 | 2,3827 |
| lastGlu | -4,4174 | 0,2981 | -5,0018 | -3,8331 | -14,8164 | 0,0121 | 0,0067 | 0,0216 |
| TAR130 | 0,3266 | 0,3135 | -0,2878 | 0,9410 | 1,0419 | 1,3862 | 0,7499 | 2,5624 |
| TAR180 | -0,1176 | 0,2806 | -0,6676 | 0,4324 | -0,4191 | 0,8891 | 0,5129 | 1,5410 |
| mealTime | -0,1958 | 0,0692 | -0,3314 | -0,0602 | -2,8305 | 0,8222 | 0,7179 | 0,9416 |
| numMeals | -0,1198 | 0,0737 | -0,2642 | 0,0246 | -1,6258 | 0,8871 | 0,7679 | 1,0249 |
| symptoms | 0,0956 | 0,2713 | -0,4361 | 0,6273 | 0,3525 | 1,1003 | 0,6466 | 1,8726 |

Tab. 4.3-1 Output del modello completo.

Modello Completo - training set

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int | |
|------------------|---------|--------|-----------|---------|----------|--------|---------------|---------|
| intercept | -3,9139 | 0,2194 | -4,3440 | -3,4838 | -17,8351 | 0,0200 | 0,0130 | 0,0307 |
| mean | -0,2667 | 1,4954 | -3,1978 | 2,6643 | -0,1784 | 0,7659 | 0,0409 | 14,3577 |
| median | 0,1328 | 0,3113 | -0,4773 | 0,7429 | 0,4266 | 1,1420 | 0,6205 | 2,1021 |
| std | -0,7586 | 2,5093 | -5,6769 | 4,1597 | -0,3023 | 0,4683 | 0,0034 | 64,0519 |
| CV | 0,0835 | 1,4716 | -2,8007 | 2,9678 | 0,0567 | 1,0871 | 0,0608 | 19,4482 |
| range | 0,1605 | 1,5013 | -2,7820 | 3,1029 | 0,1069 | 1,1740 | 0,0619 | 22,2631 |
| diff10 | -2,6686 | 0,3360 | -3,3272 | -2,0101 | -7,9421 | 0,0693 | 0,0359 | 0,1340 |
| diff20 | 0,3867 | 0,5583 | -0,7076 | 1,4811 | 0,6926 | 1,4721 | 0,4928 | 4,3977 |
| diff30 | -0,2428 | 0,2712 | -0,7743 | 0,2888 | -0,8952 | 0,7845 | 0,4610 | 1,3348 |
| MAGE | -0,2955 | 0,2412 | -0,7682 | 0,1773 | -1,2249 | 0,7442 | 0,4638 | 1,1940 |
| EF | -0,1743 | 0,1483 | -0,4651 | 0,1164 | -1,1753 | 0,8400 | 0,6281 | 1,1234 |
| lastROC | 0,7696 | 0,5980 | -0,4024 | 1,9416 | 1,2871 | 2,1590 | 0,6687 | 6,9700 |
| meanROC | 0,6627 | 0,2140 | 0,2433 | 1,0821 | 3,0969 | 1,9400 | 1,2754 | 2,9507 |

| | | | | | | | | |
|-----------------------|---------|--------|---------|---------|----------|--------|--------|----------|
| J-index | 2,2298 | 2,9256 | -3,5045 | 7,9640 | 0,7621 | 9,2976 | 0,0301 | 2875,466 |
| LBGI | 0,5626 | 0,8449 | -1,0933 | 2,2185 | 0,6659 | 1,7552 | 0,3351 | 9,1939 |
| BGRI | -0,9197 | 1,6035 | -4,0626 | 2,2232 | -0,5735 | 0,3986 | 0,0172 | 9,2368 |
| ADRR | 0,8721 | 0,5251 | -0,1572 | 1,9014 | 1,6607 | 2,3920 | 0,8546 | 6,6953 |
| minDR | -2,1587 | 0,3113 | -2,7689 | -1,5485 | -6,9334 | 0,1155 | 0,0627 | 0,2126 |
| maxDR | -0,9744 | 0,7753 | -2,4939 | 0,5451 | -1,2568 | 0,3774 | 0,0826 | 1,7248 |
| meanDR | -0,4465 | 1,0230 | -2,4516 | 1,5585 | -0,4365 | 0,6399 | 0,0862 | 4,7518 |
| maxDiff | -0,0601 | 0,1396 | -0,3337 | 0,2135 | -0,4306 | 0,9417 | 0,7163 | 1,2380 |
| minDiff | 0,0492 | 0,1418 | -0,2288 | 0,3272 | 0,3469 | 1,0504 | 0,7955 | 1,3871 |
| gluMax | 0,3371 | 1,5469 | -2,6948 | 3,3690 | 0,2179 | 1,4009 | 0,0676 | 29,0486 |
| last local max | 0,2501 | 0,2104 | -0,1622 | 0,6624 | 1,1889 | 1,2842 | 0,8503 | 1,9395 |
| prominence | -0,0658 | 0,1636 | -0,3865 | 0,2549 | -0,4023 | 0,9363 | 0,6794 | 1,2903 |
| gluPos | 0,2619 | 0,1753 | -0,0817 | 0,6055 | 1,4939 | 1,2994 | 0,9215 | 1,8321 |
| gluNeg | 0,0228 | 0,1655 | -0,3015 | 0,3471 | 0,1378 | 1,0231 | 0,7397 | 1,4150 |
| Meal2Peak Diff | 0,3019 | 0,7057 | -1,0812 | 1,6850 | 0,4278 | 1,3524 | 0,3392 | 5,3924 |
| Meal2Peak Time | 0,0353 | 0,1441 | -0,2472 | 0,3178 | 0,2451 | 1,0359 | 0,7810 | 1,3741 |
| peakAngle | 0,0668 | 0,1262 | -0,1806 | 0,3142 | 0,5294 | 1,0691 | 0,8348 | 1,3692 |
| mealGlu | 0,2446 | 0,4278 | -0,5939 | 1,0830 | 0,5717 | 1,2771 | 0,5522 | 2,9535 |
| lastGlu | -4,4854 | 0,3396 | -5,1511 | -3,8197 | -13,2066 | 0,0113 | 0,0058 | 0,0219 |
| TAR130 | 0,1213 | 0,3414 | -0,5479 | 0,7905 | 0,3553 | 1,1290 | 0,5782 | 2,2046 |
| TAR180 | -0,1284 | 0,3201 | -0,7558 | 0,4990 | -0,4013 | 0,8795 | 0,4696 | 1,6470 |
| mealTime | -0,2169 | 0,0770 | -0,3678 | -0,0659 | -2,8160 | 0,8050 | 0,6922 | 0,9362 |
| numMeals | -0,1086 | 0,0832 | -0,2718 | 0,0545 | -1,3055 | 0,8970 | 0,7620 | 1,0560 |
| symptoms | -0,1402 | 0,3056 | -0,7392 | 0,4588 | -0,4588 | 0,8692 | 0,4775 | 1,5821 |

Tab. 4.3-1 Output del modello completo (training set).

Dall'analisi dei valori di *Z-score* del modello costruito sull'intero dataset, emergono come variabili significative per la caratterizzazione delle ipoglicemie: diff10, MAGE, meanROC, ADRR, minDR, lastGlu e mealTime. Osservando i risultati sul training set, le variabili più significative risultano essere: diff10, meanROC, minDR, lastGlu e mealTime. Questo sottolinea ulteriormente l'importanza di diff10, ADRR, minDR e lastGlu, che erano emerse come caratteristiche chiave anche nell'analisi a singola variabile. Inoltre, anche le variabili come MAGE, meanROC e mealTime hanno acquisito significatività. In particolare, MAGE e mealTime, aventi $OR < 1$, diminuiscono la probabilità che si verifichi l'evento; al contrario, meanROC, con $OR > 1$, è un fattore di rischio.

A differenza del modello univariato che valuta l'effetto delle caratteristiche in maniera indipendente, l'analisi multivariata considera l'interazione e l'influenza combinata di più variabili sull'outcome, il che giustifica l'identificazione di variabili diverse dalla precedente analisi.

4.4 Modello ridotto

Per ridurre il modello iniziale, è stata utilizzata la funzione *stepwiseglm*. Questa tecnica utilizza la *mixed selection*, ovvero sia la *forward* che la *backward selection* per la selezione delle variabili e

quindi la determinazione del modello finale. Ad ogni passaggio, la funzione decide se aggiungere o rimuovere una variabile dal modello, in base al criterio di selezione scelto. Di default la funzione *stepwiseglm* utilizza come criterio la 'deviance': se il *p-value* dell'*F-statistic* o il *chi-squared statistic* è inferiore dello 0.05, allora viene aggiunta la variabile al modello. Oltre al criterio di selezione delle variabili è stato specificato il tipo di modello di partenza, la distribuzione della variabile risposta e la *link function*. In particolare, per il modello di partenza si è scelta l'opzione 'linear', ovvero il modello all'inizio contiene tutte le variabili. La distribuzione della variabile risposta è stata impostata come 'binomial', poiché la variabile Y è binaria ed è stata utilizzata la *link function* 'logit' dato che si vuole effettuare una regressione di tipo logistico.

Modello ridotto – intero dataset

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int | |
|----------------|---------|--------|-----------|---------|----------|--------|---------------|--------|
| (Intercept) | -3,8881 | 0,1712 | -4,2236 | -3,5526 | -22,7161 | 0,0205 | 0,0146 | 0,0286 |
| diff10 | -2,3732 | 0,1721 | -2,7106 | -2,0359 | -13,7886 | 0,0932 | 0,0665 | 0,1306 |
| diff20 | 0,7816 | 0,1552 | 0,4775 | 1,0857 | 5,0371 | 2,1850 | 1,6120 | 2,9616 |
| MAGE | -0,6640 | 0,1571 | -0,9719 | -0,3561 | -4,2270 | 0,5148 | 0,3784 | 0,7004 |
| meanROC | 0,6584 | 0,1180 | 0,4271 | 0,8896 | 5,5803 | 1,9317 | 1,5329 | 2,4342 |
| ADRR | 0,7981 | 0,1740 | 0,4570 | 1,1391 | 4,5864 | 2,2212 | 1,5793 | 3,1240 |
| minDR | -2,3814 | 0,1743 | -2,7230 | -2,0397 | -13,6612 | 0,0924 | 0,0657 | 0,1301 |
| Meal2Peak Time | 0,2205 | 0,0724 | 0,0787 | 0,3623 | 3,0476 | 1,2467 | 1,0819 | 1,4367 |
| lastGlu | -4,5681 | 0,2683 | -5,0940 | -4,0422 | -17,0247 | 0,0104 | 0,0061 | 0,0176 |
| TAR130 | 0,5936 | 0,1117 | 0,3746 | 0,8126 | 5,3126 | 1,8105 | 1,4544 | 2,2539 |
| mealTime | -0,1973 | 0,0648 | -0,3244 | -0,0702 | -3,0418 | 0,8210 | 0,7230 | 0,9323 |
| numMeals | -0,1234 | 0,0701 | -0,2608 | 0,0140 | -1,7608 | 0,8839 | 0,7704 | 1,0141 |

Tab. 4.4-1 Output del modello ridotto.

Modello ridotto – training set

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int | |
|-------------|---------|--------|-----------|---------|----------|--------|---------------|--------|
| (Intercept) | -3,8938 | 0,1929 | -4,2718 | -3,5157 | -20,1889 | 0,0204 | 0,0140 | 0,0297 |
| diff10 | -2,5034 | 0,2744 | -3,0413 | -1,9656 | -9,1227 | 0,0818 | 0,0478 | 0,1401 |
| MAGE | -0,5908 | 0,1699 | -0,9238 | -0,2578 | -3,4777 | 0,5539 | 0,3970 | 0,7727 |
| lastROC | 0,8761 | 0,2601 | 0,3663 | 1,3859 | 3,3683 | 2,4015 | 1,4424 | 3,9983 |
| meanROC | 0,6921 | 0,1389 | 0,4199 | 0,9643 | 4,9834 | 1,9979 | 1,5218 | 2,6229 |
| ADRR | 0,6331 | 0,1895 | 0,2617 | 1,0044 | 3,3413 | 1,8834 | 1,2991 | 2,7303 |
| minDR | -2,2911 | 0,1924 | -2,6681 | -1,9140 | -11,9096 | 0,1012 | 0,0694 | 0,1475 |
| gluPos | 0,2460 | 0,1050 | 0,0401 | 0,4518 | 2,3420 | 1,2789 | 1,0409 | 1,5712 |
| lastGlu | -4,6249 | 0,3075 | -5,2276 | -4,0222 | -15,0413 | 0,0098 | 0,0054 | 0,0179 |
| TAR130 | 0,4959 | 0,1292 | 0,2425 | 0,7492 | 3,8367 | 1,6419 | 1,2745 | 2,1153 |
| mealTime | -0,2241 | 0,0713 | -0,3639 | -0,0843 | -3,1415 | 0,7992 | 0,6950 | 0,9192 |

Tab. 4.4-2 Output del modello ridotto (training set).

Applicando la selezione automatica delle variabili tramite la funzione *stepwiseglm* al dataset intero si è ottenuto il primo modello ridotto, contenente 11 variabili. Oltre alle variabili diff10, diff20, MAGE, meanROC, ADRR, minDR, lastGlu, mealTime, già risultate significative nell'analisi di

correlazione univariata, sono state selezionate nel modello anche le variabili Meal2Peak Time, TAR130 e numMeals, anche se quest'ultima non è statisticamente significativa. Il secondo modello ridotto, ottenuto utilizzando il training set, comprende 10 variabili. In questo modello, entrano in gioco la variabile lastGlu e gluPos, mentre le variabili diff20, ADRR, Meal2Peak Time e numMeals non sono incluse rispetto al modello precedente.

Vengono riportate di seguito le performance sull'intero dataset e sul test set del modello completo e del modello ridotto.

Performance sull'intero dataset

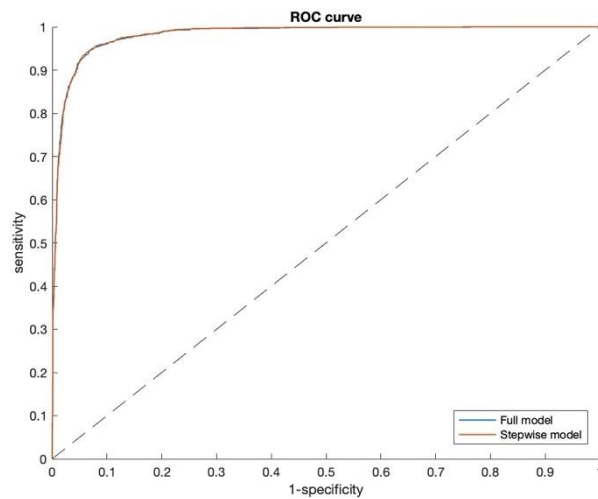


Fig. 4.4-3 ROC curve del modello completo e ridotto applicati all'intero dataset.

| | Full model | Stepwise model |
|--------------------|-------------------|-----------------------|
| AUC | 0,9817 | 0,9819 |
| Accuracy | 0,9423 | 0,9433 |
| Sensitivity | 0,9240 | 0,9157 |
| Specificity | 0,9495 | 0,9539 |
| Precision | 0,8773 | 0,8847 |

Tab. 4.4-4 Risultati di performance sull'intero dataset.

I risultati di performance per entrambi i modelli sono molto buoni: ciascun modello ha una buona capacità di discriminazione tra le due classi, prevedendo correttamente la maggior parte dei casi. La *precision* leggermente inferiore rispetto alle altre metriche indica che il modello può classificare erroneamente come positive alcune osservazioni che in realtà sono negative, ovvero porta a falsi positivi.

Performance sul test set

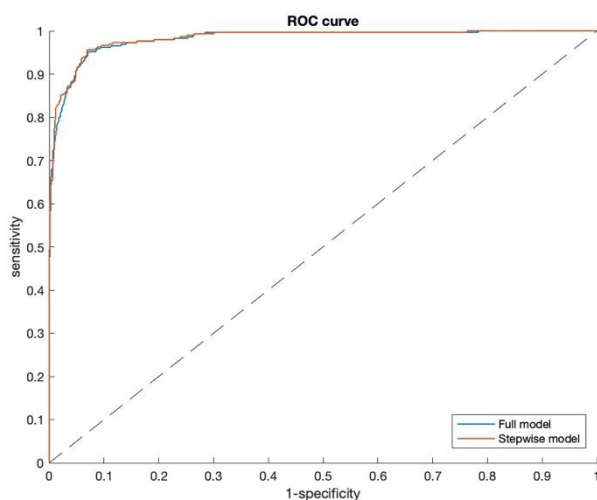


Fig. 4.4-5 ROC curve del modello completo e ridotto applicati al test set.

| | Full model | Stepwise model |
|--------------------|-------------------|-----------------------|
| AUC | 0,9815 | 0,9828 |
| Accuracy | 0,9397 | 0,9436 |
| Sensitivity | 0,8690 | 0,8519 |
| Specificity | 0,9675 | 0,9791 |
| Precision | 0,9130 | 0,9405 |

Tab. 4.4-6 Risultati di performance sul test set.

Anche le performance valutate sul test set risultano essere veramente buone. A differenza dei risultati precedenti, la *sensitivity* risulta inferiore e la *precision* più alta; ciò indica che il modello genera meno falsi positivi, ma allo stesso tempo non classifica correttamente alcuni casi positivi.

In conclusione, i risultati sono soddisfacenti dato che viene classificata correttamente la quasi totalità degli eventi ipoglicemici (*sensitivity* e *precision* > 0,85). In particolare, le feature più importanti risultano essere quelle che descrivono il cambiamento del trend del segnale glicemico (CV, diff10, diff20, diff30, lastROC, meanROC, lastGlu, MAGE, maxDiff, gluPos), quelle che descrivono il rischio (LBGI, ADRR, minDR e anche TAR130) e quelle che riguardano il pasto (mealTime, Meal2Peak Time). Non risultano invece avere potere predittivo le variabili legate al picco glicemico prandiale come maxGlu e Meal2Peak Diff.

Capitolo 5

Predizione di ipoglicemie post-prandiali

In questo capitolo, verrà presentata l'analisi del modello predittivo delle ipoglicemie post-beriatriche: nella prima parte, si presenterà lo sviluppo del framework delle ipoglicemie e l'estrazione delle feature presentate nel capitolo 3, insieme a nuove feature locali; nelle fasi successive, verranno presentati i risultati ottenuti con il modello completo e con il modello 'ridotto', in cui le feature sono state selezionate tramite *stepwise selection*, costruiti sia sull'intero dataset che sul training set. Un'ultima analisi riguarderà le performance dei modelli sviluppati, bilanciando le classi nel dataset.

5.1 Framework di predizione futura delle ipoglicemie

Dall'analisi di caratterizzazione delle ipoglicemie descritta nel capitolo precedente emerge che le feature più significative sono legate agli ultimi 20/30 minuti precedenti all'insorgenza dell'evento di ipoglicemia. In altre parole, riguardano la parte del segnale CGM caratterizzata da una discesa ripida del segnale glicemico. Tuttavia, queste informazioni non possono essere impiegate nello sviluppo di un algoritmo di predizione in real-time, poiché la loro misurazione richiederebbe l'acquisizione del segnale CGM fino a pochi istanti prima all'ipoglicemia. Di conseguenza, il guadagno temporale, ovvero il periodo in cui un paziente potrebbe intervenire per evitare l'ipoglicemia, sarebbe minimo, se non addirittura inesistente. Questo limiterebbe notevolmente l'interesse pratico dell'applicazione dell'algoritmo predittivo.

Per ovviare a tale problema, a partire dai periodi post-prandiali selezionati precedentemente, viene definita una nuova finestra temporale. Essa inizia 30 minuti prima del pasto, ma termina 20 minuti dopo il punto di massimo che corrisponde al picco iperglicemico, come mostrato in *Fig. 5.1-1*.

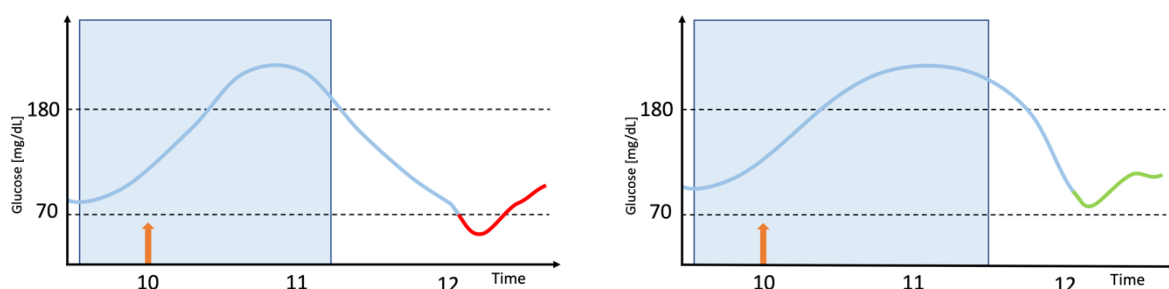


Fig. 5.1-1 Esempi di tracce CGM, con finestra da cui vengono estratte le feature per la predizione.

Oltre alle feature già presentate nel capitolo 3, ne sono state introdotte altre specifiche per il problema di predizione che si sta affrontando. In particolare, come mostrato in Fig. 5.1-2, ci si focalizza su 3 regioni del periodo post-prandiale:

- **regione pre-pasto:** da 30 min prima dell'assunzione pasto al momento del pasto;
- **regione pasto-picco glicemico:** dall'istante del pasto all'istante del picco;
- **regione 20 minuti post-picco:** regione dall'istante del picco a 20 minuti dopo.

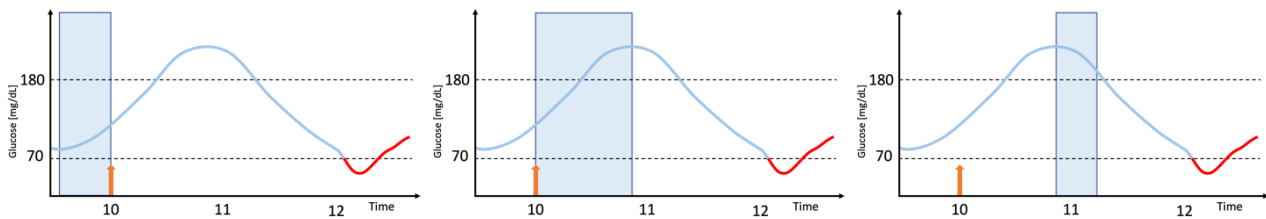


Fig. 5.1-2 esempio di tracce CGM con in evidenza i periodi da cui vengono estratte le feature locali: periodo da 30 min pre-pasto al pasto (a sinistra), dal pasto al picco (al centro), dal picco a 20 min dopo (a destra).

