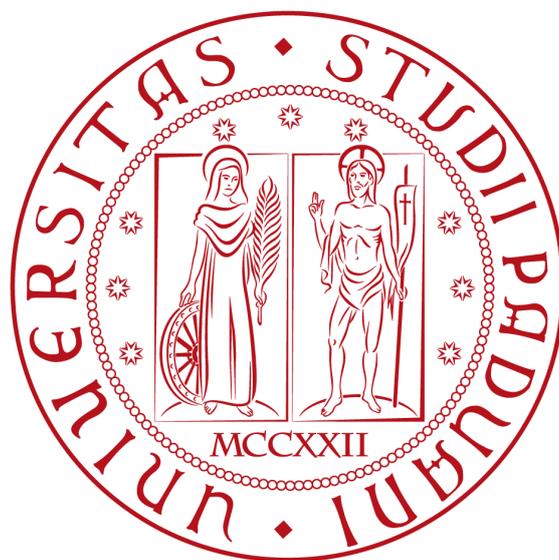


UNIVERSITÀ DEGLI STUDI DI PADOVA



TESI MAGISTRALE IN BIOINGEGNERIA

**VALUTAZIONE DI NUOVE RETI NEURALI PER LA
PREDIZIONE DELLA GLICEMIA FUTURA MEDIANTE
SENSORI CGM E MODELLI DELL'ASSORBIMENTO
DEL GLUCOSIO DOPO PASTO**

Laureando: **Luca CHERUBIN**

Relatore: **Prof. Giovanni SPARACINO**

Correlatore: **Ing. Chiara ZECCHIN**

ANNO ACCADEMICO 2011-2012

Indice

| | | |
|----------|------------------------------------------------------------------------------------------------|-----------|
| 1 | Il diabete e il monitoraggio continuo della glicemia | 9 |
| 1.1 | Classificazione del diabete | 9 |
| 1.1.1 | Diabete di tipo 1 | 9 |
| 1.1.2 | Diabete di tipo 2 | 10 |
| 1.1.3 | Altri tipi “specifici” | 10 |
| 1.1.4 | Diabete gestionale | 10 |
| 1.2 | Diagnosi del diabete | 10 |
| 1.3 | Complicanze del diabete | 11 |
| 1.4 | L’automonitoraggio del diabete | 11 |
| 1.5 | I sensori CGM minimamente invasivi | 13 |
| 1.5.1 | Il principio della glucosio ossidasi | 13 |
| 1.5.2 | Principali sensori basati sulla glucosio-ossidasi | 14 |
| 1.5.3 | Il dispositivo CGM Dexcom® 7-Plus | 14 |
| 1.6 | Dispositivi non invasivi | 15 |
| 1.7 | Aspetti di forza e critici dei sensori CGM | 16 |
| 1.8 | Uso dei sensori CGM per la predizione della glicemia futura | 17 |
| 1.8.1 | Prevenzione di episodi ipo/iperglicemici mediante l’utilizzo di algoritmi predittivi | 17 |
| 1.8.2 | Predizione: modello polinomiale e modello autoregressivo | 18 |
| 1.8.3 | Predizione con reti neurali: stato dell’arte | 18 |
| 1.9 | Scopo della tesi e sua articolazione | 19 |
| 2 | Introduzione alle reti neurali | 21 |
| 2.1 | Reti neurali monostrato | 22 |
| 2.1.1 | Il neurone biologico | 22 |
| 2.1.2 | Il neurone artificiale | 22 |
| 2.2 | Reti neurali multistrato | 27 |
| 2.2.1 | Multilayer perceptron | 27 |
| 2.2.2 | Jump network | 28 |
| 2.3 | Accenno ad altre tipologie di reti neurali | 28 |
| 2.3.1 | Reti neurali ricorsive | 28 |
| 2.3.2 | Reti neurali radial-basis function | 29 |
| 2.4 | Apprendimento errore-correzione: l’algoritmo di <i>back-propagation</i> | 30 |
| 2.4.1 | Apprendimento attraverso back-propagation | 30 |
| 2.5 | Uso di reti neurali nella predizione della glicemia futura | 34 |
| 2.5.1 | La rete di Perez-Gandia et al. | 34 |

| | | |
|----------|-------------------------------------------------------------------------------------------------|-----------|
| 2.5.2 | La rete di Pappada et al. | 34 |
| 2.5.3 | La rete di Zecchin et al. | 35 |
| 2.5.4 | Potenziati margini di miglioramento | 35 |
| 2.6 | Conclusioni | 36 |
| 3 | Approcci alla predizione della glicemia basati su jump-net | 37 |
| 3.1 | Architetture jump testate: tipologie “A” e “B” | 37 |
| 3.2 | Possibili ingressi utilizzabili nella predizione della glicemia mediante reti neurali | 39 |
| 3.2.1 | Informazioni sulla glicemia attuale, derivata ed integrale. | 39 |
| 3.2.2 | Predizione della glicemia futura mediante modello poly(1) | 40 |
| 3.2.3 | Modello per l’assorbimento del glucosio dopo pasto: la rate of appearance | 41 |
| 3.3 | Jump-net A: modello base | 42 |
| 3.4 | Algoritmo di <i>cross-validation</i> per la scelta della struttura ottimale | 42 |
| 3.5 | Metriche di valutazione delle performance della rete | 44 |
| 3.5.1 | Definizione dell’ RMSE | 44 |
| 3.5.2 | Guadagno temporale medio | 44 |
| 3.5.3 | Energia delle derivate seconde: | 44 |
| 3.5.4 | Regolarità del profilo predetto: J | 45 |
| 3.6 | Le reti candidate valutate | 45 |
| 3.7 | Da jump-net A a jump-net B: un possibile metodo di scelta | 49 |
| 3.8 | Le jump-net candidate | 50 |
| 3.9 | Conclusioni | 51 |
| 4 | Data Sets | 53 |
| 4.1 | Creazione del data-set simulato | 53 |
| 4.2 | Data Sets reali | 54 |
| 4.2.1 | Data Set reale Dexcom® 7-Plus (campionamento 5 min) | 54 |
| 4.2.2 | Data Set reale Abbott Navigator™ (campionamento 1 min) | 54 |
| 4.3 | Conclusioni | 55 |
| 5 | Risultati su data set simulato | 57 |
| 5.1 | Predizione del profilo glicemico | 57 |
| 5.2 | Reti jump di tipologia B | 63 |
| 5.2.1 | Strutturazione delle 3 reti candidate | 63 |
| 5.2.2 | Risultati | 65 |
| 5.3 | Considerazioni riassuntive | 67 |
| 6 | Risultati su data set reale Dexcom® 7-Plus | 71 |
| 6.1 | Predizione del profilo glicemico | 71 |
| 6.2 | Reti jump di tipologia B | 76 |
| 6.2.1 | Strutturazione delle 2 reti candidate | 76 |
| 6.2.2 | Risultati | 78 |
| 6.3 | Considerazioni riassuntive | 81 |

| | | |
|----------|------------------------------------------------------|-----------|
| 7 | Risultati su data set reale Abbott Navigator™ | 83 |
| 7.1 | Predizione del profilo glicemico | 83 |
| 7.2 | Reti jump di tipologia B | 88 |
| 7.2.1 | Strutturazione delle 3 reti candidate | 88 |
| 7.2.2 | Risultati | 88 |
| 7.3 | Considerazioni riassuntive | 92 |
| 8 | Conclusioni e sviluppi futuri | 93 |
| 8.1 | Riassunto del lavoro svolto | 93 |
| 8.2 | Bilancio dei risultati | 94 |
| 8.3 | Possibili sviluppi futuri | 94 |
| 8.4 | Considerazioni finali | 95 |

Sommario

Il diabete mellito è una patologia largamente diffusa a livello mondiale, la cui terapia prevede azioni decise sulla base dell'auto-monitoraggio della glicemia. Tale monitoraggio è tradizionalmente effettuato attraverso dispositivi pungi-dito, ma dagli anni 2000 sono stati proposti dei dispositivi di misura quasi-continua noti come sensori per il Continuous Glucose Monitoring (CGM). L'avvento di tali strumenti ha stimolato la ricerca e lo sviluppo di tecniche per la predizione a breve termine della glicemia futura (30-60 minuti), con lo scopo di prevenire gli episodi di rischio per il paziente durante il trattamento della patologia diabetica.