Da questi intervalli temporali vengono estratte le variabili locali: **mean**, **CV**, **lastROC**, **meanROC**, **maxDiff** e **minDiff**.

Come nel capitolo 3, per valutare l'impatto di ciascuna variabile, sono stati identificati dei modelli univariati in modo tale da includere poi nella costruzione dei modelli di predizione solo le variabili maggiormente significative. Vengono riportate di seguito le stime dei coefficienti β , gli SE, gli Z-score e gli OR per ciascun modello univariato, nelle tre diverse regioni di interesse del periodo post-prandiale. Le curve ROC con i valori di AUC sono disponibili nell'appendice.

Regione pre-pasto

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | 0,1576 | 0,1661 | 0,9489 | 1,1707 |
| mean | -0,0109 | 0,0016 | -6,7385 | 0,9892 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9900 | 0,0443 | -22,3725 | 0,3716 |
| CV | 0,0051 | 0,0046 | 1,1113 | 1,0051 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9569 | 0,0308 | -31,0293 | 0,3841 |
| lastROC | -0,0316 | 0,0258 | -1,2210 | 0,9689 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9490 | 0,0311 | -30,4850 | 0,3871 |
| meanROC | 0,0510 | 0,0305 | 1,6729 | 1,0523 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9605 | 0,0398 | -24,1418 | 0,3827 |
| maxDiff | 0,0009 | 0,0043 | 0,2050 | 1,0009 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -0,9983 | 0,0379 | -26,3284 | 0,3685 |
| minDiff | -0,0083 | 0,0042 | -1,9603 | 0,9918 |

Fig. 5.1-3 Risultati dei modelli univariati nella regione pre-pasto: stime dei coefficienti β , SE, Z-score e OR.

Dall'osservazione delle ROC curve e dai valori di Z-score l'unica feature della regione pre-pasto che sembra avere potere predittivo è la media.

Regione pasto-picco

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | -0,9810 | 0,1349 | -7,2741 | 0,3749 |
| mean | 0,0002 | 0,0010 | 0,2327 | 1,0002 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -2,2870 | 0,0928 | -24,6365 | 0,1016 |
| CV | 0,0580 | 0,0037 | 15,8355 | 1,0597 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,1869 | 0,0634 | -18,7330 | 0,3052 |
| lastROC | 0,1161 | 0,0256 | 4,5323 | 1,1231 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0758 | 0,0617 | -17,4491 | 0,3410 |
| meanROC | 0,0644 | 0,0250 | 2,5777 | 1,0665 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0148 | 0,0323 | -31,4586 | 0,3625 |
| maxDiff | 0,0306 | 0,0042 | 7,2394 | 1,0311 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,7390 | 0,0840 | -20,6920 | 0,1757 |
| minDiff | -0,0345 | 0,0033 | -10,3071 | 0,9661 |

Fig. 5.1-4 Risultati dei modelli univariati nella regione pasto-picco: stime dei coefficienti β , SE, Z-score e OR.

Nella regione pasto-picco l'unica feature con lo Z-score non significativo risulta essere la media. Tuttavia, osservando le curve ROC, le feature maggiormente significative risultano essere: CV e minDiff.

Regione 20 minuti post-picco

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|---------|--------|
| intercept | -1,1989 | 0,1277 | -9,3920 | 0,3015 |
| mean | 0,0020 | 0,0008 | 2,7221 | 1,0021 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -2,0137 | 0,0776 | -25,9528 | 0,1335 |
| CV | 0,0876 | 0,0052 | 16,9421 | 1,0916 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,7681 | 0,0681 | -25,9725 | 0,1707 |
| lastROC | -0,2984 | 0,0190 | -15,6960 | 0,7420 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,7400 | 0,0696 | -24,9991 | 0,1755 |
| meanROC | -0,3310 | 0,0225 | -14,7210 | 0,7182 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,7017 | 0,0672 | -25,3288 | 0,1824 |
| maxDiff | 0,0396 | 0,0027 | 14,7105 | 1,0404 |

| variable | beta | SE | Z-score | ODDS |
|-----------|---------|--------|----------|--------|
| intercept | -1,0828 | 0,0462 | -23,4369 | 0,3386 |
| minDiff | 0,0475 | 0,0071 | 6,7145 | 1,0486 |

Fig. 5.1-5 Risultati dei modelli univariati nella regione 20 minuti post-picco: stime dei coefficienti β , SE, Z-score e OR.

Nell'ultima regione, dal picco a 20 minuti post-picco, nonostante tutte le feature abbiano uno Z-score significativo, le feature con AUC più elevato sono CV, lastROC, meanROC e maxDiff.

Da questa analisi le feature più significative ($AUC \cong 0.6$) che vengono incluse nei modelli successivi sono: **preprandial mean, Meal2Peak CV, afterPeak CV, afterPeak meanROC, afterPeak maxDiff e Meal2Peak minDiff.**

Nota Implementativa: In seguito all'estrazione delle feature, è stato calcolato il *Prediction Horizon* (PH) per i vari periodi CGM e sono stati rimossi dal dataset tutti i periodi con $PH \leq 0$, poiché per effettuare una predizione è necessario che l'orizzonte di predizione sia maggiore di zero (Fig. 5.1-

6). Di conseguenza, il dataset, originariamente costituito da 5341 periodi, si è ridotto a 5304 tracce con 44 feature.

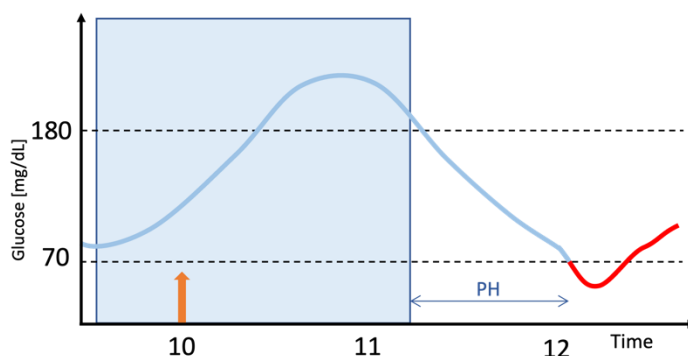


Fig. 5.1-6 Esempio di calcolo del Prediction Horizon.

5.2 Analisi modello completo e del modello con selezione delle variabili applicati all'intero dataset

In questa sezione si analizza il modello completo, con tutte le feature, che viene identificato e testato su tutto il dataset. Questa analisi consente di identificare il ruolo delle variabili nella predizione dell'ipoglicemia. Successivamente, come per l'analisi di caratterizzazione, si sviluppa e analizza il modello ridotto ottenuto attraverso l'uso della funzione *stepwiseglm*, la quale esegue una selezione delle variabili in base alla loro significatività.

Note implementative: prima di procedere con il training del modello, sono state eliminate dal dataset alcune feature. In particolare, è stata esclusa **HBGI** poiché era già stata esclusa nella precedente analisi, **typeMeal** a causa del numero elevato di 'NaN' e **symptoms** in quanto non è risultata particolarmente significativa.

5.2.1 Modello Completo

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int. | |
|------------------|---------|--------|-----------|---------|----------|--------|----------------|--------|
| intercept | -1,2288 | 0,0703 | -1,3666 | -1,0910 | -17,4748 | 0,2926 | 0,2550 | 0,3359 |
| mean | -1,0016 | 0,7404 | -2,4529 | 0,4496 | -1,3528 | 0,3673 | 0,0860 | 1,5677 |
| preprandial mean | 0,6633 | 0,1976 | 0,2760 | 1,0506 | 3,3567 | 1,9412 | 1,3178 | 2,8593 |
| median | 0,1062 | 0,1678 | -0,2226 | 0,4351 | 0,6333 | 1,1121 | 0,8004 | 1,5451 |
| std | 0,6402 | 0,8355 | -0,9975 | 2,2778 | 0,7662 | 1,8968 | 0,3688 | 9,7554 |
| CV | -0,2511 | 0,5316 | -1,2930 | 0,7908 | -0,4723 | 0,7779 | 0,2744 | 2,2052 |
| Meal2Peak CV | 0,0076 | 0,1652 | -0,3162 | 0,3314 | 0,0458 | 1,0076 | 0,7289 | 1,3929 |
| afterPeak CV | 1,3200 | 0,2038 | 0,9207 | 1,7194 | 6,4785 | 3,7436 | 2,5110 | 5,5813 |

| | | | | | | | | |
|------------------------------|---------|--------|---------|---------|---------|---------|--------|----------|
| range | 0,9518 | 0,4129 | 0,1426 | 1,7610 | 2,3054 | 2,5903 | 1,1533 | 5,8182 |
| diff10 | -0,6620 | 0,2230 | -1,0990 | -0,2249 | -2,9688 | 0,5158 | 0,3332 | 0,7986 |
| diff20 | -2,6858 | 1,0186 | -4,6823 | -0,6894 | -2,6368 | 0,0682 | 0,0093 | 0,5019 |
| diff30 | -0,0134 | 0,1428 | -0,2933 | 0,2665 | -0,0939 | 0,9867 | 0,7458 | 1,3054 |
| MAGE | -0,1637 | 0,0990 | -0,3578 | 0,0303 | -1,6539 | 0,8490 | 0,6992 | 1,0308 |
| EF | 0,0014 | 0,0671 | -0,1301 | 0,1329 | 0,0205 | 1,0014 | 0,8780 | 1,1421 |
| lastROC | -0,4120 | 0,4461 | -1,2863 | 0,4623 | -0,9237 | 0,6623 | 0,2763 | 1,5877 |
| meanROC | 0,2583 | 0,1696 | -0,0740 | 0,5907 | 1,5234 | 1,2948 | 0,9286 | 1,8053 |
| afterPeak meanROC | 1,5192 | 0,3604 | 0,8128 | 2,2255 | 4,2155 | 4,5685 | 2,2543 | 9,2585 |
| J-index | -1,2189 | 0,9789 | -3,1376 | 0,6998 | -1,2452 | 0,2956 | 0,0434 | 2,0133 |
| LBGI | -0,2744 | 0,2688 | -0,8013 | 0,2526 | -1,0206 | 0,7600 | 0,4487 | 1,2873 |
| BGRI | 1,2089 | 0,6990 | -0,1610 | 2,5788 | 1,7296 | 3,3498 | 0,8513 | 13,1819 |
| ADRR | 0,6242 | 0,3748 | -0,1103 | 1,3588 | 1,6657 | 1,8668 | 0,8956 | 3,8915 |
| minDR | -0,3373 | 0,1101 | -0,5531 | -0,1215 | -3,0636 | 0,7137 | 0,5752 | 0,8856 |
| maxDR | -0,4532 | 0,4038 | -1,2446 | 0,3383 | -1,1223 | 0,6356 | 0,2881 | 1,4025 |
| meanDR | -0,2789 | 0,4818 | -1,2232 | 0,6653 | -0,5790 | 0,7566 | 0,2943 | 1,9452 |
| maxDiff | -0,3456 | 0,2029 | -0,7432 | 0,0520 | -1,7038 | 0,7078 | 0,4756 | 1,0534 |
| afterPeak maxDiff | 0,0418 | 0,2271 | -0,4033 | 0,4869 | 0,1839 | 1,0426 | 0,6681 | 1,6272 |
| minDiff | 0,3910 | 0,1814 | 0,0354 | 0,7466 | 2,1552 | 1,4785 | 1,0360 | 2,1098 |
| Meal2Peak minDiff | -0,2649 | 0,1783 | -0,6144 | 0,0847 | -1,4852 | 0,7673 | 0,5410 | 1,0884 |
| gluMax | -1,5755 | 0,6858 | -2,9197 | -0,2314 | -2,2974 | 0,2069 | 0,0540 | 0,7934 |
| last local max | -2,5967 | 1,3989 | -5,3386 | 0,1451 | -1,8563 | 0,0745 | 0,0048 | 1,1562 |
| prominence | -0,0266 | 0,0711 | -0,1658 | 0,1127 | -0,3738 | 0,9738 | 0,8472 | 1,1193 |
| gluPos | -0,0452 | 0,1294 | -0,2989 | 0,2086 | -0,3488 | 0,9559 | 0,7417 | 1,2319 |
| gluNeg | -0,1456 | 0,0947 | -0,3313 | 0,0401 | -1,5367 | 0,8645 | 0,7180 | 1,0409 |
| Meal2Peak Time | 0,1693 | 0,1400 | -0,1052 | 0,4438 | 1,2089 | 1,1845 | 0,9002 | 1,5586 |
| peakAngle | 0,2342 | 0,0656 | 0,1056 | 0,3628 | 3,5701 | 1,2639 | 1,1114 | 1,4374 |
| mealGlu | -0,1986 | 0,1540 | -0,5005 | 0,1034 | -1,2889 | 0,8199 | 0,6062 | 1,1089 |
| lastGlu | 3,3513 | 1,1687 | 1,0607 | 5,6420 | 2,8676 | 28,5405 | 2,8883 | 282,0176 |
| TAR130 | 0,3807 | 0,1497 | 0,0874 | 0,6741 | 2,5441 | 1,4634 | 1,0913 | 1,9622 |
| TAR180 | -0,0452 | 0,1361 | -0,3120 | 0,2216 | -0,3322 | 0,9558 | 0,7320 | 1,2481 |
| mealTime | -0,2209 | 0,0369 | -0,2933 | -0,1485 | -5,9808 | 0,8018 | 0,7458 | 0,8620 |
| numMeals | -0,1699 | 0,0430 | -0,2542 | -0,0856 | -3,9509 | 0,8437 | 0,7755 | 0,9179 |

Tab. 5.2.1-1 Output del modello completo.

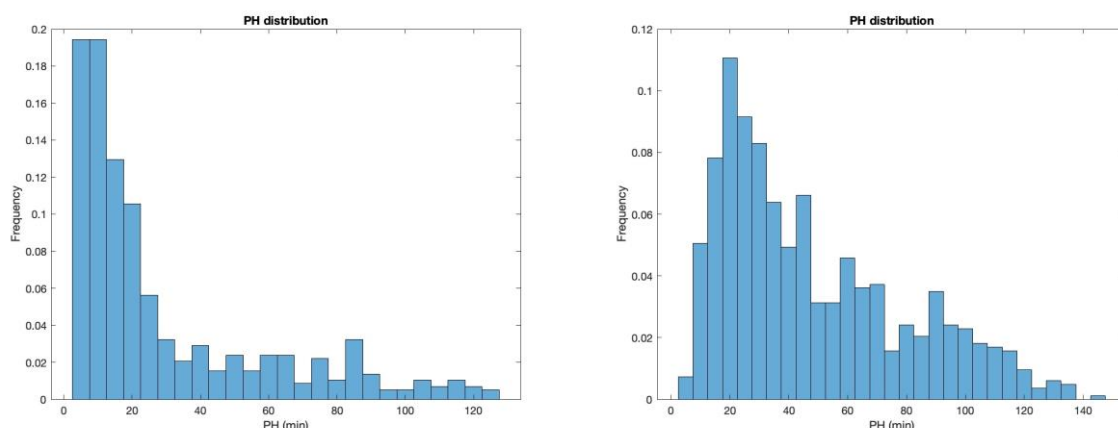


Fig. 5.2.1-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra)

Osservando i risultati del modello completo, dai valori di *Z-score* emergono le feature più significative ovvero: preprandial mean, afterPeak CV, range, diff10, diff20, afterPeak meanROC, minDR, minDiff, gluMax, peakAngle, lastGlu, TAR 130, mealTime e numMeals.

Analizzando gli *Odds Ratio*, si sono identificate le feature che hanno effetto di aumentare la probabilità che si verifichi l'evento ipoglicemico, tra cui: preprandial mean, afterPeak CV, range, afterPeak meanROC, minDiff, peakAngle, lastGlu e TAR130. Quindi, un aumento di queste variabili è associato ad una maggiore probabilità di occorrenza dell'evento. Al contrario, variabili come diff10, diff20, minDR, gluMax, mealTime e numMeals indicano una minore probabilità.

Dall'osservazione dei *prediction horizon*, si evince che il modello riesce a classificare correttamente la maggior parte delle ipoglicemie che si verificano nel breve termine (PH < 20 minuti), tuttavia, a partire dai 20 minuti in avanti, il modello riscontra maggiori difficoltà nella predizione degli eventi (Fig. 5.2.1-2).

5.2.2 Modello ridotto

| variable | beta | SE | Conf.Int. | Z-score | Odds | Odds Conf.Int. | | |
|--------------------------|---------|--------|-----------|---------|----------|----------------|--------|--------|
| (Intercept) | -1,1982 | 0,0620 | -1,3197 | -1,0767 | -19,3219 | 0,3017 | 0,2672 | 0,3407 |
| preprandial mean | 0,4751 | 0,1353 | 0,2100 | 0,7403 | 3,5126 | 1,6082 | 1,2337 | 2,0965 |
| afterPeak CV | 1,3449 | 0,1710 | 1,0097 | 1,6802 | 7,8637 | 3,8379 | 2,7448 | 5,3664 |
| range | 1,4139 | 0,2613 | 0,9017 | 1,9261 | 5,4107 | 4,1119 | 2,4638 | 6,8625 |
| diff10 | -0,8145 | 0,1400 | -1,0888 | -0,5401 | -5,8181 | 0,4429 | 0,3366 | 0,5827 |
| diff20 | -2,6028 | 0,9146 | -4,3955 | -0,8101 | -2,8457 | 0,0741 | 0,0123 | 0,4448 |
| MAGE | -0,1357 | 0,0775 | -0,2875 | 0,0162 | -1,7515 | 0,8731 | 0,7502 | 1,0163 |
| meanROC | 0,2533 | 0,1258 | 0,0066 | 0,4999 | 2,0125 | 1,2882 | 1,0066 | 1,6486 |
| afterPeak meanROC | 1,3513 | 0,2942 | 0,7747 | 1,9280 | 4,5930 | 3,8625 | 2,1699 | 6,8755 |
| J-index | -1,3740 | 0,2695 | -1,9023 | -0,8457 | -5,0978 | 0,2531 | 0,1492 | 0,4292 |

| | | | | | | | | |
|-----------------------|---------|--------|---------|---------|---------|---------|--------|----------|
| BGRI | 0,7532 | 0,1678 | 0,4244 | 1,0820 | 4,4897 | 2,1237 | 1,5286 | 2,9505 |
| minDR | -0,4102 | 0,0894 | -0,5854 | -0,2350 | -4,5899 | 0,6635 | 0,5569 | 0,7905 |
| maxDiff | -0,3045 | 0,0888 | -0,4785 | -0,1304 | -3,4283 | 0,7375 | 0,6197 | 0,8777 |
| minDiff | 0,1701 | 0,0623 | 0,0480 | 0,2923 | 2,7295 | 1,1855 | 1,0491 | 1,3395 |
| gluMax | -1,4295 | 0,4847 | -2,3795 | -0,4795 | -2,9493 | 0,2394 | 0,0926 | 0,6191 |
| last local max | -2,7293 | 1,3221 | -5,3205 | -0,1381 | -2,0644 | 0,0653 | 0,0049 | 0,8710 |
| peakAngle | 0,2546 | 0,0491 | 0,1583 | 0,3508 | 5,1852 | 1,2899 | 1,1716 | 1,4202 |
| mealGlu | -0,2764 | 0,0960 | -0,4645 | -0,0883 | -2,8807 | 0,7585 | 0,6285 | 0,9154 |
| lastGlu | 3,1847 | 1,0644 | 1,0984 | 5,2709 | 2,9919 | 24,1591 | 2,9994 | 194,5932 |
| TAR130 | 0,2668 | 0,0725 | 0,1246 | 0,4089 | 3,6791 | 1,3057 | 1,1328 | 1,5051 |
| mealTime | -0,2267 | 0,0358 | -0,2968 | -0,1565 | -6,3345 | 0,7972 | 0,7432 | 0,8551 |
| numMeals | -0,1648 | 0,0404 | -0,2439 | -0,0857 | -4,0817 | 0,8481 | 0,7836 | 0,9179 |

Tab. 5.2.2-1 Output del modello ridotto.

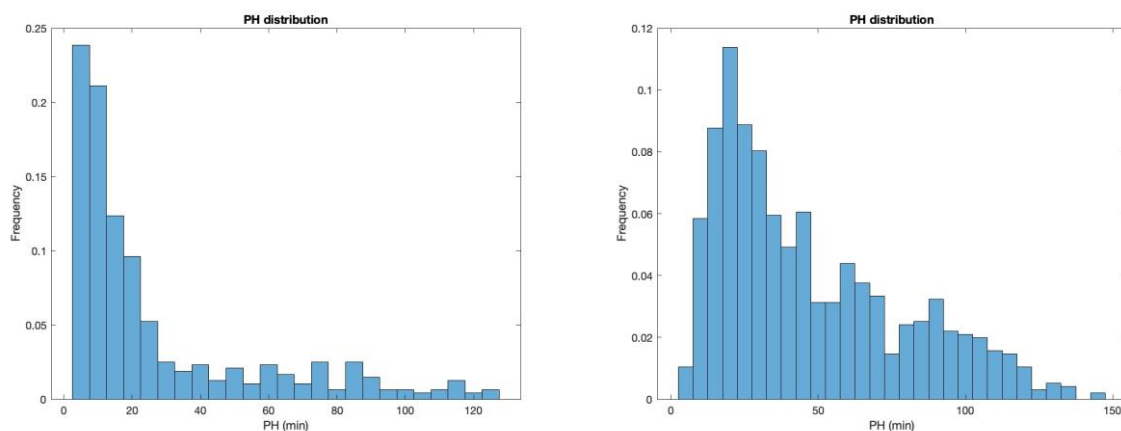


Fig. 5.2.2-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Il modello ridotto ottenuto include 21 variabili: preprandial mean, afterPeak CV, range, diff10, diff20, MAGE, meanROC, afterPeak meanROC, J-index, BGRI, minDR, maxDiff, minDiff, gluMax, last local max, peakAngle, mealGlu, lastGlu, TAR130, mealTime e numMeals.

A differenza del modello con tutte le variabili, il modello ridotto ha selezionato come variabili significative anche MAGE, meanROC, J-index, BGRI, maxDiff, last local max e mealGlu che, a parte meanROC e BGRI, hanno potere di diminuire la probabilità che si verifichi l'evento.

Vengono riportate di seguito le performance dei modelli sull'intero dataset e sul training set.

5.2.3 Performance sull'intero dataset

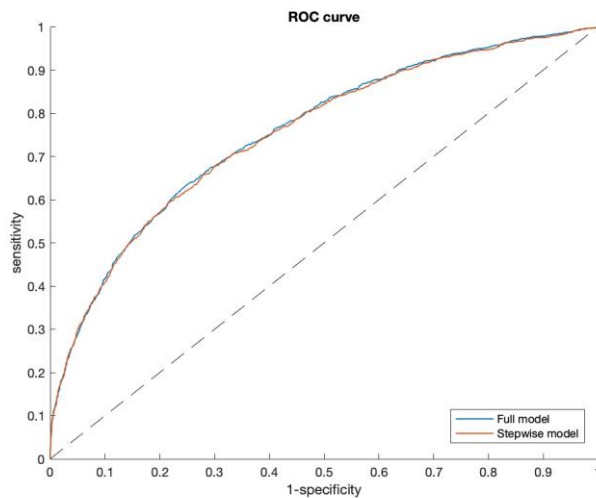


Fig. 5.2.3-1 ROC curve del modello completo e ridotto applicati all'intero dataset.

| | Full model | Stepwise model |
|---|-------------------|-----------------------|
| AUC | 0,7635 | 0,7533 |
| Accuracy | 0,7565 | 0,7588 |
| Sensitivity | 0,4140 | 0,3329 |
| Specificity | 0,9032 | 0,9375 |
| Precision | 0,6415 | 0,6908 |
| TG: median | 15 | 15 |
| (25th - 75th percentile) | (10 - 40) | (10 - 35) |

Tab. 5.2.3-2 Risultati di performance sull'intero dataset.

I risultati per entrambi i modelli sono abbastanza simili: valori di AUC e *accuracy* intorno al 75% indicano una buona capacità del modello di discriminare le due classi, riuscendo a prevedere correttamente circa 3/4 delle osservazioni. Tuttavia, la *sensitivity* bassa indica che il modello ha difficoltà ad individuare tutti i casi positivi, classificandoli erroneamente come negativi a causa dello sbilanciamento tra le due classi. Dai valori di *precision* si evince che il modello causa comunque dei falsi positivi. D'altro canto, la *specificity* elevata suggerisce che il modello riesce a classificare correttamente la maggior parte dei casi negativi.

5.3 Analisi del modello completo e del modello con selezione delle variabili applicati a training e test set

In questa sezione, si analizza il modello completo, con tutte le feature, identificato sul training set e successivamente testato sul test set. L'efficacia dell'algoritmo viene quindi valutata mediante l'analisi della sua capacità predittiva. Come in precedenza, viene presentato il modello completo e, in seguito, il modello ridotto attraverso l'uso della funzione *stepwiseglm*, la quale esegue una selezione delle variabili in base alla loro significatività.

5.3.1 Modello completo

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int. | |
|--------------------------|---------|--------|-----------|---------|----------|--------|----------------|---------|
| intercept | -1,2483 | 0,0769 | -1,3990 | -1,0976 | -16,2317 | 0,2870 | 0,2468 | 0,3337 |
| mean | -1,4583 | 0,8373 | -3,0995 | 0,1829 | -1,7415 | 0,2326 | 0,0451 | 1,2007 |
| preprandial mean | 0,8784 | 0,2259 | 0,4357 | 1,3211 | 3,8893 | 2,4070 | 1,5461 | 3,7474 |
| median | 0,0900 | 0,1923 | -0,2868 | 0,4669 | 0,4683 | 1,0942 | 0,7506 | 1,5950 |
| std | 0,6392 | 0,9111 | -1,1464 | 2,4249 | 0,7016 | 1,8950 | 0,3178 | 11,3010 |
| CV | -0,2173 | 0,5821 | -1,3582 | 0,9236 | -0,3732 | 0,8047 | 0,2571 | 2,5185 |
| Meal2Peak CV | -0,0310 | 0,1833 | -0,3903 | 0,3282 | -0,1693 | 0,9694 | 0,6769 | 1,3885 |
| afterPeak CV | 1,1405 | 0,2254 | 0,6987 | 1,5824 | 5,0594 | 3,1285 | 2,0111 | 4,8666 |
| range | 1,1073 | 0,4523 | 0,2207 | 1,9938 | 2,4480 | 3,0261 | 1,2470 | 7,3434 |
| diff10 | -0,6401 | 0,2545 | -1,1390 | -0,1413 | -2,5152 | 0,5272 | 0,3201 | 0,8682 |
| diff20 | -3,5554 | 1,0954 | -5,7023 | -1,4084 | -3,2458 | 0,0286 | 0,0033 | 0,2445 |
| diff30 | 0,0349 | 0,1583 | -0,2753 | 0,3450 | 0,2203 | 1,0355 | 0,7593 | 1,4120 |
| MAGE | -0,1856 | 0,1081 | -0,3975 | 0,0263 | -1,7168 | 0,8306 | 0,6720 | 1,0266 |
| EF | -0,0376 | 0,0746 | -0,1839 | 0,1086 | -0,5040 | 0,9631 | 0,8321 | 1,1148 |
| lastROC | -0,1699 | 0,4963 | -1,1425 | 0,8028 | -0,3423 | 0,8438 | 0,3190 | 2,2318 |
| meanROC | 0,2876 | 0,1900 | -0,0848 | 0,6600 | 1,5139 | 1,3332 | 0,9187 | 1,9347 |
| afterPeak meanROC | 1,2173 | 0,4014 | 0,4306 | 2,0039 | 3,0329 | 3,3779 | 1,5382 | 7,4180 |
| J-index | -0,8400 | 1,1045 | -3,0049 | 1,3249 | -0,7605 | 0,4317 | 0,0495 | 3,7619 |
| LBGI | -0,3497 | 0,2976 | -0,9330 | 0,2335 | -1,1754 | 0,7049 | 0,3934 | 1,2630 |
| BGRI | 1,4518 | 0,7940 | -0,1044 | 3,0079 | 1,8285 | 4,2706 | 0,9009 | 20,2450 |
| ADRR | 0,6015 | 0,4090 | -0,2002 | 1,4031 | 1,4706 | 1,8248 | 0,8186 | 4,0680 |
| minDR | -0,2566 | 0,1206 | -0,4930 | -0,0202 | -2,1277 | 0,7737 | 0,6108 | 0,9800 |
| maxDR | -0,6533 | 0,4466 | -1,5287 | 0,2221 | -1,4628 | 0,5203 | 0,2168 | 1,2487 |
| meanDR | -0,5038 | 0,5500 | -1,5819 | 0,5743 | -0,9159 | 0,6042 | 0,2056 | 1,7759 |
| maxDiff | -0,2817 | 0,2245 | -0,7217 | 0,1583 | -1,2548 | 0,7545 | 0,4859 | 1,1715 |
| afterPeak maxDiff | -0,0418 | 0,2516 | -0,5349 | 0,4514 | -0,1661 | 0,9591 | 0,5857 | 1,5704 |
| minDiff | 0,2746 | 0,1932 | -0,1041 | 0,6533 | 1,4213 | 1,3160 | 0,9011 | 1,9219 |
| Meal2Peak minDiff | -0,1888 | 0,1898 | -0,5609 | 0,1833 | -0,9946 | 0,8279 | 0,5707 | 1,2012 |
| gluMax | -2,1615 | 0,7584 | -3,6479 | -0,6751 | -2,8503 | 0,1151 | 0,0260 | 0,5091 |
| last local max | -3,1463 | 1,4931 | -6,0727 | -0,2199 | -2,1073 | 0,0430 | 0,0023 | 0,8026 |
| prominence | -0,0150 | 0,0798 | -0,1713 | 0,1414 | -0,1876 | 0,9852 | 0,8426 | 1,1518 |

| | | | | | | | | |
|-----------------------|---------|--------|---------|---------|---------|---------|--------|----------|
| gluPos | -0,0618 | 0,1437 | -0,3435 | 0,2199 | -0,4299 | 0,9401 | 0,7093 | 1,2460 |
| gluNeg | -0,1453 | 0,1051 | -0,3512 | 0,0606 | -1,3831 | 0,8648 | 0,7038 | 1,0625 |
| Meal2Peak Time | 0,2391 | 0,1552 | -0,0651 | 0,5434 | 1,5404 | 1,2701 | 0,9369 | 1,7218 |
| peakAngle | 0,2137 | 0,0748 | 0,0671 | 0,3602 | 2,8583 | 1,2382 | 1,0695 | 1,4336 |
| mealGlu | -0,2593 | 0,1755 | -0,6032 | 0,0847 | -1,4775 | 0,7716 | 0,5471 | 1,0884 |
| lastGlu | 4,0591 | 1,2535 | 1,6023 | 6,5159 | 3,2383 | 57,9237 | 4,9645 | 675,8236 |
| TAR130 | 0,3754 | 0,1685 | 0,0451 | 0,7057 | 2,2277 | 1,4556 | 1,0461 | 2,0253 |
| TAR180 | 0,0559 | 0,1537 | -0,2454 | 0,3572 | 0,3635 | 1,0575 | 0,7824 | 1,4293 |
| mealTime | -0,2060 | 0,0413 | -0,2870 | -0,1250 | -4,9860 | 0,8138 | 0,7505 | 0,8825 |
| numMeals | -0,1998 | 0,0483 | -0,2945 | -0,1051 | -4,1351 | 0,8189 | 0,7449 | 0,9002 |

Tab. 5.3.1-1 Output del modello completo (training set).

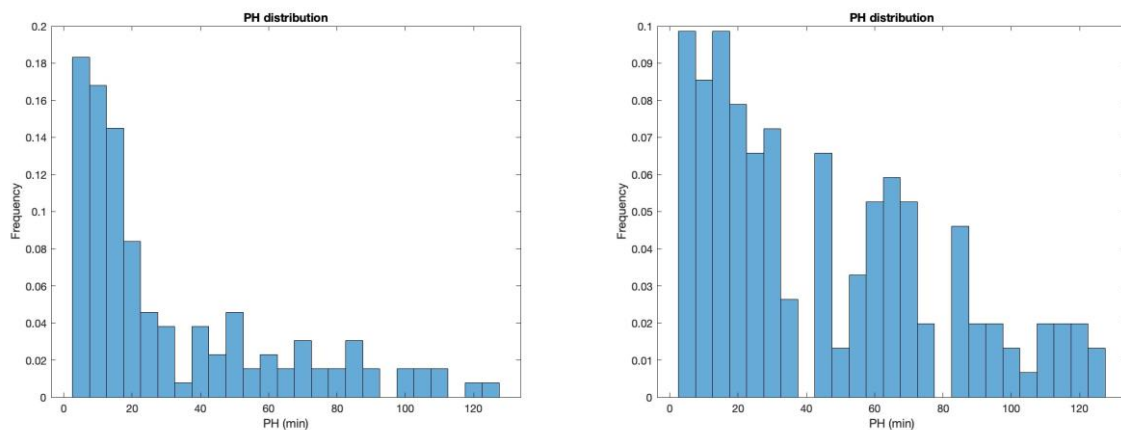


Fig. 5.3.1-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Il modello costruito solamente sul training set, a differenza del primo modello completo, considera in più come feature significativa anche last local max, ma esclude minDiff.

Dagli istogrammi di distribuzione del *prediction horizon*, si nota un cambiamento rispetto all'analisi sull'intero dataset: se in precedenza il modello tendeva a classificare correttamente gli eventi più prossimi rispetto a quelli temporalmente più distanti, ora si osserva che il PH ha un impatto minore sulle prestazioni del modello.

5.3.2 Modello ridotto

| variable | beta | SE | Conf.Int. | Z-score | Odds | Odds Conf.Int. | | |
|-------------------------|---------|--------|-----------|---------|----------|----------------|--------|--------|
| (Intercept) | -1,2399 | 0,0679 | -1,3730 | -1,1069 | -18,2613 | 0,2894 | 0,2533 | 0,3306 |
| mean | -1,1361 | 0,2786 | -1,6822 | -0,5900 | -4,0774 | 0,3211 | 0,1860 | 0,5543 |
| preprandial mean | 0,5482 | 0,1160 | 0,3210 | 0,7755 | 4,7280 | 1,7302 | 1,3785 | 2,1717 |
| afterPeak CV | 1,0998 | 0,1972 | 0,7132 | 1,4864 | 5,5758 | 3,0036 | 2,0405 | 4,4212 |
| range | 1,1367 | 0,2967 | 0,5552 | 1,7181 | 3,8316 | 3,1164 | 1,7423 | 5,5742 |
| diff10 | -0,7935 | 0,1518 | -1,0910 | -0,4960 | -5,2272 | 0,4523 | 0,3359 | 0,6090 |
| diff20 | -3,4981 | 0,9783 | -5,4155 | -1,5806 | -3,5757 | 0,0303 | 0,0044 | 0,2059 |

| | | | | | | | | |
|------------------------------|---------|--------|---------|---------|---------|---------|--------|----------|
| MAGE | -0,2453 | 0,0881 | -0,4180 | -0,0726 | -2,7838 | 0,7825 | 0,6583 | 0,9300 |
| afterPeak meanROC | 1,2034 | 0,3288 | 0,5589 | 1,8478 | 3,6600 | 3,3313 | 1,7488 | 6,3456 |
| BGRI | 0,5067 | 0,1859 | 0,1424 | 0,8711 | 2,7257 | 1,6599 | 1,1530 | 2,3896 |
| ADRR | 0,5918 | 0,3399 | -0,0744 | 1,2580 | 1,7411 | 1,8072 | 0,9283 | 3,5182 |
| minDR | -0,2759 | 0,1105 | -0,4925 | -0,0593 | -2,4968 | 0,7589 | 0,6111 | 0,9424 |
| maxDR | -0,8486 | 0,2732 | -1,3841 | -0,3132 | -3,1062 | 0,4280 | 0,2505 | 0,7311 |
| maxDiff | -0,3237 | 0,0991 | -0,5179 | -0,1295 | -3,2674 | 0,7234 | 0,5957 | 0,8785 |
| gluMax | -2,0118 | 0,5834 | -3,1552 | -0,8683 | -3,4483 | 0,1338 | 0,0426 | 0,4197 |
| last local max | -3,0125 | 1,4003 | -5,7570 | -0,2680 | -2,1514 | 0,0492 | 0,0032 | 0,7649 |
| peakAngle | 0,2498 | 0,0534 | 0,1452 | 0,3545 | 4,6809 | 1,2838 | 1,1563 | 1,4254 |
| mealGlu | -0,1385 | 0,0800 | -0,2954 | 0,0183 | -1,7312 | 0,8706 | 0,7443 | 1,0185 |
| lastGlu | 4,1100 | 1,1285 | 1,8981 | 6,3219 | 3,6420 | 60,9485 | 6,6735 | 556,6354 |
| TAR130 | 0,3086 | 0,1156 | 0,0819 | 0,5352 | 2,6686 | 1,3615 | 1,0854 | 1,7077 |
| mealTime | -0,2083 | 0,0403 | -0,2873 | -0,1293 | -5,1699 | 0,8120 | 0,7503 | 0,8787 |
| numMeals | -0,2011 | 0,0455 | -0,2903 | -0,1120 | -4,4214 | 0,8178 | 0,7481 | 0,8941 |

Tab. 5.3.2-1 Output del modello ridotto (training set).

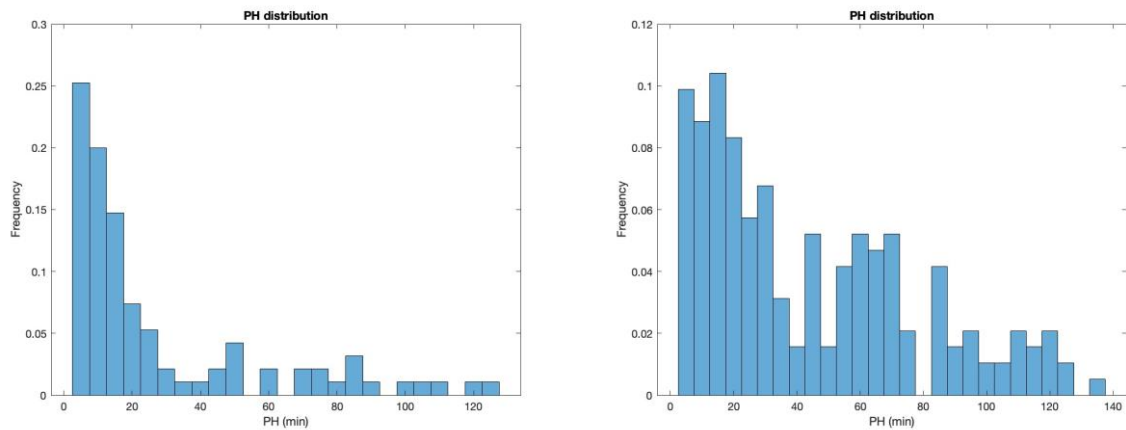


Fig. 5.3.2-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Il modello *Stepwise* comprende le variabili risultate significative nel modello completo e considera in più le variabili mean, MAGE, BGRI, ADRR, maxDR, maxDiff e mealGlu.

Vengono riportate di seguito le performance dei modelli sull'intero dataset e sul training set.

5.3.3 Performance sul test set

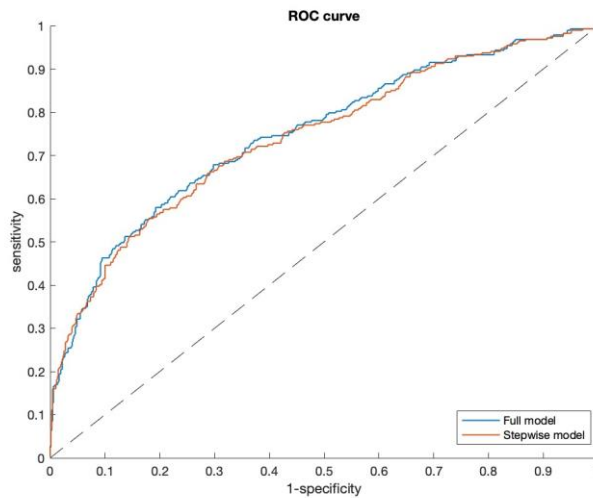


Fig. 5.3.3-1 ROC curve del modello completo e ridotto applicati al test set.

| | Full model | Stepwise model |
|---|-------------------|-----------------------|
| AUC | 0,7487 | 0,7415 |
| Accuracy | 0,7745 | 0,7671 |
| Sensitivity | 0,4629 | 0,3310 |
| Specificity | 0,9052 | 0,9514 |
| Precision | 0,6718 | 0,7422 |
| TG: median (25 th - 75 th percentile) | 20 (10 - 50) | 15 (6,25 - 33,75) |

Tab. 5.3.3-2 Risultati di performance sul test set.

Anche i risultati di performance sul test set portano alle stesse considerazioni di prima: entrambi i modelli hanno una buona accuratezza, riuscendo a prevedere in maniera corretta la quasi totalità dei casi negativi. Tuttavia, i modelli hanno la tendenza di causare falsi negativi, e nonostante non riescano a predire correttamente la maggior parte dei casi positivi, causano comunque alcuni falsi positivi. La bassa *specificity* e l'alta *sensitivity* possono dipendere dal fatto che il modello è sbilanciato, con un numero significativamente maggiore di casi appartenenti alla classe 0 rispetto a quelli della classe 1.

5.4 Bilanciamento delle classi: analisi modello completo e del modello con selezione delle variabili applicati all' intero dataset

Siccome all'interno del dataset, le due classi 1 e 0 sono sbilanciate e lo sbilanciamento può portare a distorsioni durante il processo di training e di valutazione del modello, per bilanciare le due classi è stato rimosso un numero di campioni della classe 0, in modo da ottenere lo stesso numero di campioni per entrambe le classi. Così facendo, il dataset si è ridotto a 2904 campioni.

Vengono riportati i risultati del modello completo e ridotto sull'intero dataset e sul training set.