Sono diversi in letteratura i lavori relativi alla possibilità di predire la glicemia sfruttando modelli matematici. Negli ultimi due anni, in particolare, è stato proposto l'uso delle reti neurali, soprattutto per la loro capacità d'integrare ai dati glicemici informazioni relative a domini differenti, come per esempio quelle sui pasti. Il metodo attualmente stato dell'arte (Zecchin et al, 2012), specificatamente, combina due approcci diversi per modellare in maniera ottimale le dinamiche lineari e non lineari del segnale glicemico ed è per questo indicato come NN-Linear Prediction Algorithm (NN-LPA).

In questa tesi verrà proposta una nuova tipologia di reti neurali, note come reti jump, la cui peculiarità ci consentirà di gestire per mezzo di un modello unico entrambe le dinamiche glicemiche, combinandole con dati esterni come quelli relativi ai pasti. Nella tesi verranno inizialmente presentate tre reti jump, di complessità crescente, la cui ottimizzazione in termini di struttura e numero/tipologia degli ingressi verrà svolta sfruttando un data set simulato. Le performance di queste reti verranno poi valutate su tre diversi data sets, uno simulato (vedi prima), uno reale ottenuto dal monitoraggio di 9 pazienti a mezzo di sensore Dexcom® 7-Plus e un altro ottenuto da 25 pazienti reali differenti, monitorati con Abbott Freestyle Navigator™.

Per ciascuna di queste reti verranno presentate delle reti alternative caratterizzate dall'assenza di alcune delle connessioni sinaptiche, reti che verranno ugualmente testate sui tre data set a disposizione. I risultati, infine, verranno confrontati con quelli di due predittori stato dell'arte, poly(1) (Sparacino et al. 2007 [21]) e NN-LPA (Zecchin et al. 2012 [24]), al fine di individuare pregi e difetti delle nuove architetture. Le reti jump presentate si dimostreranno in grado di gestire le dinamiche lineari e non lineari del segnale glicemico e di integrare le informazioni derivate dai pasti, grazie ad un unico modello di complessità inferiore, capace tuttavia di produrre dei risultati che competono o superano quelli dei modelli stato dell'arte considerati.

Questa tesi dimostrerà le potenzialità delle reti jump che, mantenendo le performance di modelli più complessi, sono in prospettiva meglio implementabili su una soluzione *on-chip* portabile da pazienti affetti da diabete.

Capitolo 1

Il diabete e il monitoraggio continuo della glicemia

Il diabete mellito è una patologia cronica che colpisce al mondo più di 346 milioni di persone. Solo nel 2004 è stato stimato che circa 3.4 milioni di individui sono morti a causa di essa e l'OMS (Organizzazione Mondiale della Sanità) prevede che le morti raddoppieranno nel periodo compreso tra il 2005 e il 2030 [1]. Dei soggetti affetti da questa patologia, circa il 90% soffre della forma di diabete nota come tipo 2, mentre il restante 10% soffre di diabete di tipo 1. Questi dati tuttavia rappresentano una sottostima del fenomeno poiché si stima che vi sia almeno un paziente diabetico non diagnosticato per ogni due noti [10]. In Italia i dati riportati dall'annuario statistico Istat (Istituto Nazionale di Statistica) 2010 indicano come sia diabetico il 4,9% degli italiani (5,2% delle donne e 4,5% degli uomini), con cifre che sfiorano i 3 milioni di persone.

Il diabete mellito insorge qualora il pancreas non sia in grado di produrre sufficiente insulina (tipo 1) o l'organismo non utilizzi al meglio l'insulina prodotta (tipo 2). L'insulina è un ormone secreto dalle β -cellule pancreatiche, che regola l'assorbimento e l'utilizzo del glucosio da parte delle cellule dell'organismo. Il glucosio, derivato dalla trasformazione degli amidi durante la fase digestiva, è la principale fonte d'energia dell'organismo, nonché l'unica per il cervello. Quando mangiamo, il pancreas produce una quantità d'insulina tale da permettere l'ingresso del glucosio nelle cellule. La sua mancanza o una riduzione della sua funzionalità porta ad un aumento di glucosio nel sangue oltre i livelli fisiologici (120 – 140 mg/dl). Il glucosio in eccesso viene dunque espulso attraverso le urine, che assumono un distintivo sapore dolciastro.

1.1 Classificazione del diabete

Nel 1997, un comitato internazionale di esperti dell'ADA (American Diabetes Association) e dell'OMS ha redatto un comunicato contenente le nuove raccomandazioni per la classificazione e la diagnosi del diabete mellito. Il nuovo sistema comprende quattro tipologie di diabete mellito: tipo 1, tipo 2, altri "specifici" e diabete gestazionale. Vediamo più nello specifico le varie classi di questa patologia [2].

1.1.1 Diabete di tipo 1

Il diabete di tipo 1 (noto anche come insulino-dipendente o giovanile) è caratterizzato dall'incapacità del pancreas di produrre insulina a causa di una progressiva distruzione delle β -cellule. Sebbene l'origine di tale degenerazione non sia ancora nota, essa può essere associata a diverse cause, come forme non

comuni di immuno-deficienza, infezioni (Rubella Congenita, Citomegalovirus) o altre sindromi genetiche (sindrome Down, di Wolfram e altre). Questa forma della malattia colpisce il 10% dei soggetti diabetici e in genere insorge durante l'infanzia o l'adolescenza, sebbene si presenti anche in soggetti in età avanzata. Tra i sintomi che la caratterizzano si hanno l'eccessiva secrezione urinaria (poliuria), sete (polidipsia), senso d'appetito, perdita di peso, stanchezza profonda e alterazioni della vista. In questa tipologia sono necessarie dosi giornaliere di insulina allo scopo di raggiungere e mantenere un livello glicemico prossimo a quello fisiologico, compensando così il deficit pancreatico.

1.1.2 Diabete di tipo 2

Il diabete di tipo 2 (noto anche come insulino-indipendente o adulto) è caratterizzato da insulino resistenza nei tessuti periferici e da un difetto di secrezione d'insulina delle β -cellule. Questa forma colpisce circa il 90% della popolazione diabetica ed è strettamente associata con una storia familiare di diabete, età avanzata, obesità e scarso esercizio fisico. Il diabete di tipo 2, a differenza del tipo 1, presenta una sintomatologia più sfumata che non consente quindi una diagnosi rapida. La glicemia risulta quindi elevata ma senza i segnali clinici del diabete di tipo 1.

1.1.3 Altri tipi “specifici”

Questo gruppo include tutte le persone con difetti genetici della funzione β -cellulare, noto anche come MODY (diabete dell'età adulta ad esordio giovanile) o con difetti dell'azione insulinica. Afferiscono a questo gruppo anche persone con disturbi del pancreas esocrino (pancreatiti o fibrosi cistiche), con disfunzioni associate ad altre endocrinopatie e persone con disfunzione pancreatica di origine farmacologica, chimica o infettiva.

1.1.4 Diabete gestionale

Il diabete gestionale, a differenza delle altre classi, è una classificazione operativa anziché una condizione fisiopatologica. Identifica le donne che sviluppano il diabete mellito durante la gestazione. Molte donne classificate con diabete gestazionale hanno una normale omeostasi del glucosio durante la prima metà della gravidanza e sviluppano una relativa insulinodeficienza durante l'ultima metà della gestazione, sfociante in seguito in iperglicemia. L'iperglicemia si risolve in molte donne subito dopo il parto ma le pone tra le persone ad alto rischio di sviluppare diabete di tipo 2 nel corso della vita.

1.2 Diagnosi del diabete

I criteri raccomandati per la diagnosi del diabete prevedono la positività a ciascuno dei seguenti test effettuati in giorni differenti:

- Sintomi tipici del diabete mellito (poliuria, polidipsia o inesplicabile perdita di peso) accompagnati da misure casuali¹ di:
 - concentrazione di glucosio nel plasma ≥ 200 mg /dl (11.1 mmol/l);
 - FPG² ≥ 126 mg/dl (7.0 mmol/l);

¹E' definita casuale ogni ora del giorno senza riguardo alla distanza dall'ultimo pasto

²FPG: glicemia a digiuno, cioè senza alcun apporto colorico nelle ultime 8 ore

- 2hrPPG³ \geq 200 mg per dL (11.1 mmol per L) dopo un carico di 75 g di glucosio.
- Alterata omeostasi del glucosio
 - Alterata glicemia a digiuno: FPG da 110 a 126 mg/dl (da 6.1 a 7.0 mmol/l);
 - Alterata tolleranza del glucosio: 2hrPPG da 140 a 200 mg/dl (da 7.75 a 11.1 mmol/l).

Viene considerato sano un soggetto che presenti una glicemia a digiuno inferiore ai 110 mg/dl (6.1 mmol/l) o equivalentemente inferiore a 140 mg/dl (7.75 mmol/l) due ore dopo i pasti.

1.3 Complicanze del diabete

Il diabete come patologia presenta delle complicanze che possono presentarsi sia sul breve periodo che sul lungo termine. Le prime sono principalmente dovute ad eventi ipoglicemici, quando la quantità di glucosio nel sangue scende sotto la soglia dei 70 mg/dl, mentre le seconde sono dovute al livello di glucosio superiore ai 180 mg/dl, evento noto come iperglicemia.

La complicanza più grave dovuta agli episodi ipoglicemici è il coma diabetico o chetoacidotico, dovuto all'acidificazione del sangue a seguito della produzione di corpi chetonici nel fegato per l'iperafflusso di acidi grassi, il cui metabolismo è regolato dall'insulina. Se non trattato rapidamente, il coma diabetico può portare alla morte del paziente [2].

Le complicanze croniche invece, dovute principalmente all'iperglicemia prolungata, riguardano diversi organi e tessuti come i reni, gli occhi, cuore, vasi sanguigni e nervi periferici. La patologia nota come retinopatia diabetica è legata ad un danno ai vasi che irrorano la retina che può portare alla cecità del soggetto. I pazienti diabetici presentano un rischio di contrarre malattie cardiovascolari che è dalle 2 alle 4 volte superiore rispetto alla popolazione sana e tali complicanze rappresentano il 50% delle morti per diabete. La neuropatia diabetica è invece legata ad una degenerazione dei nervi e conseguentemente del sistema nervoso periferico, con conseguenze come la perdita di sensibilità, dolore e danni agli arti, con necessità d'amputazione nei casi più gravi. Questa complicazione si manifesta per esempio nella patologia nota come piede diabetico in cui l'insorgere di gravi ulcere negli arti inferiori possono portare, a seguito di un'infezione, all'amputazione dell'arto stesso. Infine i danni a livello renale possono portare alla nefropatia diabetica in cui il rene, a causa di una iperfiltrazione glomerulare, non è più in grado di svolgere il suo ruolo di riassorbimento e filtraggio [2].

Il diabete risulta quindi una patologia con conseguenze socio-economiche importanti soprattutto per le complicanze altamente invalidanti ad esso associate. E' chiaro dunque come la diagnosi e la cura di questa malattia richiedano un impegno finanziario tale da porre il diabete tra le prime voci di spesa del sistema sanitario nazionale[10].

1.4 L'automonitoraggio del diabete

Uno degli obiettivi principali nella gestione del diabete, è quello di mantenere la glicemia il più possibile all'interno dell'intervallo di normalità nell'arco dell'intera giornata. Per raggiungere questo fondamentale obiettivo, è necessario eseguire controlli dei livelli di glucosio nel sangue, attraverso una serie di operazioni quotidiane che vanno sotto il nome di automonitoraggio.

Se fino agli anni '50 non era possibile raccogliere informazioni giornaliere durante la terapia, è con gli anni '80 che un notevole passo avanti è stato fatto grazie allo sviluppo di tecnologie e dispositivi innovativi

³2hrPPG: glicemia due ore dopo il pasto

che consentivano al paziente di monitorare più volte durante il giorno il proprio livello di glucosio nel sangue (*Self-Monitoring Blood Glucose*, SMBG). Attraverso l'ausilio di piccoli apparecchi elettronici, noti come glucometri, i pazienti hanno la possibilità di monitorare ed analizzare in breve tempo i propri valori glicemici. In pratica, si preleva una goccia di sangue da un polpastrello effettuando una puntura con appositi punge dito (Figura 1.1); dopodiché la goccia di sangue viene posta sulla striscia reattiva (o sul sensore) del glucometro che, in pochi secondi, riporterà il valore della glicemia. La frequenza con cui i pazienti dovrebbero effettuare un automonitoraggio varia da persona a persona, anche se molti esperti sostengono che sarebbe opportuno effettuare almeno 3-4 misurazioni al giorno, prima dei pasti e prima di coricarsi. In base ai valori così misurati è possibile regolare l'assunzione di farmaci ipoglicemizzanti e le dosi d'insulina da iniettare.



Figura 1.1: Glucometro per le misure SMBG

Sebbene questi strumenti abbiano portato grandi vantaggi nella terapia del diabete, presentano ancora delle forti limitazioni. In primo luogo, il prelievo del sangue attraverso lo strumento punge-dito risulta una pratica invasiva, sebbene in maniera limitata. Il limite maggiore deriva però dalla scarsa frequenza delle misure ottenute per mezzo di SMBG. Sebbene infatti la misura di glicemia riportata risulti affidabile, la sporadicità dei dati non favorisce il riconoscimento di molti eventi iper o ipoglicemici. I dati inoltre non chiariscono se la glicemia misurata sia stabile o stia aumentando piuttosto che diminuendo. Dalla Figura 1.2 si può ben vedere come le misurazioni SMBG (pallini rossi) non siano in grado di esprimere la variabilità del profilo glicemico (linea blu) ottenibile da misure quasi in continua attraverso un dispositivo per il Continuous Glucose Monitoring (CGM). Il segnale può non rimanere stabile tra due misurazioni SMBG e vi sono delle escursioni (cerchiate in rosso) nelle zone di ipoglicemia e iperglicemia.

Appare dunque chiaro come monitorare la glicemia in maniera pseudo-continua possa fornire grande supporto alla terapia del diabete. Fino al 2000 tuttavia ciò non era possibile se non in ambito ospedaliero attraverso campionamento frequente in vena (per esempio ogni 15 minuti). Tuttavia l'invasività del trattamento lo rende inadatto al monitoraggio continuo nella vita quotidiana, e sfruttabile solo per un riaggiustamento della terapia in modo retrospettivo

A partire dal 2000 sono state sviluppate nuove tecniche e dispositivi, noti come dispositivi CGM, che consentono di misurare in maniera quasi continua la glicemia nell'arco della giornata, riducendo contemporaneamente al minimo l'invasività della procedura. È evidente che la possibilità di avere a disposizione un segnale continuo porta numerosi vantaggi. In primo luogo permette un'analisi retrospettiva dei dati, dando quindi modo di comprendere meglio le dinamiche glicemiche e migliorare la terapia soggetto-specifica. La tecnologia CGM fornisce, in secondo luogo, gli strumenti per riconoscere in tempo reale eventi potenzialmente rischiosi e generare allarmi conseguenti. Infine, la quantità di dati raccolti in maniera retrospettiva dà la possibilità di predire l'insorgere di episodi ipo/iperglicemici e di generare allarmi con un certo anticipo. Questo sarà oggetto della tesi, come si vedrà tra poco alla sezione **1.9**.

SMBG vs CGM

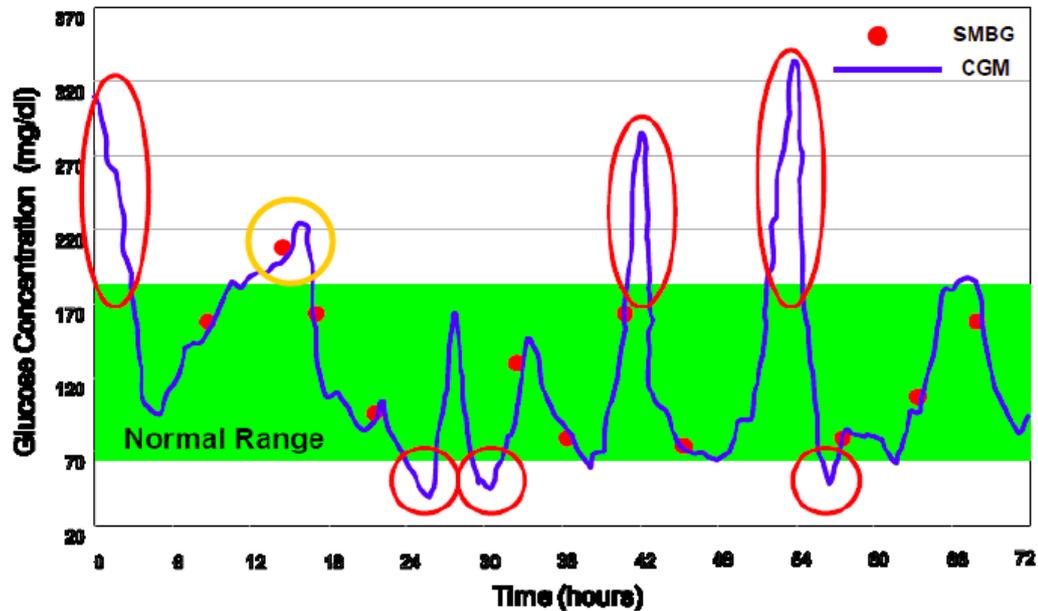


Figura 1.2: Confronto tra misurazioni SMBG (pallini rossi) e misurazioni CGM (linea continua blu). Dalla serie temporale SMBG si coglie solo una (cerchio arancione) delle escursioni (cerchi rossi) al di fuori del range di normalità (fascia verde) che invece il segnale CGM consente di individuare completamente.