5.4.1 Modello Completo

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int. | |
|-------------------|---------|--------|-----------|---------|---------|--------|----------------|---------|
| intercept | 0,0832 | 0,0704 | -0,0549 | 0,2212 | 1,1807 | 1,0867 | 0,9466 | 1,2476 |
| mean | -0,3592 | 0,9190 | -2,1605 | 1,4421 | -0,3909 | 0,6982 | 0,1153 | 4,2295 |
| preprandial mean | 0,5548 | 0,2615 | 0,0421 | 1,0674 | 2,1211 | 1,7415 | 1,0430 | 2,9078 |
| median | 0,1586 | 0,2129 | -0,2587 | 0,5758 | 0,7450 | 1,1718 | 0,7721 | 1,7786 |
| std | 0,2076 | 1,0172 | -1,7860 | 2,2013 | 0,2041 | 1,2307 | 0,1676 | 9,0365 |
| CV | 0,0558 | 0,6416 | -1,2019 | 1,3134 | 0,0869 | 1,0573 | 0,3006 | 3,7188 |
| Meal2Peak CV | 0,1756 | 0,2191 | -0,2539 | 0,6051 | 0,8014 | 1,1920 | 0,7758 | 1,8315 |
| afterPeak CV | 1,4188 | 0,2816 | 0,8669 | 1,9707 | 5,0386 | 4,1321 | 2,3795 | 7,1755 |
| range | 1,3252 | 0,5211 | 0,3038 | 2,3466 | 2,5430 | 3,7630 | 1,3550 | 10,4502 |
| diff10 | -0,6732 | 0,2731 | -1,2085 | -0,1379 | -2,4651 | 0,5101 | 0,2987 | 0,8711 |
| diff20 | -1,4715 | 1,2605 | -3,9422 | 0,9991 | -1,1674 | 0,2296 | 0,0194 | 2,7159 |
| diff30 | 0,0729 | 0,1798 | -0,2794 | 0,4252 | 0,4055 | 1,0756 | 0,7562 | 1,5299 |
| MAGE | -0,0871 | 0,1250 | -0,3321 | 0,1579 | -0,6967 | 0,9166 | 0,7174 | 1,1711 |
| EF | -0,0720 | 0,0888 | -0,2462 | 0,1021 | -0,8109 | 0,9305 | 0,7818 | 1,1075 |
| lastROC | -0,4683 | 0,5472 | -1,5408 | 0,6043 | -0,8557 | 0,6261 | 0,2142 | 1,8300 |
| meanROC | 0,1153 | 0,2222 | -0,3203 | 0,5509 | 0,5189 | 1,1222 | 0,7260 | 1,7348 |
| afterPeak meanROC | 1,5357 | 0,4677 | 0,6191 | 2,4524 | 3,2836 | 4,6448 | 1,8572 | 11,6167 |
| J-index | -1,4266 | 1,2206 | -3,8190 | 0,9658 | -1,1688 | 0,2401 | 0,0219 | 2,6268 |
| LBGI | -0,1613 | 0,3732 | -0,8927 | 0,5701 | -0,4322 | 0,8510 | 0,4095 | 1,7685 |
| BGRI | 0,9266 | 0,8964 | -0,8304 | 2,6835 | 1,0336 | 2,5258 | 0,4359 | 14,6369 |
| ADRR | 0,5015 | 0,4365 | -0,3540 | 1,3571 | 1,1490 | 1,6513 | 0,7018 | 3,8850 |
| minDR | -0,3836 | 0,1732 | -0,7230 | -0,0442 | -2,2150 | 0,6814 | 0,4853 | 0,9568 |
| maxDR | 0,3135 | 0,5346 | -0,7343 | 1,3613 | 0,5864 | 1,3682 | 0,4798 | 3,9013 |
| meanDR | -0,4162 | 0,6441 | -1,6787 | 0,8462 | -0,6462 | 0,6595 | 0,1866 | 2,3308 |
| maxDiff | -0,5556 | 0,2670 | -1,0790 | -0,0323 | -2,0808 | 0,5737 | 0,3399 | 0,9683 |
| afterPeak maxDiff | 0,2848 | 0,2983 | -0,2998 | 0,8695 | 0,9548 | 1,3295 | 0,7409 | 2,3857 |
| minDiff | 0,3119 | 0,2379 | -0,1544 | 0,7783 | 1,3110 | 1,3661 | 0,8569 | 2,1778 |
| Meal2Peak minDiff | -0,1567 | 0,2341 | -0,6156 | 0,3022 | -0,6693 | 0,8550 | 0,5403 | 1,3528 |
| gluMax | -1,6154 | 0,8890 | -3,3578 | 0,1270 | -1,8172 | 0,1988 | 0,0348 | 1,1354 |
| last local max | -1,6924 | 1,7097 | -5,0435 | 1,6587 | -0,9898 | 0,1841 | 0,0065 | 5,2526 |
| prominence | -0,0206 | 0,0900 | -0,1971 | 0,1559 | -0,2290 | 0,9796 | 0,8211 | 1,1687 |

| | | | | | | | | |
|-----------------------|---------|--------|---------|---------|---------|--------|--------|----------|
| gluPos | 0,1244 | 0,1598 | -0,1888 | 0,4376 | 0,7784 | 1,1324 | 0,8279 | 1,5489 |
| gluNeg | -0,0524 | 0,1193 | -0,2863 | 0,1814 | -0,4395 | 0,9489 | 0,7511 | 1,1989 |
| Meal2Peak Time | -0,0720 | 0,1824 | -0,4294 | 0,2854 | -0,3948 | 0,9305 | 0,6509 | 1,3303 |
| peakAngle | 0,3450 | 0,0889 | 0,1707 | 0,5193 | 3,8793 | 1,4120 | 1,1861 | 1,6808 |
| mealGlu | -0,0992 | 0,2033 | -0,4977 | 0,2993 | -0,4879 | 0,9056 | 0,6079 | 1,3489 |
| lastGlu | 2,0173 | 1,4995 | -0,9216 | 4,9563 | 1,3454 | 7,5183 | 0,3979 | 142,0647 |
| TAR130 | 0,4038 | 0,1743 | 0,0622 | 0,7454 | 2,3166 | 1,4975 | 1,0641 | 2,1073 |
| TAR180 | 0,0254 | 0,1684 | -0,3046 | 0,3554 | 0,1509 | 1,0257 | 0,7374 | 1,4267 |
| mealTime | -0,2292 | 0,0466 | -0,3206 | -0,1379 | -4,9195 | 0,7951 | 0,7257 | 0,8712 |
| numMeals | -0,1679 | 0,0560 | -0,2777 | -0,0581 | -2,9972 | 0,8454 | 0,7575 | 0,9435 |

Tab. 5.4.1-1 Output del modello completo.

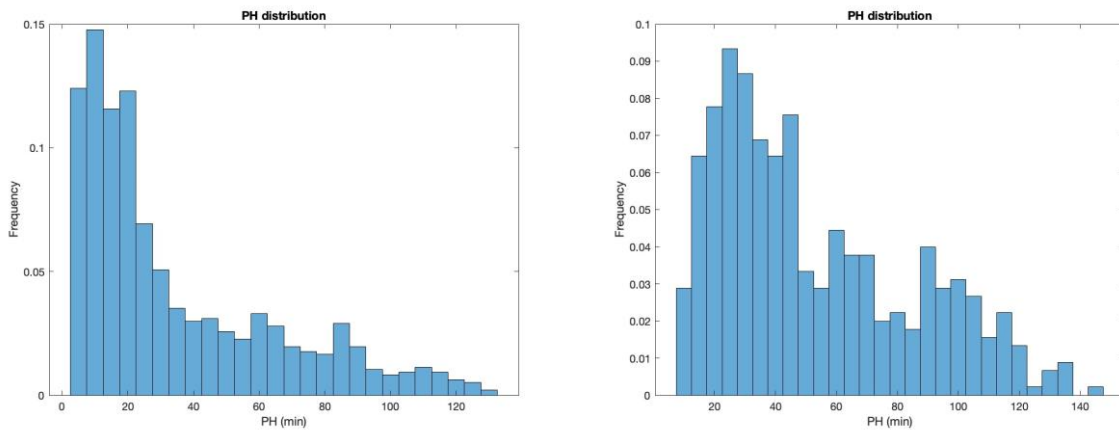


Fig. 5.4.1-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Andando a bilanciare le classi si è ottenuto un nuovo modello completo. Dai valori di Z-score emergono le feature significative, ovvero preprandial mean, afterPeak CV, range, diff10, afterPeak meanROC, minDR, maxDiff, peakAngle, TAR130, mealTime e numMeals. Rispetto ai modelli completi precedenti le feature diff20, gluMax e lastGlu perdono di importanza. Le considerazioni sul *prediction horizon* sono analoghe a quelle precedentemente fatte per i modelli costruiti sull'intero dataset.

5.4.2 Modello Ridotto

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int. | |
|--------------------------|---------|--------|-----------|---------|---------|--------|----------------|--------|
| (Intercept) | 0,1448 | 0,0461 | 0,0545 | 0,2350 | 3,1435 | 1,1558 | 1,0560 | 1,2649 |
| preprandial mean | 0,2473 | 0,1000 | 0,0512 | 0,4434 | 2,4720 | 1,2806 | 1,0526 | 1,5580 |
| Meal2Peak CV | 0,1954 | 0,0940 | 0,0112 | 0,3796 | 2,0796 | 1,2158 | 1,0113 | 1,4617 |
| afterPeak CV | 1,3368 | 0,2223 | 0,9012 | 1,7725 | 6,0147 | 3,8070 | 2,4626 | 5,8855 |
| range | 1,3208 | 0,3188 | 0,6960 | 1,9456 | 4,1431 | 3,7463 | 2,0056 | 6,9979 |
| diff10 | -0,9950 | 0,1443 | -1,2778 | -0,7121 | -6,8951 | 0,3697 | 0,2787 | 0,4906 |
| afterPeak meanROC | 1,1234 | 0,2703 | 0,5935 | 1,6532 | 4,1557 | 3,0752 | 1,8104 | 5,2236 |
| J-index | -1,3444 | 0,3522 | -2,0346 | -0,6541 | -3,8175 | 0,2607 | 0,1307 | 0,5199 |

| | | | | | | | | |
|------------------|---------|--------|---------|---------|---------|--------|--------|--------|
| BGRI | 0,8258 | 0,2057 | 0,4227 | 1,2289 | 4,0156 | 2,2837 | 1,5261 | 3,4174 |
| minDR | -0,5071 | 0,1389 | -0,7794 | -0,2349 | -3,6509 | 0,6022 | 0,4587 | 0,7907 |
| maxDiff | -0,3229 | 0,1114 | -0,5412 | -0,1045 | -2,8979 | 0,7241 | 0,5820 | 0,9008 |
| minDiff | 0,1581 | 0,0744 | 0,0123 | 0,3038 | 2,1256 | 1,1712 | 1,0124 | 1,3550 |
| gluMax | -0,9276 | 0,5356 | -1,9773 | 0,1221 | -1,7320 | 0,3955 | 0,1384 | 1,1299 |
| peakAngle | 0,3204 | 0,0654 | 0,1922 | 0,4486 | 4,8987 | 1,3776 | 1,2119 | 1,5661 |
| lastGlu | 0,6397 | 0,3034 | 0,0450 | 1,2343 | 2,1083 | 1,8958 | 1,0460 | 3,4361 |
| TAR130 | 0,3237 | 0,0847 | 0,1578 | 0,4897 | 3,8233 | 1,3823 | 1,1709 | 1,6318 |
| mealTime | -0,2167 | 0,0445 | -0,3040 | -0,1294 | -4,8659 | 0,8052 | 0,7379 | 0,8786 |
| numMeals | -0,1925 | 0,0518 | -0,2940 | -0,0909 | -3,7146 | 0,8249 | 0,7453 | 0,9131 |

Tab. 5.4.2-1 Output del modello ridotto

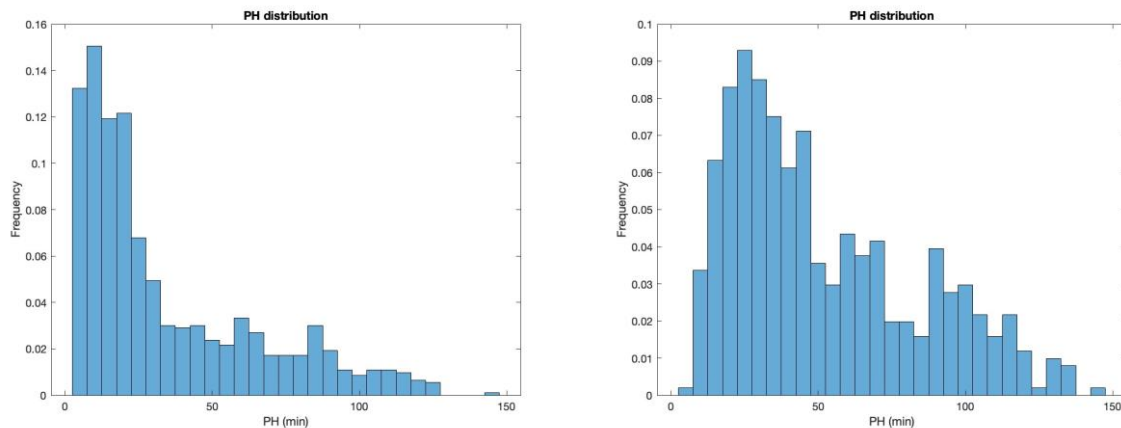


Fig. 5.4.2-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Il modello ridotto comprende 17 variabili: preprandial mean, afterPeak CV, range, diff10, afterPeak meanROC, minDR, maxDiff, peakAngle, TAR130, mealTime e numMeals, già risultate significative nel modello completo, e anche Meal2Peak CV, J-index, BGRI, minDiff, gluMax e lastGlu.

5.4.3 Performance sull'intero dataset

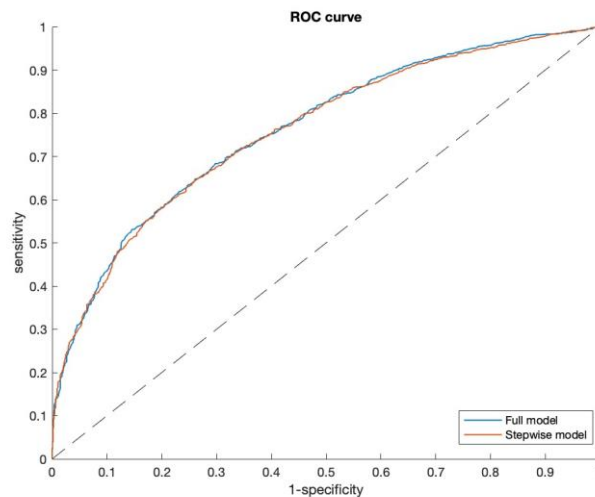


Fig. 5.4.3-1 ROC curve del modello completo e ridotto applicati all'intero dataset.

| | Full model | Stepwise model |
|---|-----------------|-----------------|
| AUC | 0,7626 | 0,7593 |
| Accuracy | 0,6925 | 0,6912 |
| Sensitivity | 0,6827 | 0,6476 |
| Specificity | 0,7034 | 0,7396 |
| Precision | 0,7181 | 0,7340 |
| TG: median (25 th - 75 th percentile) | 20 (10 - 50) | 20 (10 - 50) |

Tab. 5.4.3-2 Risultati di performance sull'intero dataset.

Confrontando i risultati ottenuti con quelli in cui è stato utilizzato il dataset sbilanciato, l'AUC e la *precision* risultano simili, l'*accuracy* leggermente inferiore, la *sensitivity* più alta e *specificity* più bassa. Ciò indica che il modello ha una buona capacità di individuare correttamente sia i casi positivi che negativi, causando un numero limitato di falsi positivi.

5.5 Bilanciamento delle classi: analisi del modello completo e del modello con selezione delle variabili applicati a training e test set

Come nella sezione precedente, si procede alla costruzione del modello completo e ridotto sul training set, seguita dalla fase di valutazione sul test set. È stato eseguito anche un test di robustezza, variando in maniera casuale l'estrazione di training e test set. Di seguito, vengono presentati i risultati dei modelli.

5.5.1 Modello Completo

| variable | beta | SE | Conf.Int. 1 | | Z-score | Odds | Odds Conf.Int. 1 | |
|-------------------------|---------|--------|-------------|---------|---------|--------|------------------|---------|
| intercept | 0,0751 | 0,0773 | -0,0763 | 0,2265 | 0,9718 | 1,0780 | 0,9265 | 1,2542 |
| mean | -0,2349 | 1,0413 | -2,2759 | 1,8061 | -0,2256 | 0,7907 | 0,1027 | 6,0868 |
| preprandial mean | 0,6643 | 0,2967 | 0,0827 | 1,2459 | 2,2387 | 1,9431 | 1,0862 | 3,4762 |
| median | 0,0917 | 0,2437 | -0,3860 | 0,5694 | 0,3763 | 1,0961 | 0,6798 | 1,7673 |
| std | 0,5829 | 1,1536 | -1,6781 | 2,8439 | 0,5053 | 1,7912 | 0,1867 | 17,1822 |
| CV | -0,2396 | 0,7304 | -1,6712 | 1,1920 | -0,3280 | 0,7870 | 0,1880 | 3,2937 |
| Meal2Peak CV | 0,1561 | 0,2404 | -0,3151 | 0,6274 | 0,6493 | 1,1690 | 0,7297 | 1,8727 |
| afterPeak CV | 1,2611 | 0,3136 | 0,6464 | 1,8758 | 4,0211 | 3,5293 | 1,9087 | 6,5262 |
| range | 1,4673 | 0,5647 | 0,3606 | 2,5740 | 2,5986 | 4,3374 | 1,4341 | 13,1183 |
| diff10 | -0,6818 | 0,3058 | -1,2810 | -0,0825 | -2,2297 | 0,5057 | 0,2777 | 0,9208 |
| diff20 | -2,0564 | 1,3961 | -4,7927 | 0,6799 | -1,4730 | 0,1279 | 0,0083 | 1,9736 |
| diff30 | 0,1314 | 0,1996 | -0,2598 | 0,5226 | 0,6582 | 1,1404 | 0,7712 | 1,6864 |
| MAGE | -0,2029 | 0,1403 | -0,4778 | 0,0720 | -1,4464 | 0,8164 | 0,6201 | 1,0747 |
| EF | -0,0208 | 0,0987 | -0,2143 | 0,1726 | -0,2113 | 0,9794 | 0,8071 | 1,1883 |
| lastROC | -0,1364 | 0,6063 | -1,3248 | 1,0519 | -0,2250 | 0,8725 | 0,2659 | 2,8631 |
| meanROC | 0,1564 | 0,2485 | -0,3306 | 0,6435 | 0,6296 | 1,1693 | 0,7185 | 1,9030 |

| | | | | | | | | |
|--------------------------|---------|--------|---------|---------|---------|--------|--------|----------|
| afterPeak meanROC | 1,3496 | 0,5194 | 0,3317 | 2,3676 | 2,5986 | 3,8559 | 1,3933 | 10,6713 |
| J-index | -1,5407 | 1,3875 | -4,2603 | 1,1789 | -1,1104 | 0,2142 | 0,0141 | 3,2509 |
| LBGI | -0,0631 | 0,4080 | -0,8627 | 0,7366 | -0,1546 | 0,9389 | 0,4220 | 2,0888 |
| BGRI | 1,1427 | 1,0191 | -0,8549 | 3,1402 | 1,1212 | 3,1352 | 0,4253 | 23,1088 |
| ADRR | 0,5598 | 0,4886 | -0,3979 | 1,5176 | 1,1457 | 1,7504 | 0,6717 | 4,5612 |
| minDR | -0,2876 | 0,1921 | -0,6641 | 0,0889 | -1,4972 | 0,7501 | 0,5147 | 1,0930 |
| maxDR | 0,1575 | 0,5904 | -0,9996 | 1,3146 | 0,2668 | 1,1706 | 0,3680 | 3,7234 |
| meanDR | -0,7571 | 0,7142 | -2,1570 | 0,6428 | -1,0600 | 0,4690 | 0,1157 | 1,9018 |
| maxDiff | -0,5190 | 0,3006 | -1,1082 | 0,0702 | -1,7263 | 0,5951 | 0,3302 | 1,0728 |
| afterPeak maxDiff | 0,2150 | 0,3347 | -0,4410 | 0,8711 | 0,6423 | 1,2399 | 0,6434 | 2,3894 |
| minDiff | 0,2676 | 0,2567 | -0,2355 | 0,7708 | 1,0426 | 1,3069 | 0,7902 | 2,1615 |
| Meal2Peak minDiff | -0,1032 | 0,2516 | -0,5963 | 0,3900 | -0,4100 | 0,9020 | 0,5508 | 1,4770 |
| gluMax | -1,8586 | 0,9897 | -3,7984 | 0,0812 | -1,8779 | 0,1559 | 0,0224 | 1,0846 |
| last local max | -1,5467 | 1,8904 | -5,2520 | 2,1586 | -0,8182 | 0,2130 | 0,0052 | 8,6589 |
| prominence | -0,0917 | 0,1025 | -0,2927 | 0,1092 | -0,8946 | 0,9124 | 0,7463 | 1,1154 |
| gluPos | -0,0048 | 0,1790 | -0,3557 | 0,3462 | -0,0266 | 0,9953 | 0,7007 | 1,4136 |
| gluNeg | -0,0658 | 0,1341 | -0,3286 | 0,1969 | -0,4912 | 0,9363 | 0,7199 | 1,2176 |
| Meal2Peak Time | 0,0651 | 0,2034 | -0,3335 | 0,4637 | 0,3202 | 1,0673 | 0,7164 | 1,5900 |
| peakAngle | 0,2757 | 0,0986 | 0,0825 | 0,4689 | 2,7970 | 1,3174 | 1,0860 | 1,5982 |
| mealGlu | -0,0642 | 0,2335 | -0,5218 | 0,3934 | -0,2751 | 0,9378 | 0,5934 | 1,4820 |
| lastGlu | 2,2084 | 1,6542 | -1,0337 | 5,4505 | 1,3351 | 9,1011 | 0,3557 | 232,8842 |
| TAR130 | 0,2372 | 0,1973 | -0,1496 | 0,6240 | 1,2019 | 1,2677 | 0,8610 | 1,8664 |
| TAR180 | -0,0103 | 0,1904 | -0,3835 | 0,3629 | -0,0539 | 0,9898 | 0,6815 | 1,4375 |
| mealTime | -0,2089 | 0,0522 | -0,3111 | -0,1067 | -4,0053 | 0,8115 | 0,7326 | 0,8988 |
| numMeals | -0,1899 | 0,0633 | -0,3140 | -0,0658 | -3,0004 | 0,8270 | 0,7306 | 0,9363 |

Tab. 5.5.1-1 Output del modello completo (training set).

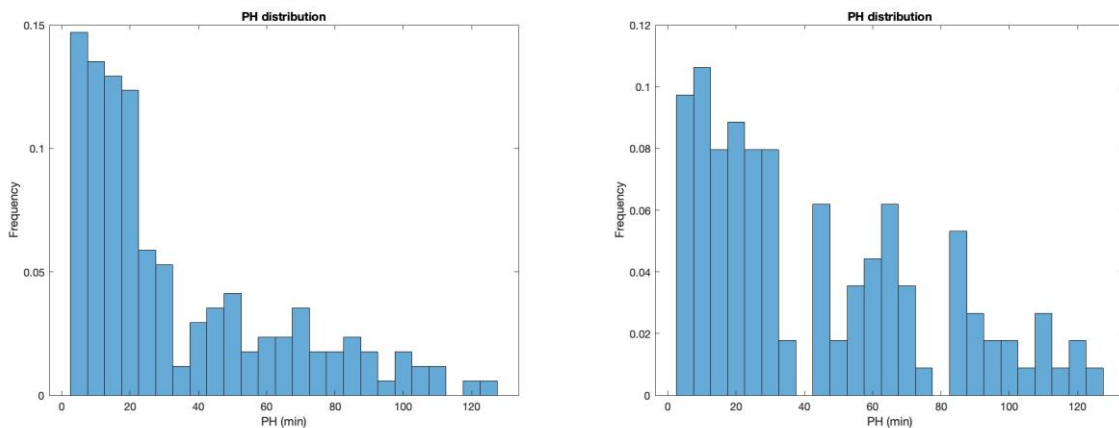


Fig. 5.5.1-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Nel modello costruito sul training set bilanciato le uniche feature statisticamente significative risultano essere preprandial mean, afterPeak CV, range, diff10, afterPeak meanROC, peakAngle, mealTime e numMeals.