1.5 I sensori CGM minimamente invasivi

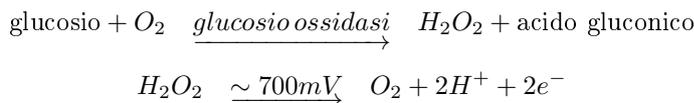
I dispositivi CGM sono in grado di effettuare una misura quasi continua della glicemia per diversi giorni consecutivi (fino ad una settimana) con frequenze anche molto elevate (ogni 1-5 minuti). Grazie alla loro tecnologia possono quindi portare molti vantaggi rispetto ai sistemi tradizionali SMBG.

I dispositivi minimamente invasivi sono così chiamati perché in grado di misurare la concentrazione di glucosio nello spazio tra le cellule e i capillari, il cosiddetto fluido interstiziale, piuttosto che direttamente nel sangue. I dispositivi CGM minimamente invasivi sono in generale costituiti da:

- un sensore impiantato nel sottocute, solitamente nel braccio o nella zona addominale
- un dispositivo di visualizzazione indossabile, come ad esempio un piccolo monitor che può essere fissato al polso
- un dispositivo tascabile di memorizzazione ed elaborazione, che può comunicare col sensore e col display via cavo o con tecnologia wireless

1.5.1 Il principio della glucosio ossidasi

La misura ottenuta dai dispositivi CGM minimamente invasivi non è generalmente di glicemia reale, ma piuttosto di un segnale grezzo (tipicamente una corrente in mV o mA) legato alla reazione enzimatica della glucosio-ossidasi, misurata per mezzo di sensori amperometrici. Questa reazione, a partire da glucosio ed ossigeno, è in grado di formare perossido di idrogeno:



Il sensore, tuttavia, non misura una corrente proporzionale alla concentrazione di glucosio nel sito dell'elettrodo, ma a quella dell'ossigeno. Affinché la reazione avvenga, infatti, è necessaria la presenza di una molecola di ossigeno per ogni molecola di glucosio. Alcune difficoltà che si riscontrano con questo tipo di misura sono la quantità di ossigeno in situ molto inferiore a quella del glucosio, rendendo di fatto l'ossigeno un reagente limitante, e la presenza di altre sostanze elettro-attive come l'acido urico o l'acido ascorbico che possono inficiare la misura amperometrica. Per questi dispositivi risulta quindi necessaria una procedura di calibrazione che permetta di convertire la misura di concentrazione dell'interstizio in una stima della concentrazione di glucosio nel plasma. Come vedremo nel Paragrafo 1.7, il processo di calibrazione è uno degli aspetti critici legati all'utilizzo di questa tecnologia.

1.5.2 Principali sensori basati sulla glucosio-ossidasi

Tra i principali dispositivi minimamente invasivi possiamo elencare:

- CGMS® System Gold™ (Medtronic MiniMed, Northridge, CA), che ha ricevuto l'approvazione della FDA nel Marzo del 2001 [4];
- Guardian® Real-Time Medtronic MiniMed, Northridge, CA), che ha ricevuto l'approvazione della FDA nel Giugno del 2005 [8];
- GlucoDay® (Menarini Diagnostic, Firenze, ITA), che ha ricevuto nel Giugno 2007 il marchio CE in Europa (non è stata fatta richiesta di approvazione alla FDA) [5];
- Seven Plus® (Dexcom, San Diego, CA), che ha ricevuto l'approvazione della FDA nel Febbraio del 2009. Questo sensore, con il quale sono stati raccolti i dati reali utilizzati in questa tesi, verrà descritto dettagliatamente nella sezione 1.5.3 [6];
- FreeStyle Navigator™ (Abbott Laboratories, Alameda, CA), che ha ricevuto l'approvazione della FDA nel Marzo del 2008 [7].

1.5.3 Il dispositivo CGM Dexcom® 7-Plus

Il dispositivo CGM Dexcom® Seven® Plus, utilizzato per la raccolta dei dati reali utilizzati in questa tesi, rappresenta la terza generazione dei sistemi per il monitoraggio continuo del glucosio per il quale la Dexcom(tm) ha ricevuto l'approvazione dall'FDA nel febbraio del 2009. Il sistema è costituito di tre componenti: un sensore, un trasmettitore ed un ricevitore, ed è in grado di fornire una misura di glucosio ogni 5 minuti per un periodo di 7 giorni consecutivi d'utilizzo del dispositivo. Il sistema è composto da:

- un sensore (Figura 1.3a) costituito da un filo flessibile miniaturizzato che viene inserito nel sottocute e fissato alla pelle per mezzo di un speciale adesivo e misura il glucosio interstiziale
- un trasmettitore (Figura 1.3b) delle dimensioni di una moneta, in grado di inviare al ricevitore attraverso tecnologia wireless la misura di glicemia ogni 5 minuti. Esso viene inserito all'interno del sensore a formare un piccolo contenitore discreto

- un ricevitore (Figura 1.3c) delle dimensioni di un cellulare che svolge le funzioni di un piccolo computer. E' in grado di mostrare sul display l'andamento della glicemia nelle ultime 24 ore, con indicazioni grafiche sulla direzione stessa del trend, e di memorizzare i dati fino a 30 giorni consecutivi.



Figura 1.3: Il sistema Dexcom® 7-Plus; (a) Il sensore; (b) Il trasmettitore; (c) Il ricevitore; tratte da [6]

Il Dexcom® 7-Plus viene anche corredato di un software (Dexcom Data Manager 3 - DM3) in grado di trasferire i dati raccolti dal sensore al computer. All'interno di questo software è possibile corredare le misure di glicemia le informazioni relative ai pasti, all'insulina, allo stato di salute e all'attività fisica, permettendo così al paziente di tenere traccia in maniera completa della propria terapia. Oltre a questo, è possibile anche aggiungere le informazioni relative ai valori dell'emoglobina glicosilata HbA1c per il monitoraggio e la visualizzazione dei miglioramenti.

Per risolvere il problema legato alla scarsità di ossigeno presente nel sito misura, questo dispositivo incorpora una membrana che funge da barriera. Tale membrana riduce la quantità di glucosio che fluisce all'interno del sito, pareggiandola a quella dell'ossigeno. Di conseguenza anche la quantità di perossido d'idrogeno si riduce, garantendo una maggiore durabilità del sensore stesso [20].

1.6 Dispositivi non invasivi

I dispositivi CGM non invasivi sono tutt'ora in fase di sviluppo e ci vorranno ancora diversi anni di ricerca per migliorare la qualità delle misure fornite. A differenza dei dispositivi minimamente invasivi che sfruttano le informazioni interstiziali per la misura della glicemia, i dispositivi non invasivi usano diverse tecniche di misura transcutanea, come l'irradiazione del tessuto, l'analisi delle radiazioni assorbite e di scatter, metodi ottici come le radiazioni infrarosse, la spettroscopia di Raifman, la polarimetria, la luce scatter e la spettroscopia fotoacustica. Di seguito viene riportato un breve elenco di dispositivi non invasivi con il loro principio di funzionamento:

- GlucoWatch G2 Biographer (Cygnus, Inc., Redwood City, CA), approvato dalla FDA nel 2001 [17]: il dispositivo utilizza una carica elettrica per espellere il sodio e il cloro all'esterno della pelle del paziente; il glucosio viene anch'esso espulso passivamente con l'acqua di idratazione dei sali. La soluzione così estratta dal paziente viene ossidata e ne viene misurato il contenuto di glucosio. Tale prodotto è stato però ritirato dal mercato nel 2006 a causa dei numerosi casi di irritazioni cutanee causate dal suo utilizzo.
- Aprise (Glucon Inc., Boulder, CO): tale dispositivo, proposto nel 2007, si basa su un interessante metodo cosiddetto foto acustico [18]. Una luce laser viene applicata alla pelle in corrispondenza di un vaso sanguigno, causando un piccolo ma rapido incremento della temperatura del vaso stesso e producendo un leggero suono. Il dispositivo è in grado di "catturare" il suono e di determinare il livello di glucosio nel sangue in base alle caratteristiche acustiche del suono. A differenza degli altri dispositivi, l'Aprise non misura il glucosio nell'interstizio ma direttamente nel sangue. Il dispositivo tuttavia non è ancora stato approvato dalla FDA.
- NBM (OrSense Ltd., Nes Ziona, Israele): il dispositivo si basa sulla spettroscopia di occlusione [19]. La misurazione avviene attraverso una sonda posizionata sul dito. La sonda è dotata di sorgenti di luce e sensori operanti nel campo del vicino infrarosso (red/near-infrared, RNIR) e di dispositivi pneumatici che producono una pressione ultra sistolica in grado di occludere il vaso. La tecnologia è basata sugli effetti diretti del glucosio sulle proprietà di scattering dell'organo: il glucosio infatti abbassa il coefficiente di scattering del tessuto e, a causa del minor numero di fotoni assorbiti, l'intensità della luce aumenta. Il dispositivo tuttavia è ancora allo stadio prototipale e non è stata ancora richiesta l'approvazione né della CE né della FDA.
- Pendra (Pendragon Medical, Svizzera): introdotto nel 2000 dalla società svizzera Pendragon Medical nel 2000 e approvato con marchio CE nel maggio 2003, questo dispositivo è stato poi ritirato dal mercato nel primo periodo del 2005 [29]. Considerabile di fatto come il primo dispositivo non invasivo per l'automonitoraggio del glucosio, il Pendra utilizzava per misurare il consumo di glucosio una tecnologia nota come spettroscopia d'impedenza. Attraverso l'applicazione di un piccolo campo magnetico, il dispositivo era in grado di misurare il flusso di sodio attraverso la membrana mitocondriale, associabile appunto al consumo di glucosio. Tuttavia, test effettuati per valutare la qualità delle misure ottenibili con tale dispositivo hanno messo in luce alcuni suoi limiti, tra i quali la possibilità di fornire indicazioni terapeutiche potenzialmente dannose per il paziente.
- Solianis Multisensor (Solianis Monitoring AG, Zürich, CH): il dispositivo utilizza la Spettroscopia d'Impedenza (IS) per rilevare le variazioni della concentrazione di glucosio [4]. La tecnologia IS utilizza frequenze comprese tra i 100 kHz e gli 8 GHz per misurare gli effetti sulla conduttività del dispositivo provocati da cambiamenti del livello di glucosio nel sangue, nelle cellule e nel fluido interstiziale. L'azienda Solianis Monitoring è stata recentemente assorbita dal gruppo Biovotion, e con essa anche la tecnologia del dispositivo Solianis Multisensor.

1.7 Aspetti di forza e critici dei sensori CGM

I dispositivi CGM hanno il grande vantaggio di fornire sia al medico curante che al paziente un'informazione molto più completa e dettagliata di quella ottenuta grazie a glucometro pungi-dito. La possibilità di avere a disposizione il profilo glicemico in tempo reale e di conoscere anche l'informazione sul trend sono molto utili per l'auto-monitoraggio, la gestione della dieta e del dosaggio d'insulina. Questi dispositivi

sono anche in grado di generare allarmi visivi e acustici, qualora il livello di glicemia superi le soglie del range di normalità. Esistono tuttavia diversi aspetti critici per l'uso corretto e accurato dei dispositivi CGM. In primo luogo bisogna ricordare che viene misurato un segnale proporzionale al glucosio nell'interstizio e non al glucosio plasmatico. Esso presenta delle caratteristiche diverse e in particolare il glucosio interstiziale si presenta come una versione distorta del glucosio plasmatico, dove la distorsione del segnale consiste di un shift orizzontale, dovuto al ritardo fisiologico con cui il glucosio diffonde dal plasma all'interstizio, e di uno shift verticale dovuto al deterioramento del sensore nel tempo (Figura 1.4).

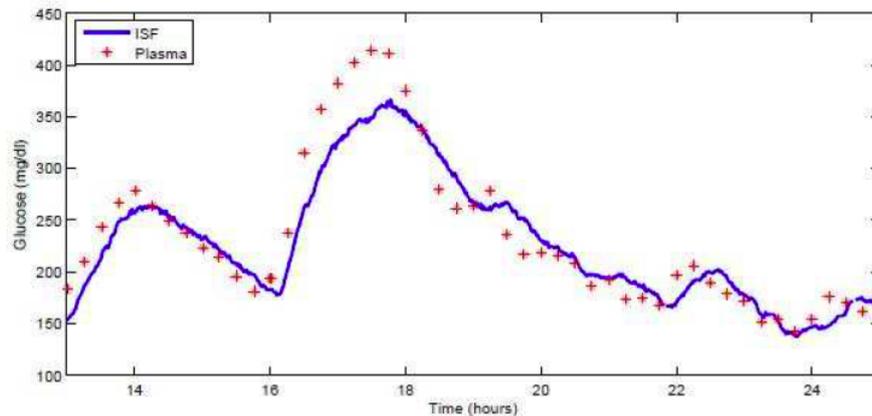


Figura 1.4: Confronto tra misurazione plasmatiche di riferimento (rosso) e misure da sensore CGM (blu)

Il problema della calibrazione del sensore risulta allora di fondamentale importanza per correggere la distorsione. Oltre alla calibrazione sono necessari anche opportuni algoritmi di filtraggio e denoising per il trattamento del rumore di misura che è intrinseco al segnale [21].

1.8 Uso dei sensori CGM per la predizione della glicemia futura

Un interessante applicazione dei sensori CGM che tutt'ora è in fase di studio riguarda il riconoscimento anticipato di episodi critici quali ipo/iperglicemie mediante tecniche di predizione

1.8.1 Prevenzione di episodi ipo/iperglicemici mediante l'utilizzo di algoritmi predittivi

L'avvento dei nuovi dispositivi per il monitoraggio continuo della glicemia ha reso possibile lo sviluppo di nuove tecniche per la cura del diabete. In particolare l'informazione continua sulla glicemia viene utilizzata per aggiustare la terapia e generare allarmi nell'istante in cui si verificano episodi pericolosi. Tuttavia sarebbe molto più utile poter prevedere in anticipo l'insorgere di questi episodi critici dando al paziente il tempo per intervenire. I dati ricavati dai sensori CGM si presentano come delle serie temporali adatte ad essere elaborate per fini predittivi, e alcuni dei dispositivi ad oggi in commercio, come anche il Dexcom Seven Plus, sono già dotati di alcuni algoritmi di proiezione che rilevano i trend degli ultimi campioni generando un allarme sulla base del valore di glicemia predetto. Gli algoritmi integrati all'interno di questi dispositivi però non sono molto sofisticati, tanto che diversi studi sono stati effettuati nel tentativo di adottare tecniche più raffinate per la predizione della glicemia futura a partire dai dati raccolti. Nei seguenti si paragrafi descriveranno alcune recenti tecniche, stato dell'arte per la

predizione della glicemia a breve termine (massimo 60 minuti) basato su dati monitorati frequentemente. Particolare attenzione verrà prestata ai metodi basati su reti neurali.