5.5.2 Modello Ridotto

| variable | beta | SE | Conf.Int. | | Z-score | Odds | Odds Conf.Int. | |
|--------------------------|---------|--------|-----------|---------|---------|--------|----------------|--------|
| (Intercept) | 0,0825 | 0,0634 | -0,0417 | 0,2068 | 1,3019 | 1,0860 | 0,9591 | 1,2297 |
| preprandial mean | 0,4828 | 0,1096 | 0,2680 | 0,6976 | 4,4058 | 1,6207 | 1,3074 | 2,0090 |
| afterPeak CV | 1,1756 | 0,2672 | 0,6519 | 1,6992 | 4,4003 | 3,2401 | 1,9193 | 5,4698 |
| range | 1,5406 | 0,3657 | 0,8237 | 2,2575 | 4,2122 | 4,6674 | 2,2790 | 9,5589 |
| diff10 | -0,7686 | 0,1881 | -1,1374 | -0,3999 | -4,0854 | 0,4637 | 0,3207 | 0,6704 |
| diff20 | -1,3985 | 0,5750 | -2,5255 | -0,2714 | -2,4321 | 0,2470 | 0,0800 | 0,7623 |
| MAGE | -0,2629 | 0,1116 | -0,4817 | -0,0441 | -2,3553 | 0,7688 | 0,6177 | 0,9568 |
| afterPeak meanROC | 1,3627 | 0,4146 | 0,5501 | 2,1753 | 3,2870 | 3,9067 | 1,7335 | 8,8046 |
| J-index | -0,9473 | 0,5117 | -1,9502 | 0,0557 | -1,8511 | 0,3878 | 0,1422 | 1,0573 |
| BGRI | 1,1083 | 0,2915 | 0,5369 | 1,6797 | 3,8017 | 3,0293 | 1,7107 | 5,3641 |
| ADRR | 0,4229 | 0,3685 | -0,2994 | 1,1451 | 1,1476 | 1,5263 | 0,7413 | 3,1427 |
| minDR | -0,3014 | 0,1778 | -0,6499 | 0,0471 | -1,6952 | 0,7398 | 0,5221 | 1,0482 |
| meanDR | -1,1239 | 0,3860 | -1,8804 | -0,3674 | -2,9118 | 0,3250 | 0,1525 | 0,6926 |
| maxDiff | -0,3104 | 0,1256 | -0,5565 | -0,0643 | -2,4719 | 0,7332 | 0,5732 | 0,9377 |
| minDiff | 0,1485 | 0,0832 | -0,0145 | 0,3115 | 1,7852 | 1,1601 | 0,9856 | 1,3654 |
| gluMax | -1,8904 | 0,7202 | -3,3021 | -0,4788 | -2,6248 | 0,1510 | 0,0368 | 0,6195 |
| peakAngle | 0,2557 | 0,0713 | 0,1159 | 0,3955 | 3,5846 | 1,2913 | 1,1229 | 1,4851 |
| lastGlu | 1,4373 | 0,3739 | 0,7044 | 2,1702 | 3,8436 | 4,2093 | 2,0226 | 8,7603 |
| mealTime | -0,1988 | 0,0501 | -0,2970 | -0,1007 | -3,9697 | 0,8197 | 0,7430 | 0,9042 |
| numMeals | -0,2099 | 0,0587 | -0,3249 | -0,0948 | -3,5758 | 0,8107 | 0,7226 | 0,9095 |

Tab. 5.5.2-1 Output del modello ridotto (training set).

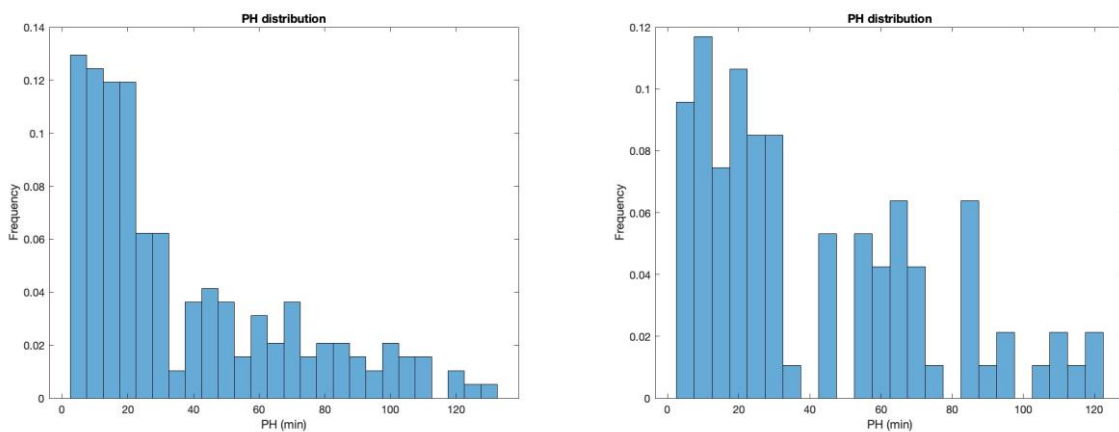


Fig. 5.5.2-2 Distribuzione PH per casi appartenenti alla classe 1, predetti correttamente (a sinistra) e scorrettamente (a destra).

Osservando i risultati del modello ridotto, si nota che entrano in gioco anche le variabili diff20, MAGE, J-index, BGRI, ADRR, minDR, meanDR, maxDiff, minDiff, gluMax e lastGlu, non risultate significative nel modello completo.

5.5.3 Performance sul test set

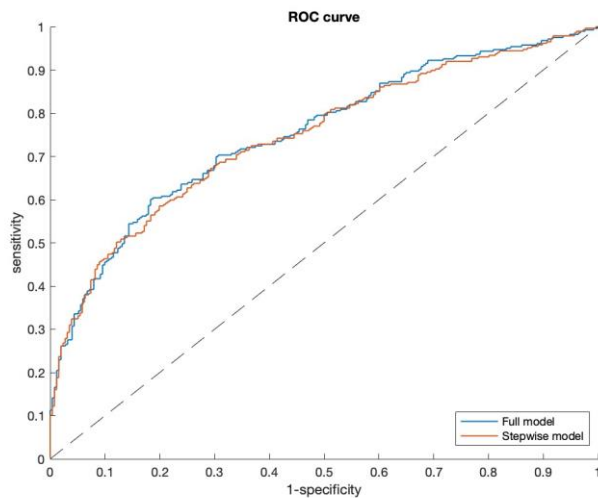


Fig. 5.5.3-1 ROC curve del modello completo e ridotto applicati al test set.

| | Full model | Stepwise model |
|---|-------------------|-----------------------|
| AUC | 0,7536 | 0,7469 |
| Accuracy | 0,7022 | 0,6888 |
| Sensitivity | 0,6007 | 0,6725 |
| Specificity | 0,8167 | 0,7070 |
| Precision | 0,7870 | 0,7201 |
| TG: median (25 th - 75 th percentile) | 20 (10 - 50) | 25 (10 - 55) |

Tab. 5.5.3-2 Risultati di performance sul test set.

Le performance sul test set sono altrettanto soddisfacenti: l'*accuracy*, la *specificity* e la *precision* si mantengono sul 70% e la *sensitivity* leggermente inferiore alle altre metriche indica che il modello ha sempre la tendenza di sottostimare i casi positivi.

È stato effettuato anche un test di robustezza, variando in maniera casuale l'estrazione del training set e del test set, in modo da costruire il modello su differenti training set e testare ciascun modello su un test set diverso. Viene riportato, di seguito, lo spaghetti plot contenente le curve ROC risultanti dall'applicazione dei modelli completi.

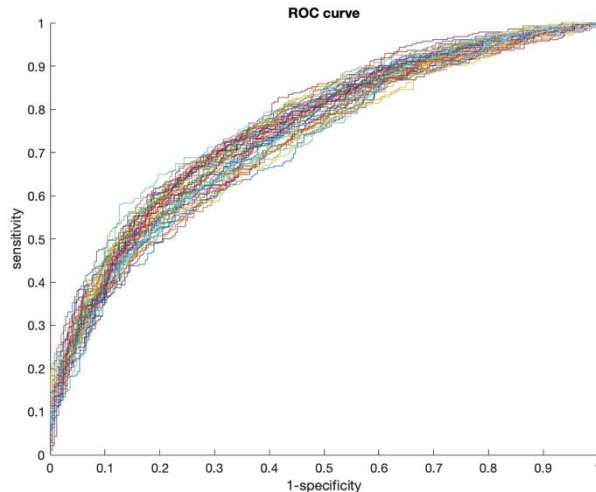


Fig. 5.5.3-3 Spaghetti plot con ROC curve dei modelli completi.

Dallo spaghetti plot si osserva che le curve ROC ottenute risultano simili, evidenziando una buona robustezza del modello di fronte alle variazioni nei training e test set.

5.6 Risultati principali

Dall'analisi svolta si nota che i modelli costruiti sul dataset le cui classi non sono bilanciate hanno complessivamente una *accuracy* più elevata (75%), grazie al valore elevato di *specificity* (90%), il che è influenzato dalla predominanza di casi appartenenti alla classe 0 rispetto a quelli della classe 1. Tuttavia, questo comporta però un inferiore *sensitivity* (35%), indicando che i modelli hanno la tendenza a classificare le nuove osservazioni come appartenenti alla classe 0, ovvero che non portano ad un evento di ipoglicemia. In termini di *Time Gain*, per i modelli su dataset sbilanciato il valore mediano è 15 minuti, per i modelli su dataset bilanciato il guadagno temporale mediano è 20 minuti.

Di conseguenza, risulta preferibile fare il training del modello sul dataset bilanciato, poiché si ottiene un miglioramento della *sensitivity* (65%), mantenendo al contempo prestazioni complessive buone (*accuracy* \cong 70%, *specificity* e *precision* \cong 73%).

Inoltre, tra i risultati del modello completo e ridotto non ci sono differenze significative: nonostante il modello ridotto presenti una complessità inferiore, la performance di predizione rimane comunque soddisfacente. Infine, dall'analisi delle distribuzioni dei *Prediction Horizon* emerge che la lunghezza del PH non influisce particolarmente sulle prestazioni del modello: non si osserva una grande propensione a classificare scorrettamente le osservazioni con PH elevato rispetto a quelle con PH più breve.

Vengono riportati di seguito due esempi di periodi post-prandiali classificati correttamente come appartenenti alla classe 1 (Fig. 5.6-1). Considerando che tra le feature significative, quelle che hanno l'effetto di aumentare la probabilità che si verifichi l'evento ipoglicemico sono preprandial mean, afterPeak CV, range, afterPeak meanROC, peakAngle; nel primo periodo giocano un ruolo fondamentale nell'aumentare la probabilità dell'evento le feature afterPeak CV e range, poiché hanno valori superiori alla media. Nel secondo esempio, invece, le feature che influiscono maggiormente sono preprandial mean, afterPeak meanROC e range.

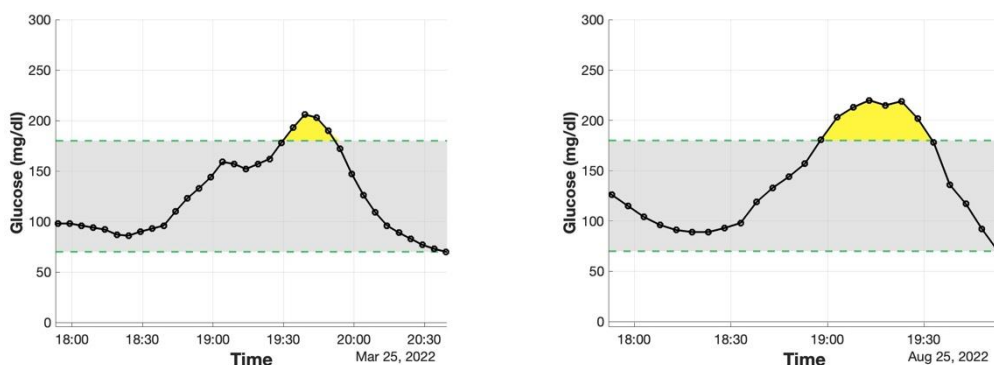


Fig. 5.6-1 Esempi di periodi post-prandiali predetti correttamente come appartenenti alla classe 1.

Dal confronto con le prestazioni ottenute dagli algoritmi per la predizione delle ipoglicemie post-beriatriche proposti da Laguna et al. e Prendin et al., si evince che il modello di Laguna ha una *precision* notevolmente più bassa rispetto al modello ARIMA di Prendin e al modello logistico sviluppato in questo lavoro (Mod. Laguna: 24.7% e Mod. ARIMA: 79.1%). La *sensitivity* di entrambi i modelli risulta però più elevata rispetto al modello logistico (Mod Laguna: 81.2% e Mod. ARIMA: 100%). Tuttavia, è importante sottolineare che l'elevata *sensitivity* del modello ARIMA è attribuibile al filtraggio dei dati CGM. Al contrario, l'analisi condotta in questo lavoro si è basata sui dati grezzi.

L'ottimizzazione della *sensitivity* potrebbe essere un obiettivo futuro, in modo da ridurre il numero di falsi negativi e migliorare la capacità del modello di riconoscere correttamente gli episodi ipoglicemici.

Capitolo 6

Conclusioni

In seguito all'operazione di bypass gastrico, una delle complicanze che si manifesta a distanza di alcuni anni dall'intervento è la comparsa di eventi ipoglicemici, solitamente da 1 a 3 ore dopo l'assunzione di un pasto. La capacità di predire in tempo reale questi episodi post-prandiali è di fondamentale importanza per l'applicazione di azioni preventive atte a mitigare la loro frequenza e durata.

L'obiettivo di questo lavoro di tesi è stato quello di sviluppare un algoritmo basato sulla regressione logistica per caratterizzare e poi predire in tempo reale le ipoglicemie post-prandiali in soggetti che hanno subito un intervento di bypass gastrico. Questo algoritmo ha utilizzato variabili estratte dall'analisi di dati CGM nel periodo post-prandiale, integrate con informazioni sul pasto.

Nell'analisi preliminare del dataset, sono state calcolate diverse metriche in modo da esaminare l'andamento glicemico nei pazienti PBH. I risultati hanno indicato che, per la maggior parte del tempo, la glicemia di questi pazienti si mantiene nel range normoglicemico (70-180 mg/dL), con episodi sporadici di iperglicemia e ipoglicemia. Tuttavia, è emersa una significativa variabilità glicemica. Dato che gli individui affetti da ipoglicemie post-bariatriche soffrono di una grande variabilità glicemica dopo l'assunzione di pasti, si è deciso di effettuare l'analisi di correlazione e successivamente di predizione delle ipoglicemie post-prandiali. Per fare ciò si sono sviluppati dei framework ad hoc, tramite la definizione di opportune finestre temporali post-prandiali, al fine di estrarre e identificare le feature maggiormente correlate alla presenza dell'evento ipoglicemico. Successivamente, sono stati sviluppati dei modelli a singola variabile per valutare l'effetto di ciascuna variabile sull'evento, e modelli completi e ridotti (selezionando le variabili più significative), sia su tutto il dataset che sul training set, per valutare l'effetto complessivo delle variabili sull'outcome.

I risultati principali di questo lavoro di tesi sono i seguenti:

1. Nell'analisi di caratterizzazione delle ipoglicemie, come evidenziato dall'analisi delle curve ROC e dai valori di Z-score, le feature più rilevanti risultano essere:
 - a. Le variabili descrittive del cambiamento del trend del segnale glicemico (CV, diff10, diff20, diff30, lastROC, meanROC, lastGlu, MAGE, maxDiff, gluPos);

- b. Le metriche di 'rischio' (LBGI, ADRR, minDR e anche TAR130);
- c. le feature riguardanti il pasto (mealTime, Meal2Peak Time).

Valutando i modelli sul dataset e sul test set si ottengono ottimi risultati di performance: *accuracy*, *sensitivity*, *specificity* e *precision* con valori intorno al 90% indicano una buona capacità di classificare correttamente gli eventi ipoglicemici e non.

2. Per quel che riguarda la predizione delle ipoglicemie, sono state calcolate delle feature locali legate a particolari regioni del segnale. Sono stati quindi sviluppati dei modelli completi e ridotti, inizialmente sull'intero dataset e poi solo sul training set (bilanciato e non), per predire gli eventi ipoglicemici futuri. Dall'analisi è emerso che i modelli costruiti sul dataset e sul training set bilanciato presentano le migliori performance di predizione, garantendo in media un'accuratezza del 70%, una *specificity* e una *precision* superiori al 70% e una *sensitivity* media del 65%. È importante notare che, sebbene i risultati complessivi siano validi, la *sensitivity* inferiore alle altre metriche indica che i modelli hanno la tendenza di sottostimare gli eventi ipoglicemici, causando diversi falsi negativi. Il guadagno temporale mediano ottenuto è di 20 minuti.

Sviluppi futuri

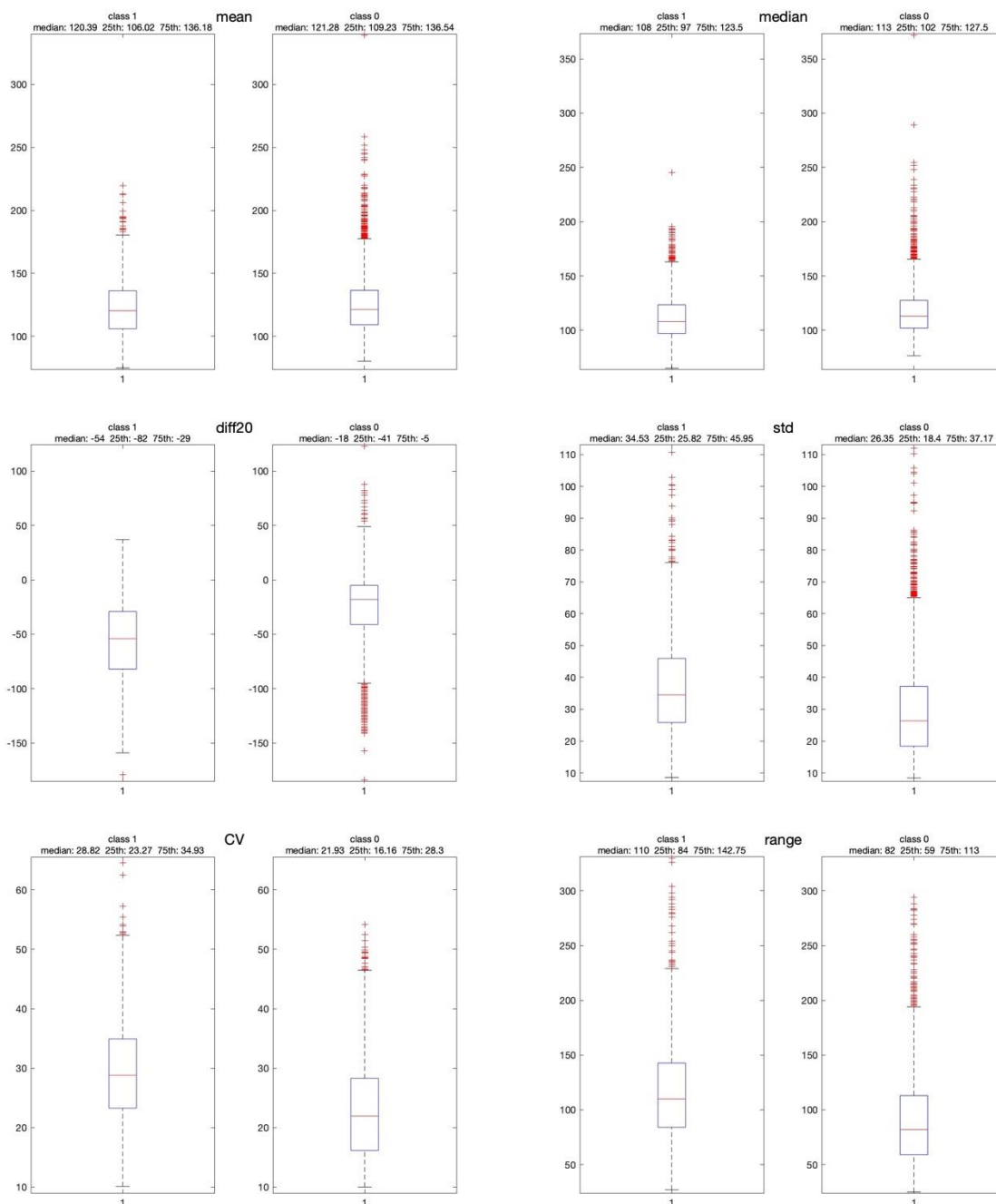
In prosecuzione del lavoro svolto si potrebbero sviluppare modelli personalizzati per descrivere la grande variabilità interindividuale presente nei soggetti. Inoltre, si potrebbe indagare l'uso di altre tecniche di machine learning, quali Random Forest, SVM, o anche metodologie di deep learning come le reti neurali. Un altro punto da migliorare riguarda le tecniche di pre-processing che consentono di ridurre al minimo le anomalie delle tracce CGM. Altro elemento importante può essere l'integrazione di informazioni disponibili con altri dati derivanti da sensori wearable, ad esempio dati di *heart rate*.

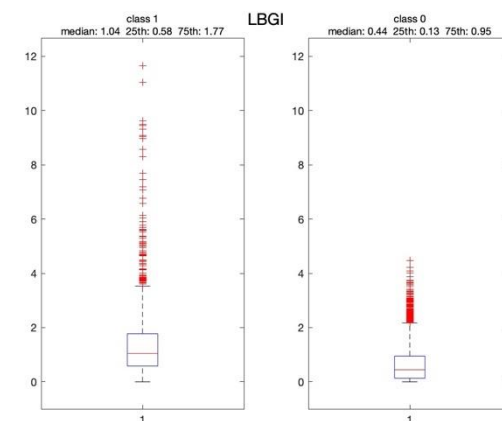
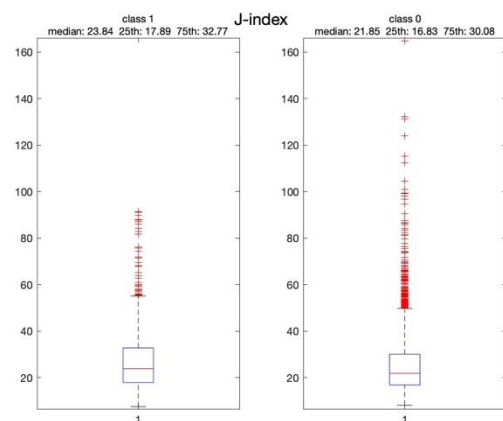
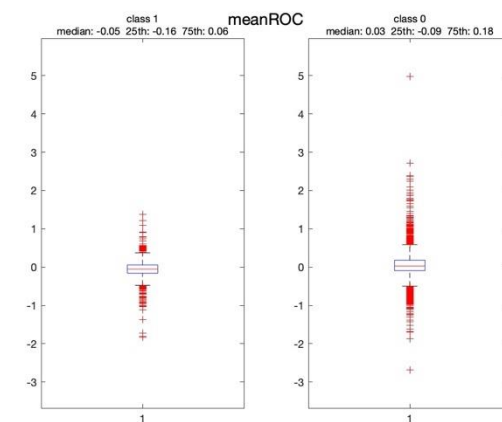
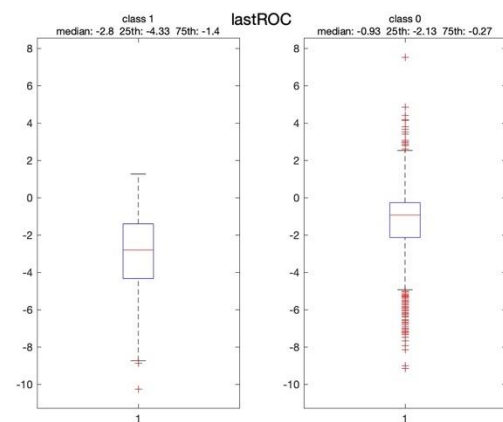
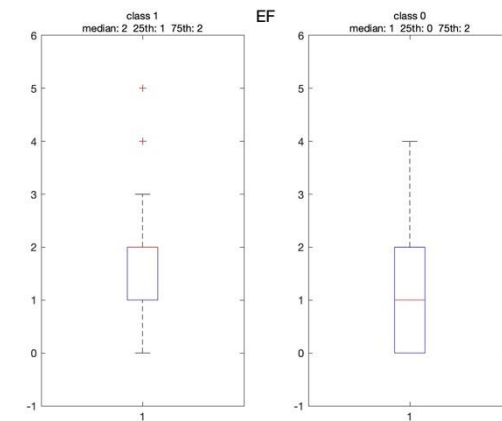
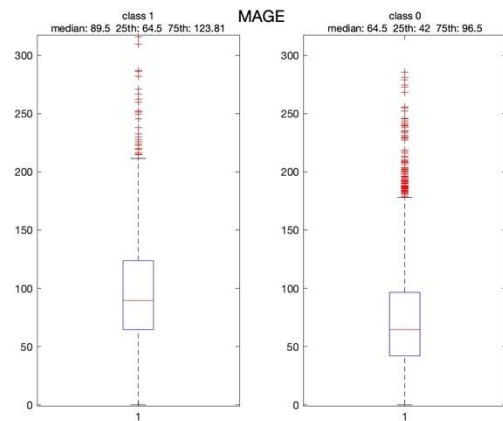
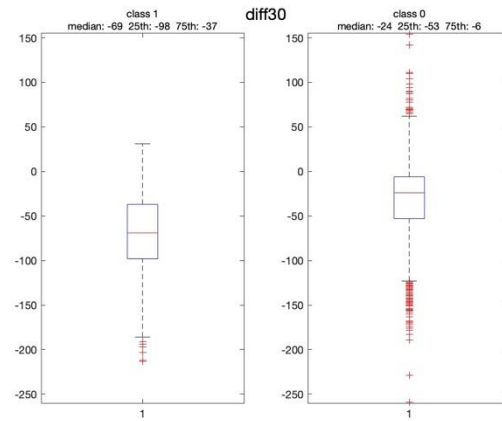
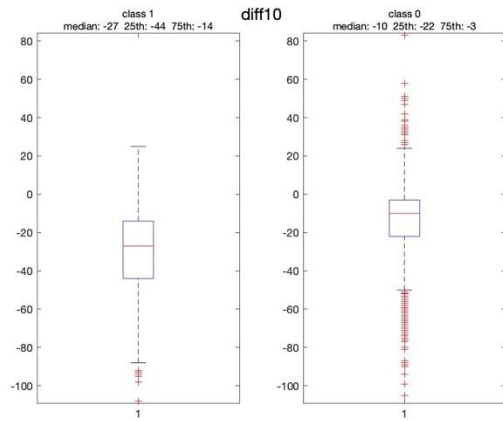
La validità delle performance dei modelli sviluppati suggerisce un possibile impiego futuro nell'ambito clinico. L'inclusione di tali modelli nella pratica clinica potrebbe offrire un supporto prezioso per la gestione delle ipoglicemie post-prandiali. Questo potrebbe tradursi in aggiustamenti mirati nella gestione dei pasti e delle terapie, contribuendo a prevenire situazioni di ipoglicemia, e di conseguenza, apportando benefici alla salute dei pazienti.

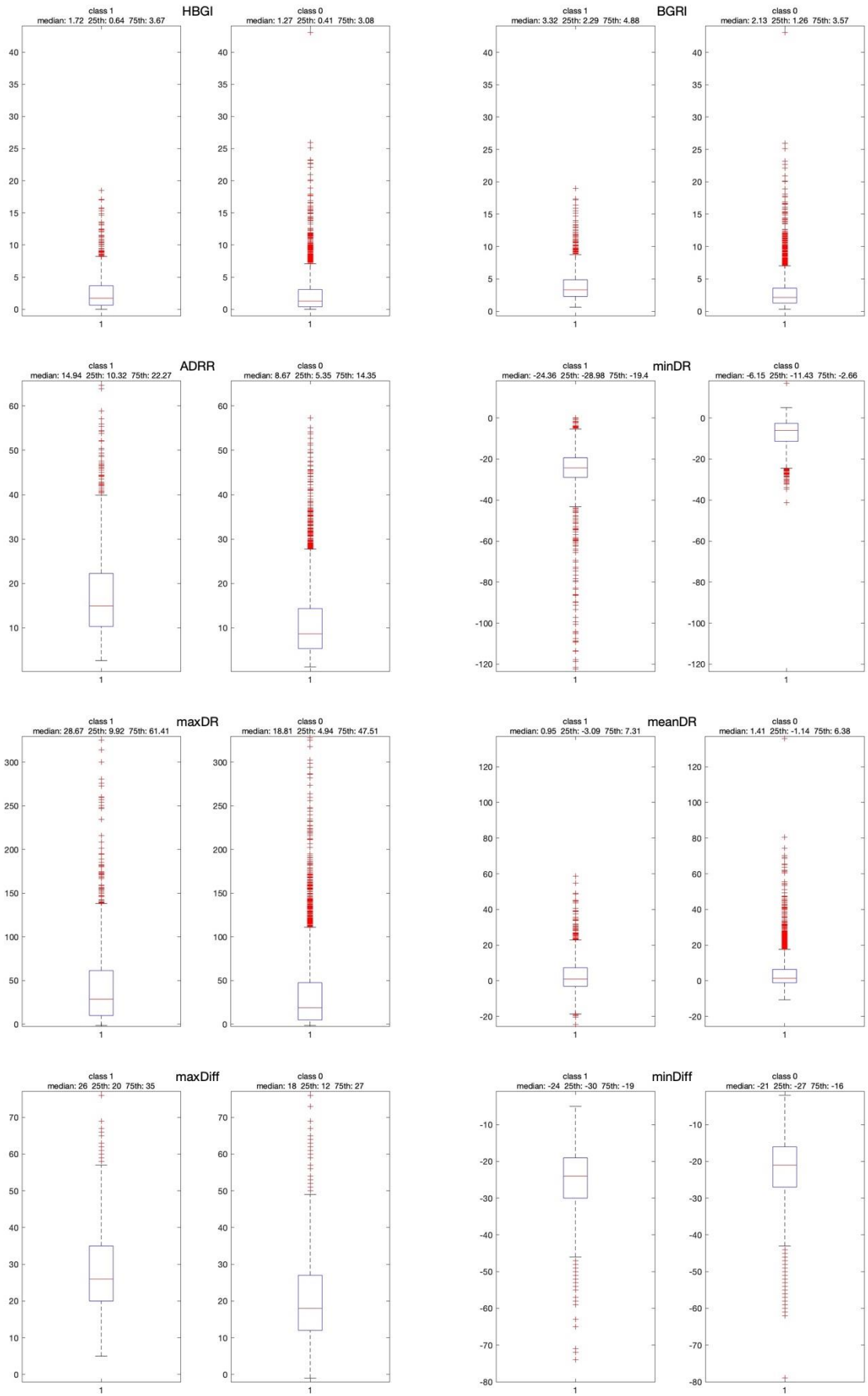
Appendice A

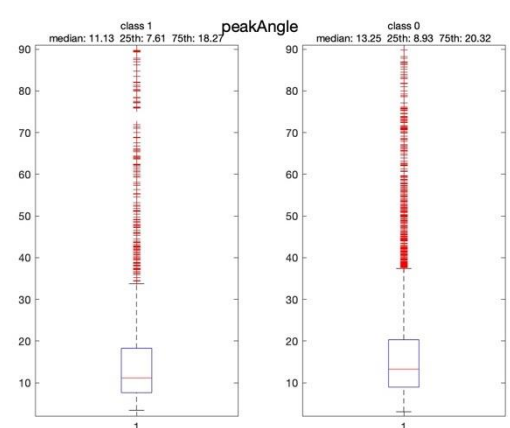
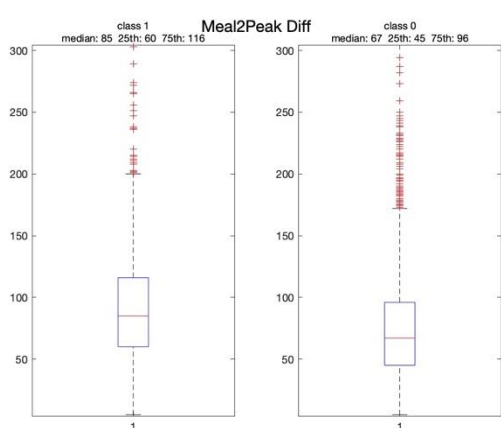
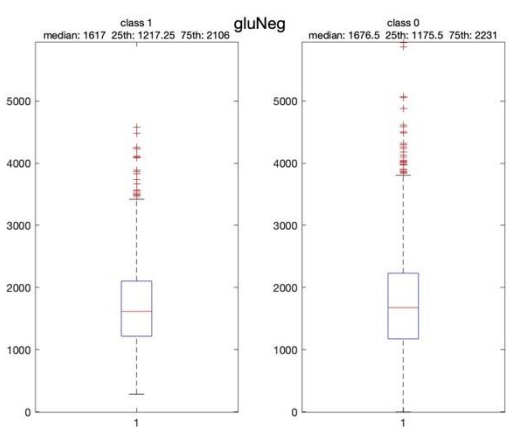
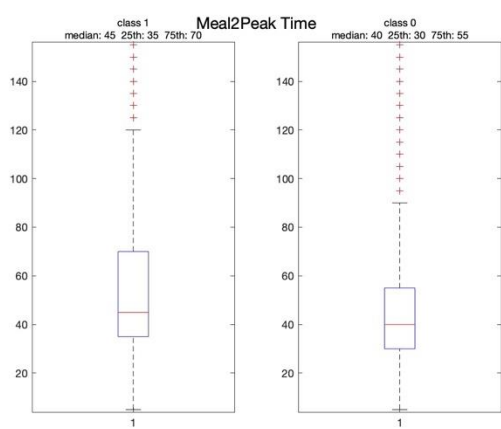
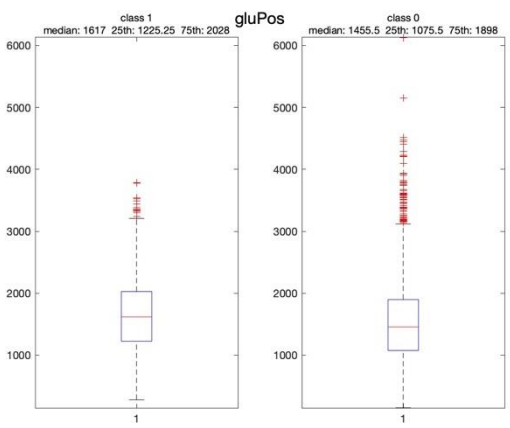
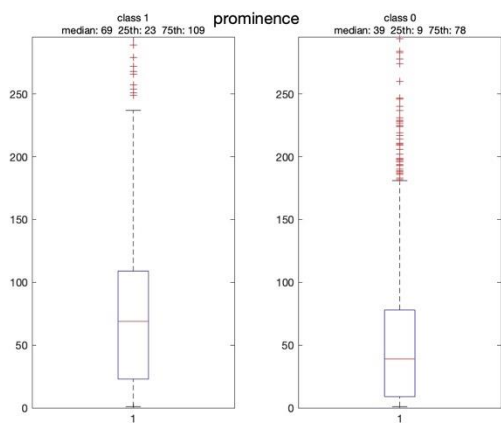
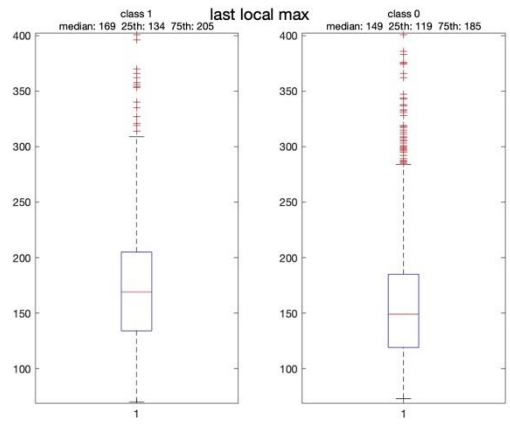
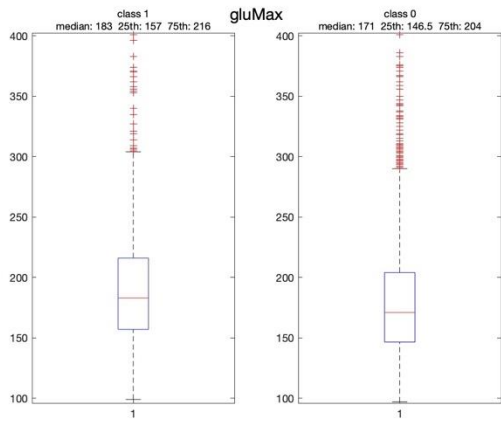
A.1 Distribuzione delle variabili nelle due classi

Nella prima parte dell'appendice vengono riportati i boxplot relativi alle variabili nelle due classi, ottenuti in seguito al pre-processing dei dati e menzionati nel capitolo 3. Questi boxplot sono stati utilizzati per confrontare le distribuzioni delle variabili tra le due classi.









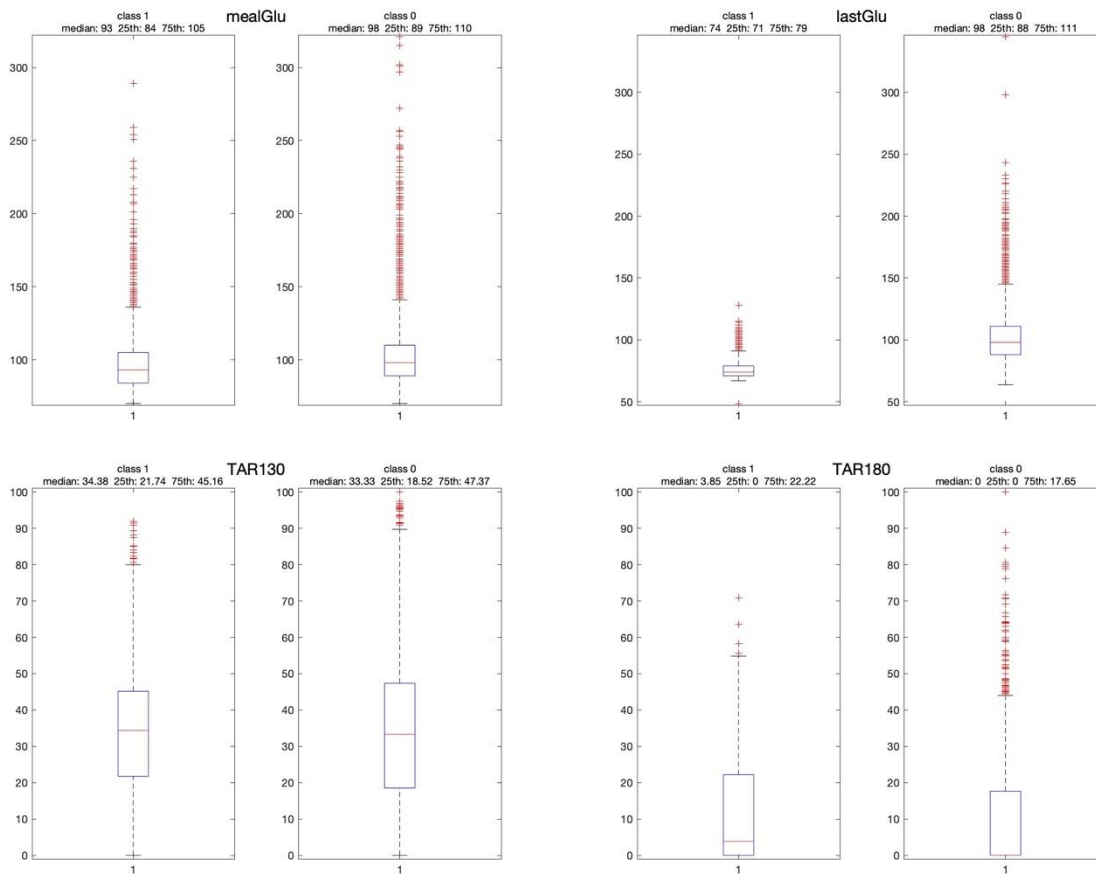
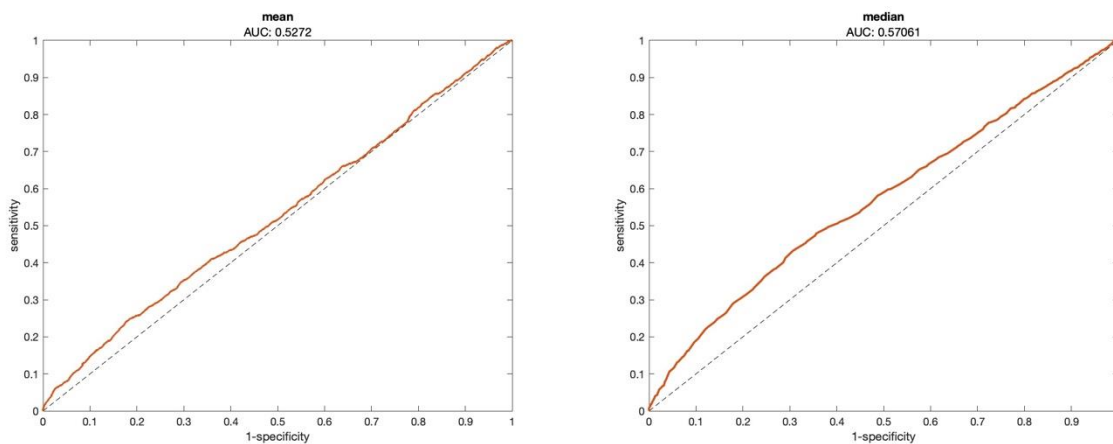
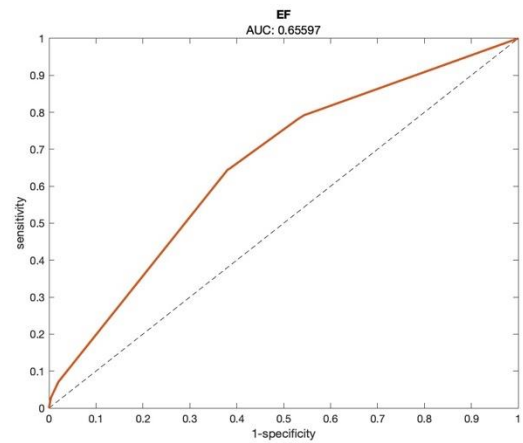
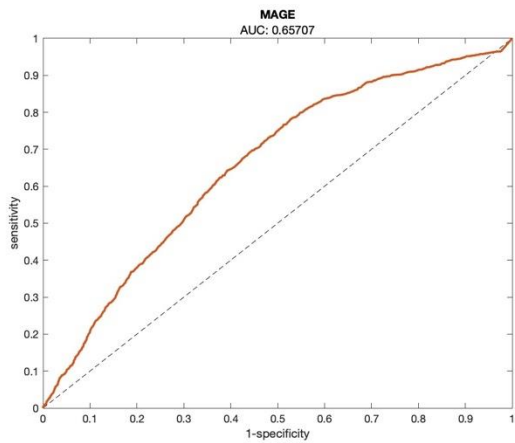
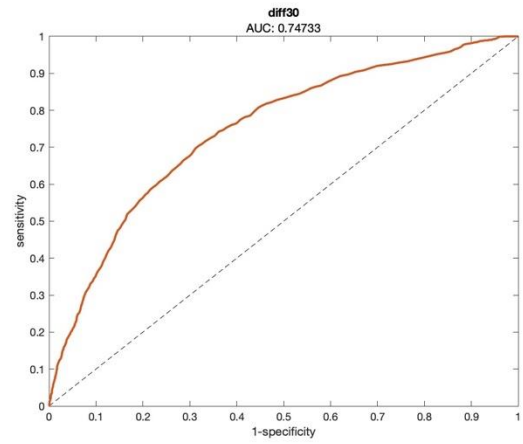
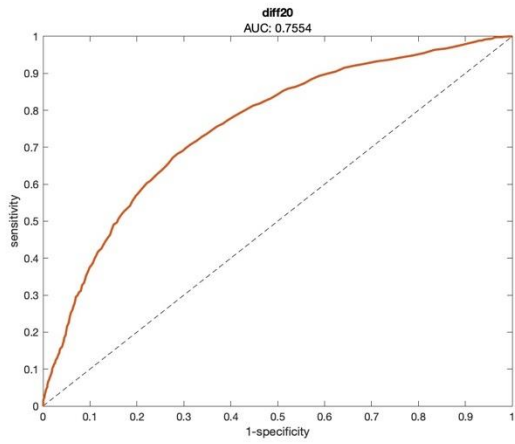
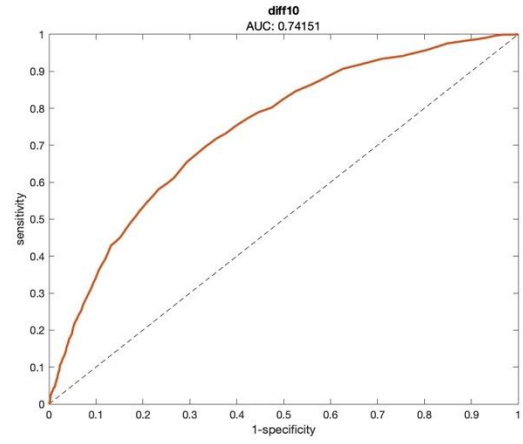
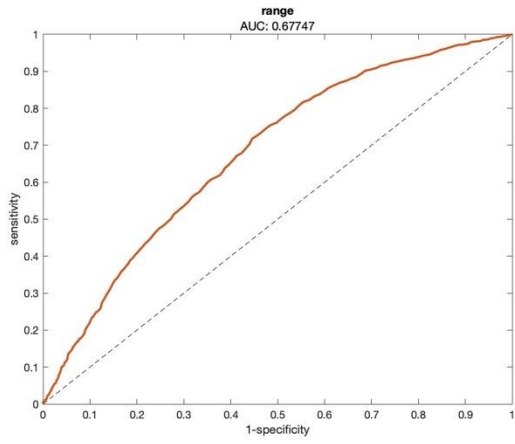
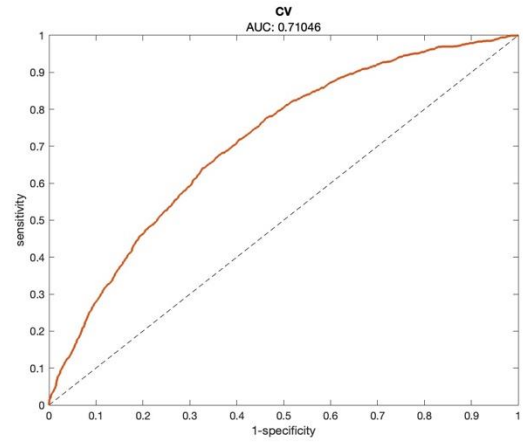
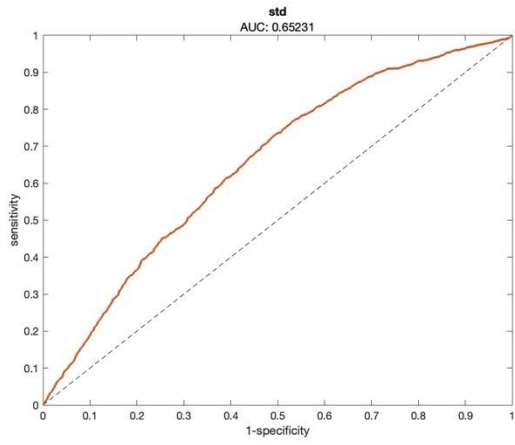