1.8.2 Predizione: modello polinomiale e modello autoregressivo

In [21], Sparacino et al. propongono un confronto tra due modelli diversi per la predizione: un modello polinomiale del primo ordine, $\text{poly}(1)$, e un modello autoregressivo del primo ordine, $\text{AR}(1)$. Entrambi i modelli vengono identificati ricorsivamente sui campioni recenti ottenuti da dispositivo CGM minimamente invasivo Glucoday. I modelli in particolare presentano una struttura fissa ma a parametri variabili che, per ogni nuovo campione ottenuto, vengono riottimizzati sfruttando le misure CGM passate, opportunamente pesate per mezzo di un coefficiente d'oblio μ che determina la "memoria" del sistema. Utilizzando un orizzonte di predizione di 30 minuti, i due modelli proposti da Sparacino et al. sono stati in grado di predire l'attraversamento delle soglie critiche con un anticipo di 20-25 minuti d'anticipo. Ulteriori modelli sono stati proposti per la predizione della glicemia futura. Eren-Oruklu et al. , in [32], hanno proposto sia un modello AR di ordine 3 che un modello AR a media mobile di ordine (3,1) in cui i parametri (tempo-varianti) vengono identificati ricorsivamente attraverso l'applicazione di un algoritmo di RLS con un coefficiente d'oblio μ variabile grazie ad un metodo per il riconoscimento dei cambiamenti di trend glicemico. Reifman et al. [31] hanno adottato invece un approccio diverso alla predizione, proponendo un modello AR tempo-invariante di ordine elevato (10) identificato una sola volta durante un intervallo di *burn-in* e utilizzato successivamente per la predizione di tutta la serie temporale. Tale approccio è stato verificato solamente su 9 serie temporali CGM campionate ad 1 minuto. In Reifman et al. [31] viene inoltre suggerito l'uso di un modello AR con parametri "di popolazione", lavoro successivamente sviluppato anche da Gani et al. in [33].

Un ulteriore approccio di tipo stocastico non parametrico è stato invece presentato da Palerm e Baquette in [35]. In questo lavoro viene proposto l'utilizzo un filtro alla Kalman i cui parametri vengono empiricamente determinati in maniera retrospettiva massimizzando la sensibilità e la specificità del modello stesso. Vengono prese in considerazione anche alcune informazioni note a priori relative alla regolarità del segnale glicemico, inglobate nel modello attraverso un processo di rumore bianco pluri-integrato. Tale approccio è stato applicato su 13 set di dati ottenuti da misure CGMS[®] (campionamento a 5 min) durante un clamp ipoglicemico (4 ore di osservazione).

1.8.3 Predizione con reti neurali: stato dell'arte

La predizione della glicemia futura basata su reti neurali è un ambito di ricerca tutt'ora in fase di sviluppo. Tra le varie soluzioni che sono individuabili in letteratura, riporto di seguito quelle più interessanti ai fini di questa tesi. In [22], Pérez-Gandía et al. propongono un nuovo metodo basato su reti neurali per la predizione on-line della glicemia futura sulla base di dati raccolti da sensori CGM. Il predittore proposto sfrutta una rete neurale alimentata con dati raccolti negli ultimi 20 minuti e la cui uscita è il valore della glicemia stimato per un determinato orizzonte di predizione (PH). Questo lavoro mette in luce le potenzialità delle reti neurali nella predizione in tempo reale della concentrazione futura di glucosio a partire da dati CGM, ottenendo risultati paragonabili o migliori rispetto ad un modello $\text{AR}(1)$. In [23], Pappada et al. suggeriscono un modello di rete neurale dove, oltre ai dati ottenuti mediante sensore CGM, si utilizza la capacità intrinseca delle reti di gestire ingressi multipli ed eterogenei considerando i segnali relativi ai dosaggi insulinici, la misura in tempo reale della velocità di variazione della glicemia ed anche informazioni relative all'apporto nutrizionale, allo stile di vita e a fattori di tipo emozionale. Infine, in [24], Zecchin et al. propongono una nuova tipologia di rete in cui oltre ai dati ottenuti con

campionamento a 1 minuti da sensore CGM si tiene conto dell'informazione sui pasti processata attraverso un modello fisiologico che genera la glucose rate of appearance (RA_G), ovvero un segnale proporzionale alla velocità con cui il glucosio entra effettivamente in circolo dopo un pasto. La rete neurale in questione utilizza inoltre le informazioni ottenute attraverso un predittore lineare del primo ordine del tipo indicato nel Paragrafo 1.8.2. Questo particolare modello di rete cerca quindi di slegare le componenti lineari del segnale CGM, elaborate per mezzo del modello polinomiale, da quelle non lineari, trattate invece attraverso la rete stessa. I risultati ottenuti con questo modello si sono rivelati molto incoraggianti, con un miglioramento rispetto alla rete di Perez-Gandia sia in termini di aderenza al profilo target che di guadagno temporale (~ 5 minuti maggiore) e di regolarità del profilo predetto.

1.9 Scopo della tesi e sua articolazione

I sensori per il monitoraggio continuo del glucosio sono un ottimo strumento per migliorare la terapia del diabete, rispetto al solo monitoraggio attraverso misure SMBG. Tuttavia i dati ottenuti da questi dispositivi forniscono solo una misura del valore della glicemia attuale o una semplice proiezione futura grazie ad alcuni algoritmi di estrapolazione inseriti nei dispositivi stessi. Diversi studi hanno però dimostrato come l'uso di raffinate tecniche di predizione permettano una stima più corretta della glicemia con un orizzonte temporale breve (30 - 75 minuti), dando quindi la possibilità di conoscere in anticipo l'eventuale verificarsi di eventi ipo/iperglicemici e di agire di conseguenza per evitarli, riducendo così il rischio di danni a breve e lungo termine dovuti a questi episodi critici. I modelli basati su reti neurali sono particolarmente interessanti, poiché permettono di sfruttare ingressi appartenenti a domini diversi (come glicemia, informazione sui pasti e insulina) per predire la glicemia futura e presentano, per loro natura, un comportamento non lineare. Scopo di questa tesi è valutare le potenzialità di una diversa tipologia di rete neurale, nota come Jump-net, in cui la struttura della rete stessa si differenzia dalle classiche reti feedforward per la presenza di connessioni dirette tra gli ingressi e le uscite, oltre alle connessioni "in avanti" tra strati successivi. In particolare le reti presentate in questo elaborato sono state ottimizzate con un'orizzonte di predizione di 30 minuti, per fornire al paziente un tempo utile da consentirgli di gestire eventuali episodi di ipo e iperglicemia. Come si vedrà nei capitoli successivi, diverse strutture sono state testate, con l'obiettivo di individuare un modello semplice in termini di richiesta di risorse ma in grado di competere con gli algoritmi dello stato dell'arte.

In particolare, nel Capitolo 2 verrà fatta un'introduzione alle reti neurali descrivendone in dettaglio alcune delle diverse architetture disponibili, i modelli ad oggi proposti in ambito di predizione della glicemia futura e le modalità di allenamento. Nel Capitolo 3 particolare attenzione verrà rivolta alle reti jump utilizzate in questo elaborato. Verranno definite le diverse tipologie da noi utilizzate, i possibili ingressi per alimentarle e le modalità con cui sono state scelte le reti candidate per la fase di test. Nel Capitolo 4 verranno presentati i diversi data sets utilizzati per l'allenamento e il test delle reti jump. Nel Capitolo 5 le reti jump candidate verranno testate su data set simulato e confrontate con i due predittori stato dell'arte poly(1)[21] e NN-LPA[24]. I medesimi predittori verranno testati successivamente nel Capitolo 6 su data set con campionamento a 5 minuti ottenuto grazie a sensore Dexcom® 7-Plus, mentre nel Capitolo 7 verranno analizzate le performance dei modelli in competizione su data set ottenuto da sensore Abbott Navigator™ con campionamento a 1 minuto. Nel Capitolo 8, infine, verranno presentati i risultati complessivi ottenuti nell'ambito di questa ricerca e proposti alcuni possibili sviluppi futuri.

Capitolo 2

Introduzione alle reti neurali

Facendo riferimento a quanto riportato da Ajoy e Dovriboje [12], possiamo definire rete neurale artificiale, o più semplicemente rete neurale, un sistema di elaborazione parallela a carattere fortemente distribuito, nato da alcune analogie col cervello umano. L'idea alla base di questo sistema è scaturita dall'osservazione delle differenze sostanziali tra il modo di gestire le informazioni del cervello e quello degli attuali elaboratori elettronici. L'encefalo umano può essere concepito come un calcolatore ad altissima complessità che opera naturalmente in maniera non lineare e fortemente parallela. Esso ha la capacità di riorganizzare la propria struttura, composta di neuroni tra loro connessi da sinapsi, per svolgere compiti come la percezione, il controllo motorio o il riconoscimento di pattern, con una velocità di gran lunga superiore a quella di un computer. Una rete neurale, dal canto suo, sfrutta una fitta rete di semplici unità tra loro fortemente interconnesse, note anch'esse come neuroni, con lo scopo di emulare le modalità con cui il cervello svolge determinate funzioni. Di particolare interesse sono le reti neurali che, in seguito a un processo d'apprendimento o allenamento riescono a svolgere compiti di elevata complessità. Volendo quindi dare una definizione finale che integri le diverse caratteristiche di questi sistemi, possiamo riferirci a [11] dicendo che:

Una rete neurale è un sistema di elaborazione che fa un uso massimo del calcolo parallelo e distribuito. E' composto da unità semplici che hanno la naturale propensione ad immagazzinare esperienza e renderla disponibile all'uso. Una rete neurale ricorda il cervello per due aspetti:

- *La conoscenza viene acquisita dall'ambiente esterno attraverso l'apprendimento.*
- *Le forze che legano i neuroni, nota anche come pesi sinaptici, vengono utilizzati dalla rete per immagazzinare la conoscenza acquisita.*

Il processo attraverso cui avviene l'apprendimento prende il nome di algoritmo di apprendimento o di allenamento, la cui funzione è modificare ordinatamente i pesi sinaptici della rete per portare a compimento un obiettivo predeterminato. Questo riaggiustamento dei pesi sinaptici rappresenta il metodo standard attraverso cui le reti vengono allenate. La storia delle reti neurali, dalla preliminare idea proposta da Hebb nel 1949 fino ai giorni nostri, ha visto soprattutto negli ultimi decenni la nascita di modelli innovativi che sono stati impiegati in svariati campi. Ad oggi tuttavia, in particolare nel settore dell'ingegneria, le strutture largamente più utilizzate e oggetto di approfondite ricerche e innovative applicazioni sono le multilayer perceptron network (MLPN) e le radial basis function networks (RBFN). Delle prime si parlerà in maniera più approfondita nel paragrafo 2.2.1.

2.1 Reti neurali monostrato

2.1.1 Il neurone biologico

I neuroni, o cellule nervose, sono i principali elementi di cui è composto il sistema nervoso ed il ruolo da essi svolto consiste nella conduzione dei segnali nervosi noti come potenziali d'azione. Sono cellule altamente specializzate con la particolarità di essere amitotiche, non in grado dunque di rigenerarsi in caso di danno o deterioramento. Ciascun neurone, come si vede dalla Figura 2.1, è composto di tre parti

1. Il corpo cellulare, detto anche soma
2. Uno o più dendriti
3. Un unico assone

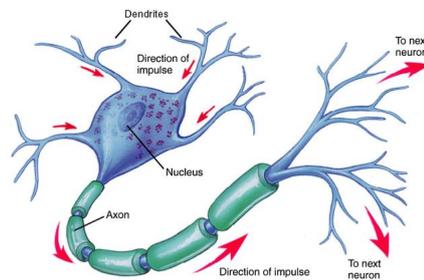


Figura 2.1: Il neurone biologico (immagine tratta da [16])

Il segnale nervoso giunge al neurone attraverso i dendriti, il cui scopo è proprio quello di raccogliere il segnale in ingresso alla cellula. Il potenziale d'azione, di natura elettrica, viene poi elaborato nel soma e ritrasmesso attraverso l'assone verso altri neuroni. In particolare la trasmissione da un neurone ad un altro avviene a livello delle sinapsi che, attraverso meccanismi di natura chimica o elettrica, permettono il passaggio del segnale dal neurone afferente a quello ricevente. Le sinapsi possono essere di tipo eccitatorio o inibitorio, a seconda che favoriscano o meno l'attivazione del neurone a valle della sinapsi stessa. Infatti, affinché questo possa generare un nuovo potenziale d'azione, è necessario che i segnali in ingresso si combinino nel soma in modo tale da permettere al potenziale di membrana il superamento di una soglia di attivazione. Una volta generato e trasmesso il potenziale d'azione, la cellula nervosa presenta un breve periodo (1 ms) di refrattarietà durante il quale risulta insensibile ad ulteriori stimoli provenienti dall'esterno. Sebbene i singoli neuroni quindi siano strutturalmente semplici e funzionalmente lenti, il grande numero di cellule nervose e di sinapsi garantisce al cervello una velocità di elaborazione molto elevata.

2.1.2 Il neurone artificiale

Il neurone artificiale rappresenta l'unità di elaborazione fondamentale per le reti neurali artificiali. In Figura 2.2 viene presentato un modello per il neurone artificiale descritto a mezzo di diagramma a blocchi. In tale diagramma è possibile identificare diversi elementi:

1. Una serie di *sinapsi* o *connessioni*, ciascuna delle quali è caratterizzata da un *peso*. Nello specifico il segnale x_j in ingresso al neurone k viene moltiplicato per il suo peso w_k che, a differenza del caso

biologico, può assumere sia valori positivi che negativi, in modo tale da poter esprimere sia sinapsi di tipo eccitatorio che di tipo inibitorio;

2. Un *sommatore*, che combina linearmente i vari ingressi secondo i rispettivi pesi sinaptici;
3. Un blocco *funzione d'attivazione*, il cui scopo è limitare l'uscita del neurone ad un range finito di valori.
4. Un termine di bias, indicato con b_k , il quale può aumentare o diminuire il segnale in ingresso alla funzione d'attivazione a seconda che assuma un valore positivo o negativo.

Volendo descrivere in termini matematici il funzionamento del neurone, ecco che:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.1)$$

e

$$y_k = \varphi(u_k + b_k) \quad (2.2)$$

dove x_1, x_2, \dots, x_m sono i segnali d'ingresso; $w_{k1}, w_{k2}, \dots, w_{km}$ sono i pesi sinaptici del neurone k ; u_k è il risultato della combinazione lineare dei segnali d'ingresso; b_k è il bias; $\varphi(\cdot)$ è la funzione d'attivazione; e y_k è il segnale d'uscita del neurone. L'uso del bias comporta l'applicazione di una trasformazione lineare dell'uscita u_k , come mostrato da

$$v_k = u_k + b_k \quad (2.3)$$

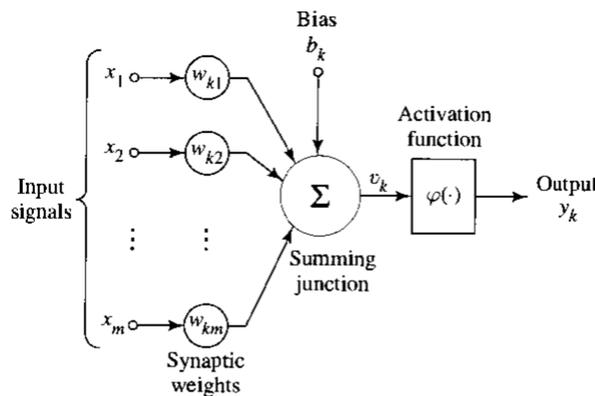


Figura 2.2: Modello non lineare del neurone (immagine tratta da [11])

In particolare, a seconda che il bias assuma un valore positivo o negativo, il legame tra il *campo locale indotto* o *potenziale d'attivazione* v_k e la combinazione lineare degli ingressi u_k del neurone k viene modificato come in Figura 2.3. E' da notare che a seguito della trasformazione affine, il grafico di v_k contro u_k non attraversa più l'origine.

Le equazioni sopra riportate possono essere riformulate nel seguente modo andando a includere il termine di bias tra gli ingressi della rete:

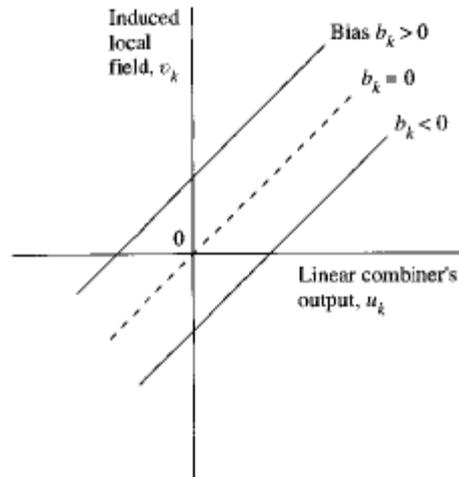


Figura 2.3: Trasformazione affine prodotta dal bias; Quando $u_k = 0$ si ha $v_k = b_k$. (immagine tratta da [11])