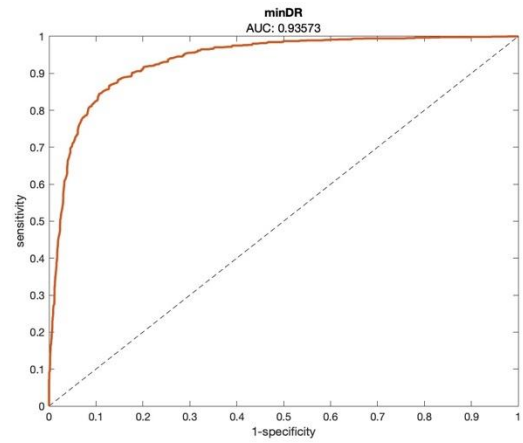
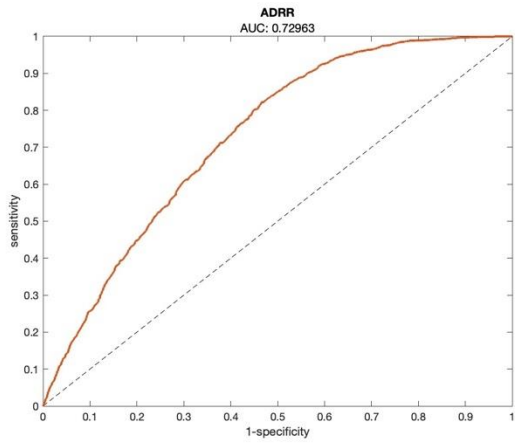
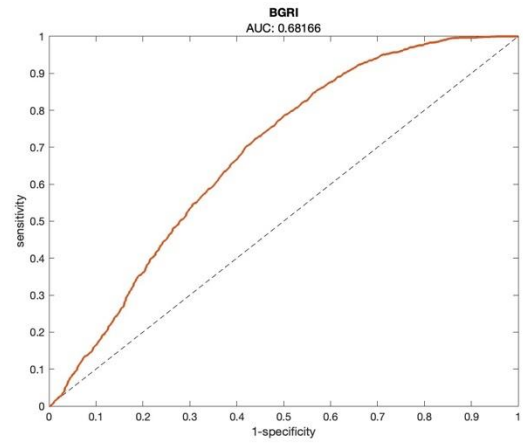
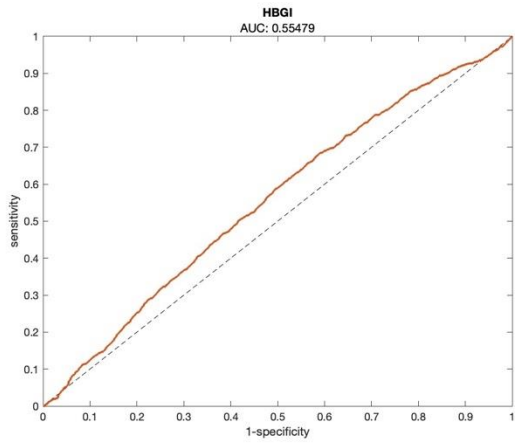
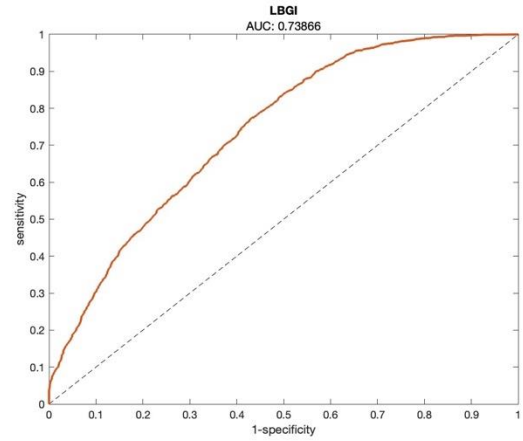
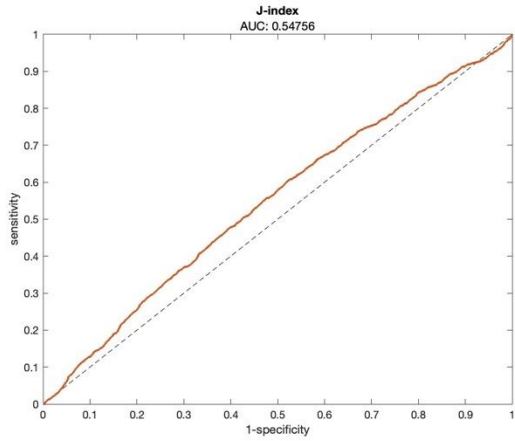
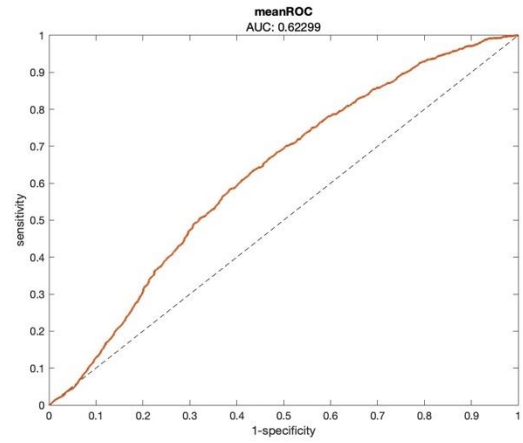
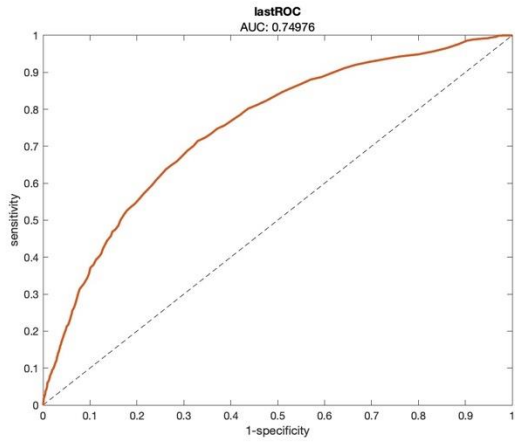
Fig. A.1 Boxplot delle variabili estratte per la classe 1 e 0, con valori di mediana, 25° e 75° percentile.

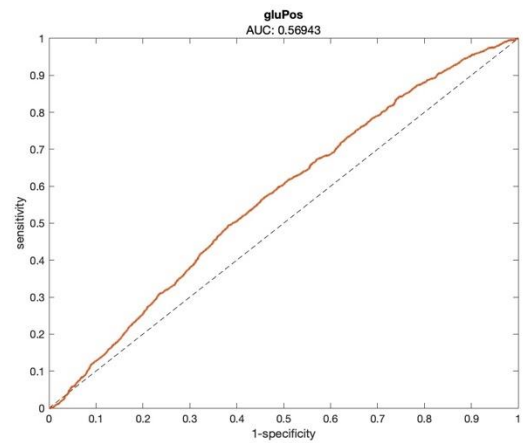
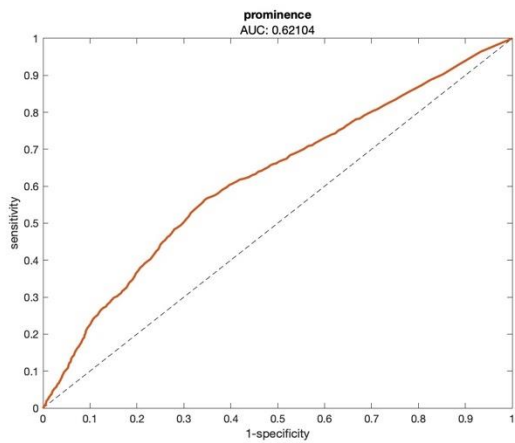
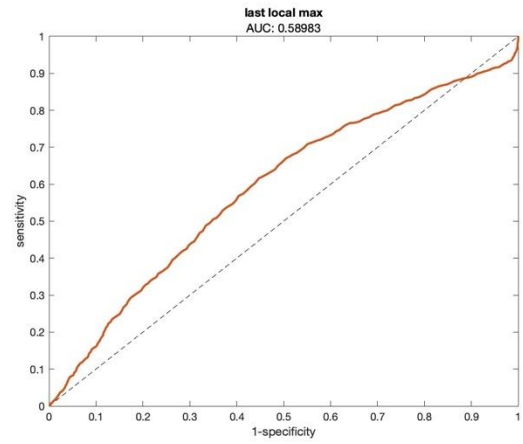
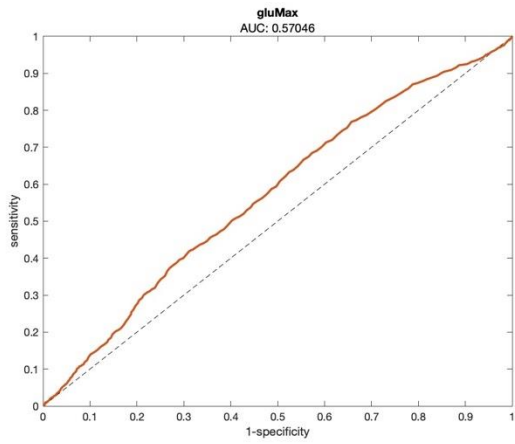
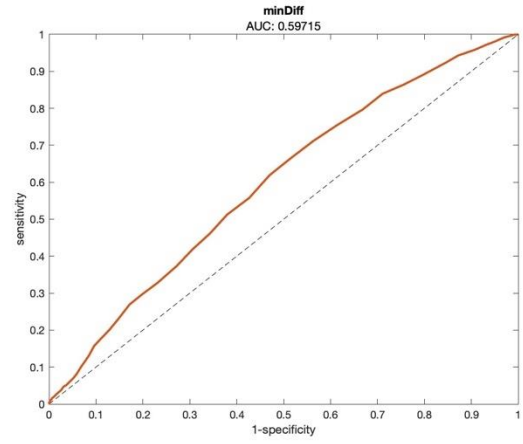
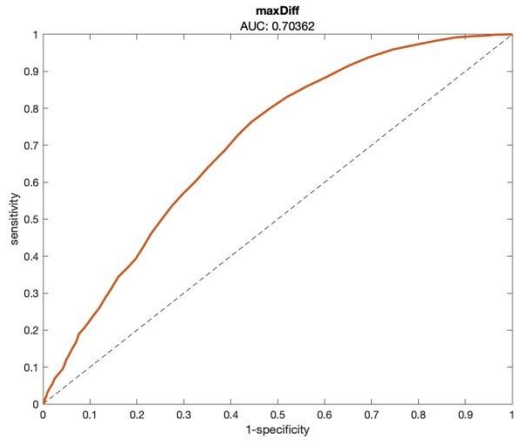
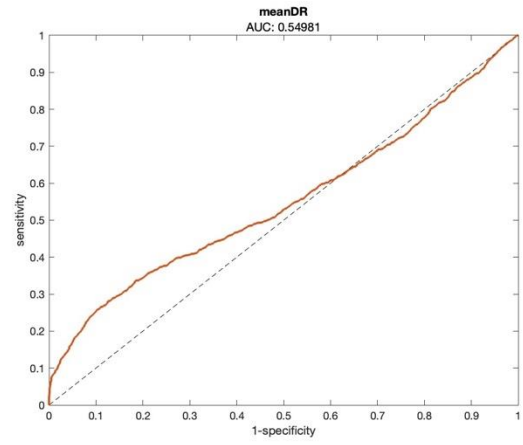
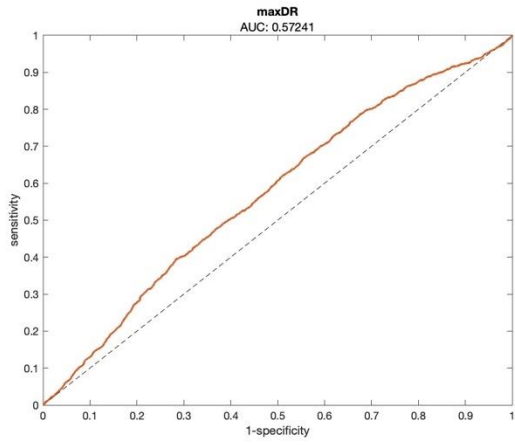
A.2 Curve ROC dei modelli univariati per l'analisi di correlazione

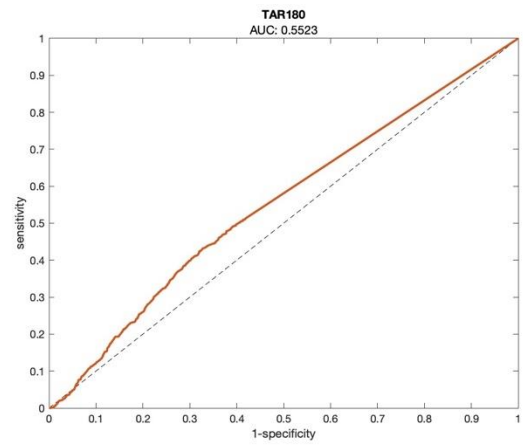
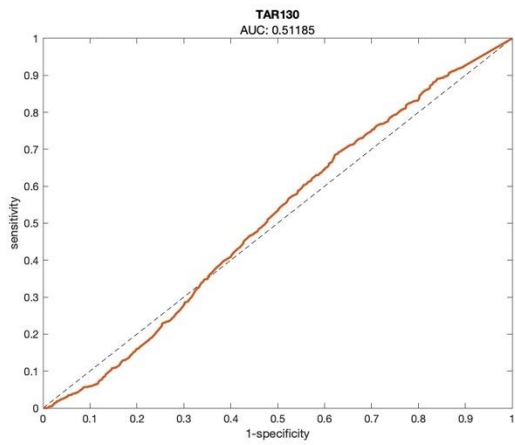
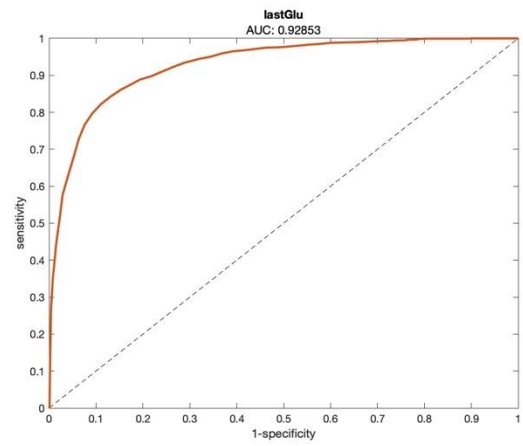
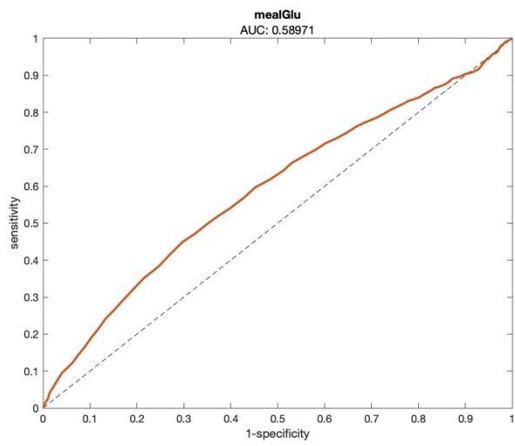
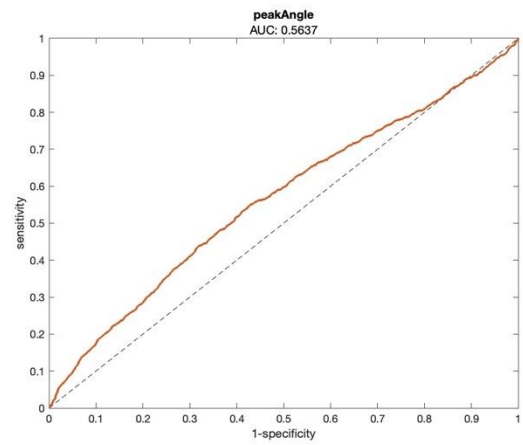
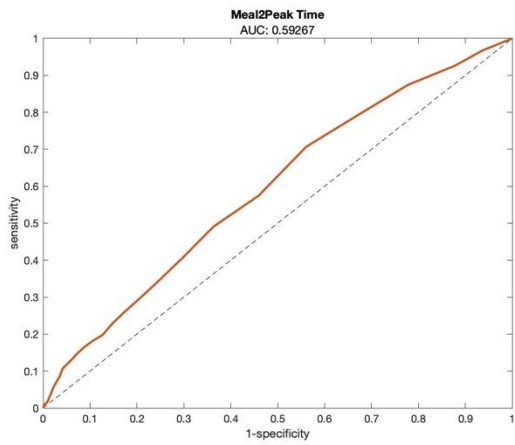
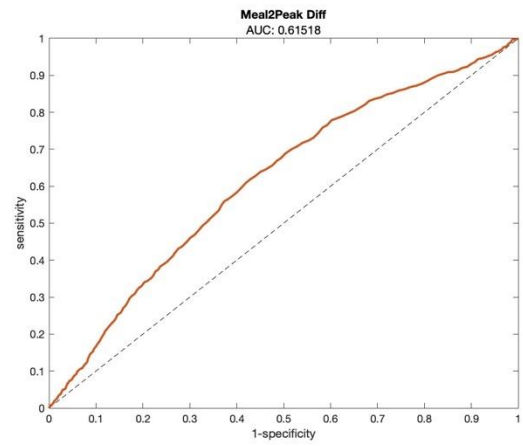
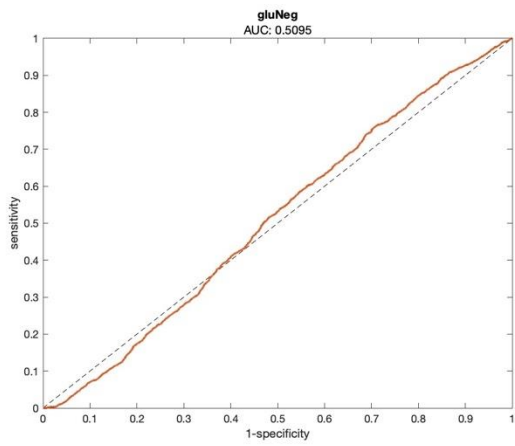
In questa sezione vengono riportate le curve ROC relative ai modelli univariati, sviluppati nel capitolo 4 per la caratterizzazione delle ipoglicemie.











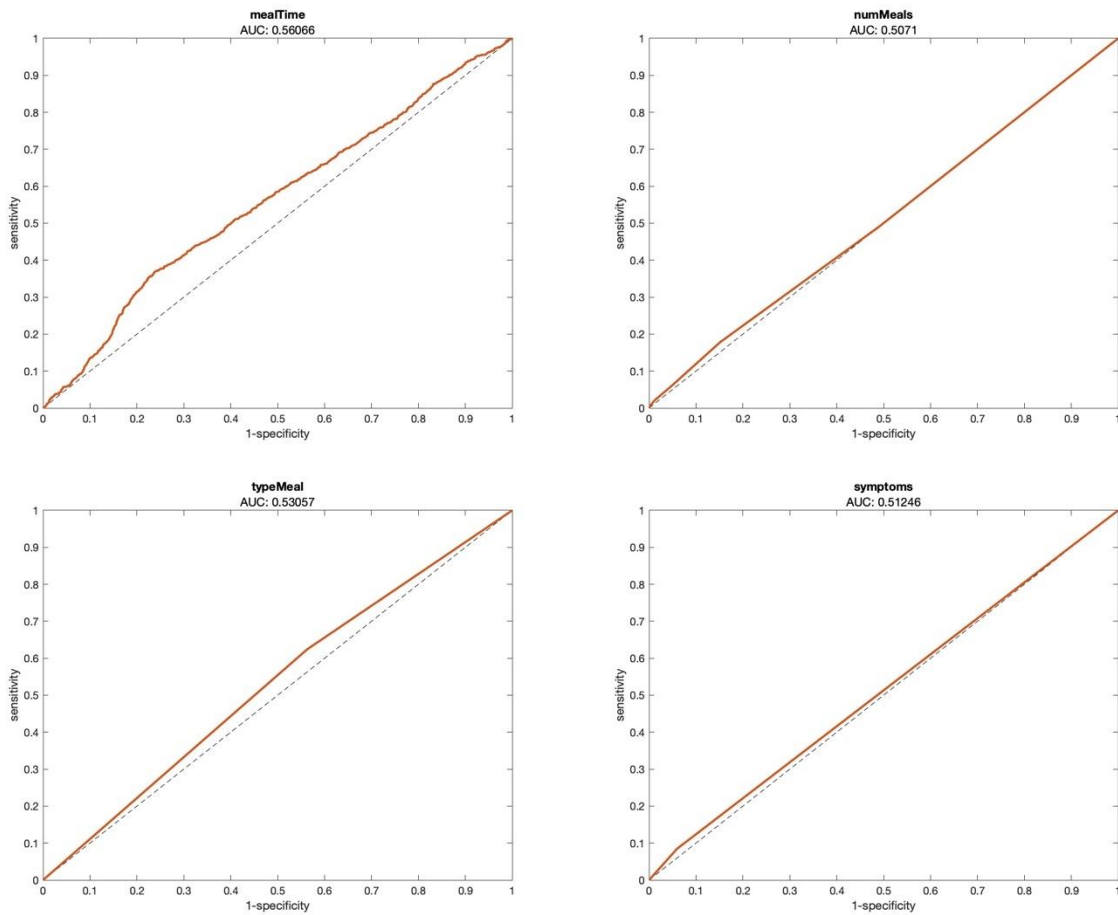
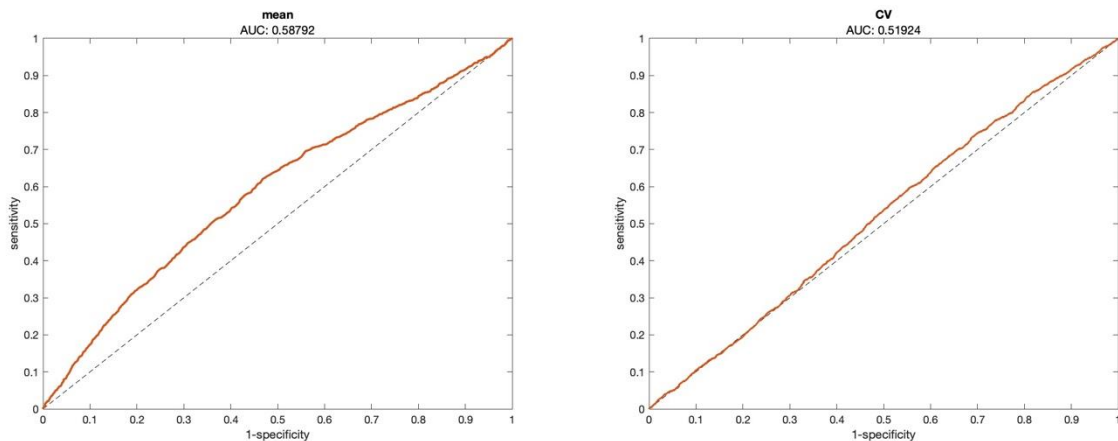


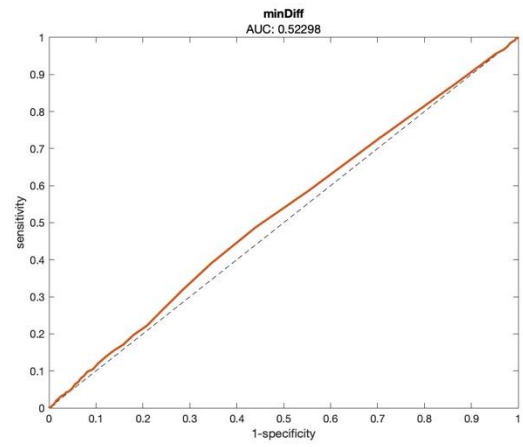
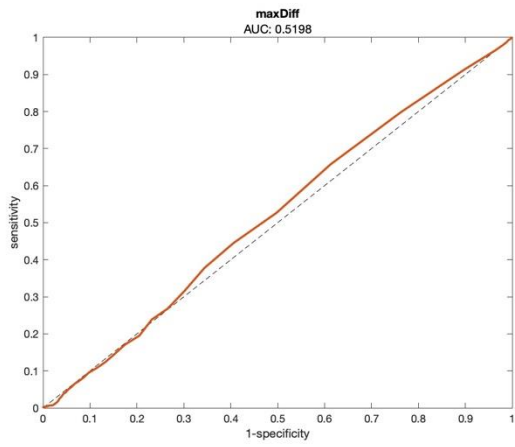
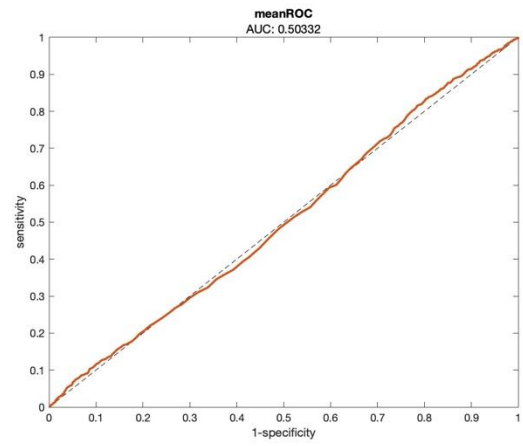
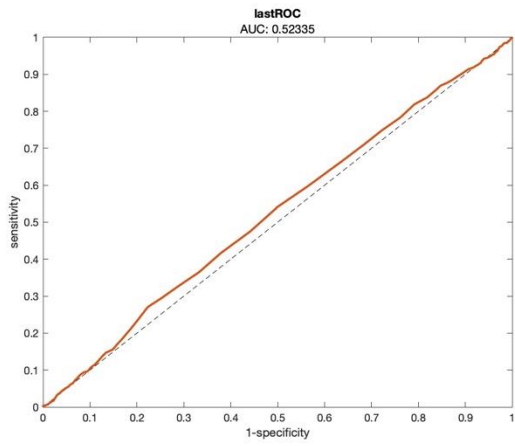
Fig. A.2 ROC curve con AUC dei modelli univariati.

A.3 Curve ROC dei modelli univariati per le feature locali

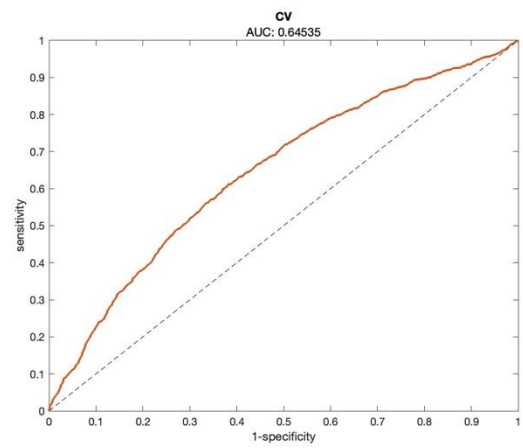
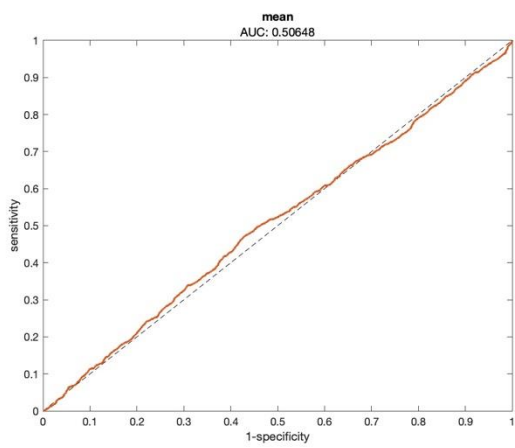
In quest'ultima sezione vengono riportate le curve ROC con i rispettivi valori di AUC per i modelli univariati delle feature locali legate a specifici periodi del segnale, considerate nell'analisi di predizione delle ipoglicemie.

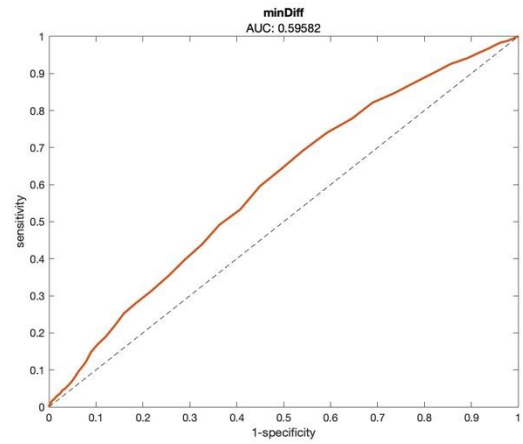
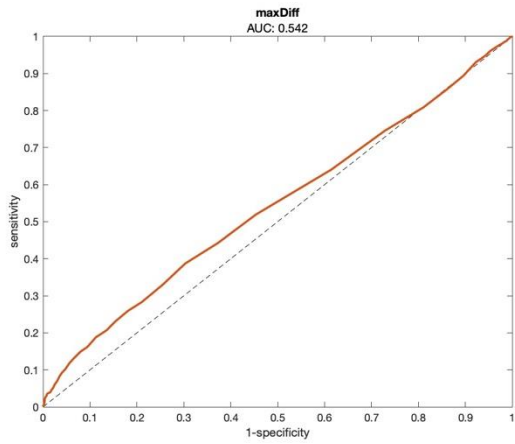
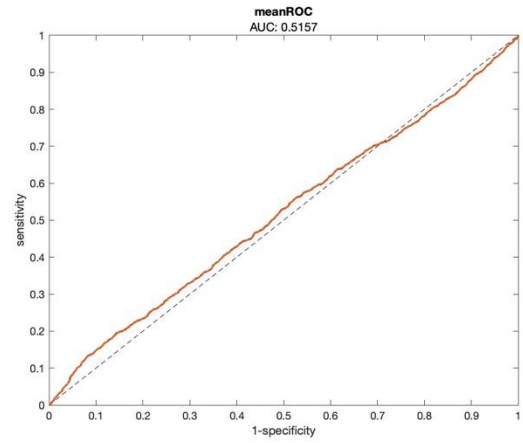
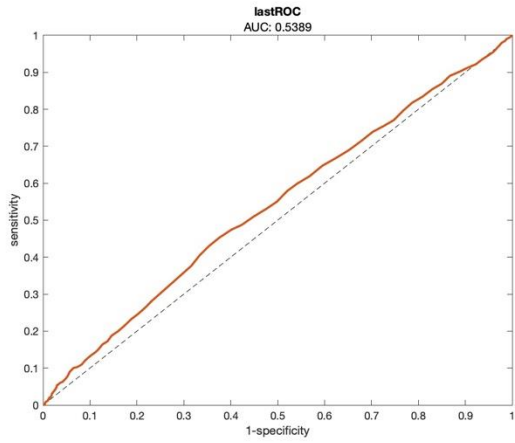
Periodo pre-prandiale



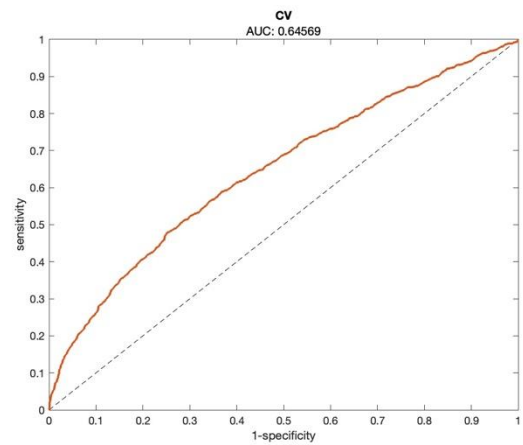
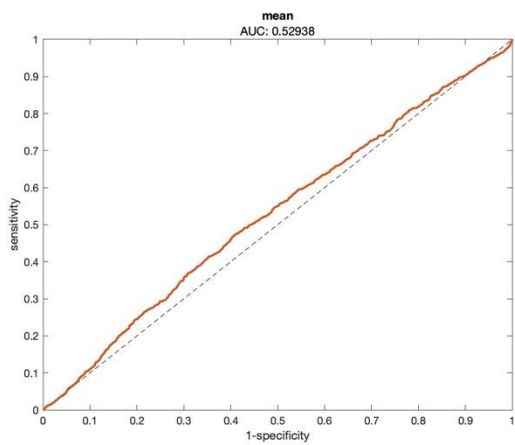


Periodo dal pasto al picco





Periodo dal picco a 20 minuti dopo



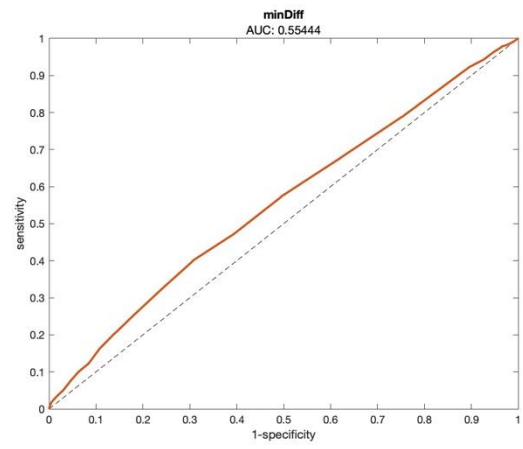
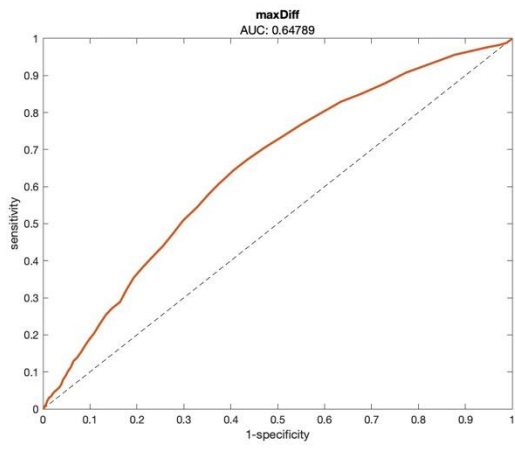
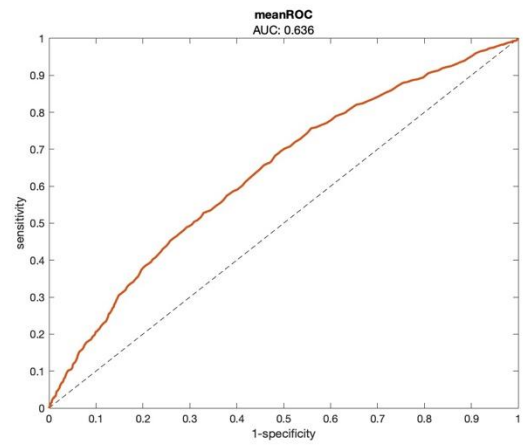
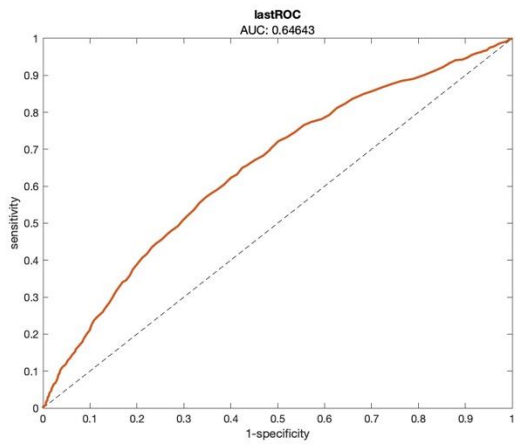


Fig. A.3 ROC curve con AUC dei modelli univariati per le feature locali.

Bibliografia

- [1] World Health Organization. Regional Office for Europe, *WHO European Regional Obesity : Report 2022*.
- [2] Cheryl D. Fryar, Margaret D. Carroll, and Joseph Afful, “Prevalence of Overweight, Obesity, and Severe Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2017–2018,” *NATIONAL CENTER FOR HEALTH STATISTICS - Health E-Stats*. American Medical Association, Sep. 22, 2020. doi: 10.1001/jama.2020.14590.
- [3] J. L. Hinkle and K. H. Cheever, *Brunner-Suddarth, Infermieristica medico-chirurgica*, Fifth., vol. 1. Casa editrice Ambrosiana, 2017.
- [4] S.I.C.OB. - Società italiana di Chirurgia dell’Obesità e delle malattie metaboliche, “Linee Guida di chirurgia dell’obesità,” 2016.
- [5] Azienda ospedaliera di Padova, “Scheda informativa Intervento chirurgia bariatrica e metabolica.”
- [6] A. B. Goldfine and M. E. Patti, “How common is hypoglycemia after gastric bypass?,” *Obesity*, vol. 24, no. 6. Blackwell Publishing Inc., pp. 1210–1211, Jun. 01, 2016. doi: 10.1002/oby.21520.
- [7] T. Battelino *et al.*, “Continuous glucose monitoring and metrics for clinical trials: an international consensus statement,” *The Lancet Diabetes and Endocrinology*, vol. 11, no. 1. Elsevier Ltd, pp. 42–57, Jan. 01, 2023. doi: 10.1016/S2213-8587(22)00319-9.
- [8] M. Emous *et al.*, “Prevalence of hypoglycaemia in a random population after roux-en-y gastric bypass after a meal test,” *Endocr Connect*, vol. 8, no. 7, pp. 969–978, Jul. 2019, doi: 10.1530/EC-19-0268.
- [9] R. Kefurt, F. B. Langer, K. Schindler, S. Shakeri-Leidenmühler, B. Ludvik, and G. Prager, “Hypoglycemia after Roux-En-Y gastric bypass: Detection rates of continuous glucose monitoring (CGM) versus mixed meal test,” *Surgery for Obesity and Related Diseases*, vol. 11, no. 3, pp. 564–569, May 2015, doi: 10.1016/j.soard.2014.11.003.
- [10] F. Halperin, M. E. Patti, M. Skow, M. Bajwa, and A. B. Goldfine, “Continuous glucose monitoring for evaluation of glycemic excursions after gastric bypass,” *J Obes*, vol. 2011, 2011, doi: 10.1155/2011/869536.
- [11] E. Suhl, S. E. Anderson-Haynes, C. Mulla, and M. E. Patti, “Medical nutrition therapy for post-bariatric hypoglycemia: practical insights,” *Surgery for Obesity and Related Diseases*, vol. 13, no. 5, pp. 888–896, May 2017, doi: 10.1016/j.soard.2017.01.025.
- [12] K. A. Schönenberger *et al.*, “Digital Solutions to Diagnose and Manage Postbariatric Hypoglycemia,” *Frontiers in Nutrition*, vol. 9. Frontiers Media S.A., Apr. 07, 2022. doi: 10.3389/fnut.2022.855223.
- [13] M. Salehi, A. Vella, T. McLaughlin, and M. E. Patti, “Hypoglycemia after gastric bypass surgery: Current concepts and controversies,” *Journal of Clinical Endocrinology and Metabolism*, vol. 103, no. 8. Oxford University Press, pp. 2815–2826, 2018. doi: 10.1210/jc.2018-00528.
- [14] G. Cappon, G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, “Wearable continuous glucose monitoring sensors: A revolution in diabetes treatment,” *Electronics (Switzerland)*, vol. 6, no. 3, Sep. 2017, doi: 10.3390/electronics6030065.
- [15] A. J. Laguna Sanz *et al.*, “Design and Clinical Evaluation of a Novel Low-Glucose Prediction Algorithm with Mini-Dose Stable Glucagon Delivery in Post-Bariatric Hypoglycemia,” *Diabetes Technol Ther*, vol. 20, no. 2, pp. 127–139, Feb. 2018, doi: 10.1089/dia.2017.0298.
- [16] F. Prendin, G. Cappon, A. Tripyla, D. Herzig, L. Bally, and A. Facchinetti, “Forecasting postbariatric hypoglycaemia in patients after Roux-en-Y gastric bypass using model-based algorithms fed by continuous glucose monitoring data: A proof-of-concept study,” *Diabetes*,

- Obesity and Metabolism*, vol. 24, no. 10. John Wiley and Sons Inc, pp. 2061–2065, Oct. 01, 2022. doi: 10.1111/dom.14783.
- [17] G. Cappon, L. Cossu, F. Boscari, D. Bruttomesso, G. Sparacino, and A. Facchinetti, “An Integrated Mobile Platform for Automated Data Collection and Real-Time Patient Monitoring in Diabetes Clinical Trials,” *J Diabetes Sci Technol*, vol. 16, no. 6, pp. 1555–1559, Nov. 2022, doi: 10.1177/19322968211024620.
- [18] G. Cappon, G. Sparacino, and A. Facchinetti, “AGATA: A Toolbox for Automated Glucose Data Analysis,” *J Diabetes Sci Technol*, 2023, doi: 10.1177/19322968221147570.
- [19] J. M. Wójcicki, “‘J’-Index. A New Proposition of the Assessment of Current Glucose Control in Diabetic Patients,” *Hormone and Metabolic Research*, 1995.
- [20] F. John Service, G. D. Molnar, J. W. Rosevear, E. Ackerman, L. C. Gatewood, and W. F. Taylor, “Mean Amplitude of Glycemic Excursions, a Measure of Diabetic Instability,” *Diabetes*, vol. 19, no. 9, 1970, [Online]. Available: <http://diabetesjournals.org/diabetes/article-pdf/19/9/644/345631/19-9-644.pdf>
- [21] P. A. Baghurst, “Calculating the mean amplitude of glycemic excursion from continuous glucose monitoring data: An automated algorithm,” *Diabetes Technol Ther*, vol. 13, no. 3, pp. 296–302, Mar. 2011, doi: 10.1089/dia.2010.0090.
- [22] C. R. Marling, J. H. Shubrook, S. J. Vernier, M. T. Wiley, and F. L. Schwartz, “Characterizing Blood Glucose Variability Using New Metrics with Continuous Glucose Monitoring Data,” *J Diabetes Sci Technol*, vol. 5, no. 4, 2011, [Online]. Available: www.journalofdst.org
- [23] W. Clarke and B. Kovatchev, “Statistical Tools to Analyze Continuous Glucose Monitor Data,” *Diabetes Technol Ther*, vol. 11, 2009.
- [24] B. P. Kovatchev, E. Otto, D. Cox, L. Gonder-Frederick, and W. Clarke, “Evaluation of a new measure of blood glucose variability in diabetes,” *Diabetes Care*, vol. 29, no. 11, pp. 2433–2438, Nov. 2006, doi: 10.2337/dc06-1085.
- [25] S. Guerra, G. Sparacino, A. Facchinetti, M. Schiavon, C. D. Man, and C. Cobelli, “A dynamic risk measure from continuous glucose monitoring data,” *Diabetes Technol Ther*, vol. 13, no. 8, pp. 843–852, Aug. 2011, doi: 10.1089/dia.2011.0006.
- [26] G. (Gareth M. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning : with applications in R*, 2013th ed. 2013.