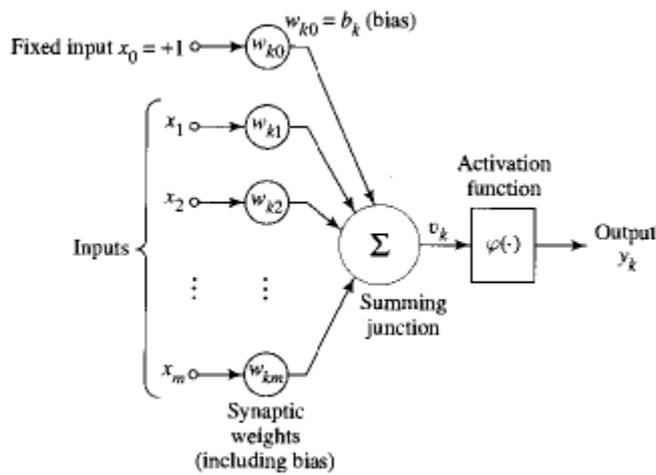


Figura 2.4: Un modello alternativo per il neurone non lineare (immagine tratta da [11])

$$v_k = \sum_{j=0}^k w_{kj} x_j \quad (2.4)$$

$$y_k = \varphi(v_k) \quad (2.5)$$

dove, all'interno dell'equazione 2.4 è stato aggiunto una nuova sinapsi in cui l'ingresso e il peso sinaptico assumo i seguenti valori:

$$x_0 = +1 \quad (2.6)$$

$$w_{k0} = b_k \quad (2.7)$$

Ecco allora che è possibile riformulare il modello del neurone k come in Figura 2.4. Si può vedere il contributo del bias preso in considerazione (1) aggiungendo un nuovo ingresso con valore costante pari a +1, e (2) aggiungendo un nuovo peso sinaptico pari a b_k . Sebbene i due modelli presentati possano sembrare diversi, dal punto di vista matematico essi risultano equivalenti.

Tipologie di funzione d'attivazione:

Le funzioni d'attivazione, denotate con $\varphi(v)$, stabiliscono l'uscita del neurone in funzione, appunto, del potenziale d'attivazione v . Ne esistono varie tipologie diverse, tra cui le tre più utilizzate sono:

1. *Funzione a soglia*, riportata in Figura 2.5a e descritta dall'equazione

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases} \quad (2.8)$$

per la quale l'uscita del neurone risulta:

$$y_k = \begin{cases} 1, & \text{se } v_k \geq 0 \\ 0, & \text{se } v_k < 0 \end{cases} \quad (2.9)$$

Con questa tipologia di funzione il segnale generato dal neurone si assesta al valore costante +1 se il potenziale d'attivazione è non-negativo e a 0 altrimenti.

2. *Funzione lineare*, riportata in Figura 2.5b e descritta dall'equazione

$$\varphi(v) = \begin{cases} 1, & v \geq +\frac{1}{2} \\ v, & +\frac{1}{2} > v > -\frac{1}{2} \\ 0, & v \leq -\frac{1}{2} \end{cases} \quad (2.10)$$

in cui è stato adottato un fattore d'amplificazione unitario per la funzione d'attivazione.

3. *Funzione sigmoideale*, riportata in Figura 2.5c e il cui grafico assume una caratteristica forma ad *S*. Questa classe risulta la più utilizzata nella progettazione di una rete neurale. La funzione si presenta come strettamente crescente, continua e differenziabile. L'equazione 2.11 descrive una funzione sigmoideale in cui l'uscita del neurone è compresa nel range $[0, 1]$. È da notare la presenza del parametro a che permette di regolare l'inclinazione della funzione.

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (2.11)$$

Delle funzioni sigmoideali fa parte anche la funzione *tangente iperbolica*, descritta dall'equazione (2.12), in cui l'uscita non assume solo valori non-negativi, ma copre tutto il range $[-1, +1]$

$$\varphi(v) = \frac{1 - e^{-2av}}{1 + e^{-2av}} \quad (2.12)$$

Esistono anche altre tipologie di funzioni d'attivazione e per una visione più approfondita si rimanda a [11].

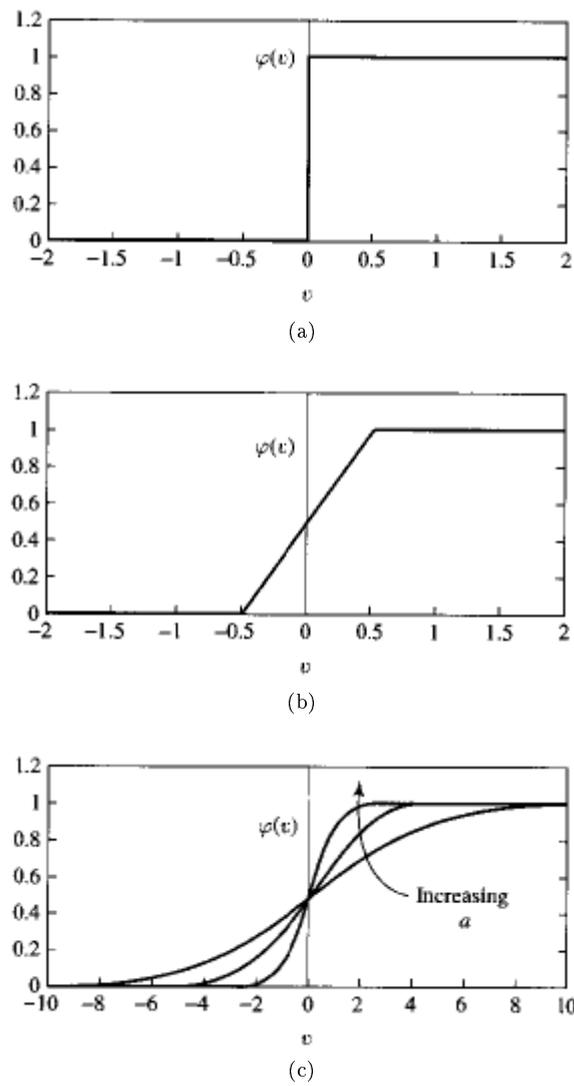


Figura 2.5: Alcune possibili funzioni d'attivazione; (a) Funzione d'attivazione a soglia; (b) Funzione d'attivazione lineare; (c) Funzione d'attivazione Sigmoidale (immagini tratte da [11])

2.2 Reti neurali multistrato

Sebbene nel tempo la varietà di reti neurali proposte sia cresciuta, la rete multistrato nota anche come *multilayer perceptron (MLP)* è rimasta la struttura ampiamente più diffusa ed utilizzata. In particolare quanto detto è valido per la struttura che prevede l'uso di tre livelli in cui il livello d'ingresso e quello d'uscita sono tra di loro connessi per mezzo di un livello cosiddetto nascosto (Figura 2.6). Questa particolare tipologia di reti presenta una capacità intrinseca di mappare un arbitrario set di ingressi-uscite, rendendole in tal modo un buono strumento nella predizione di serie temporali [12].

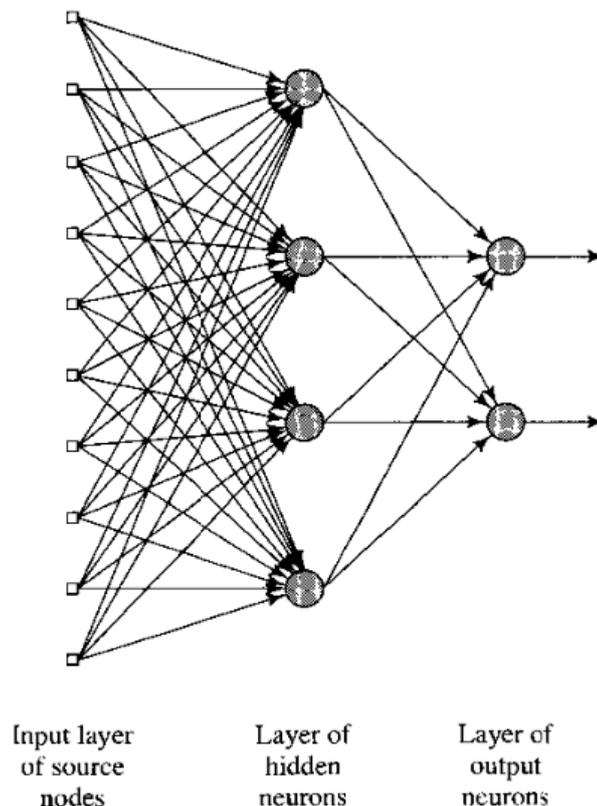


Figura 2.6: Rete neurale feedforward (immagine tratta da [11])

2.2.1 Multilayer perceptron

L'MLP fa parte di quella categoria di reti neurali dette feedforward, in cui il segnale può propagarsi sequenzialmente lungo le sinapsi solo dal livello d'ingresso al livello d'uscita, passando attraverso i livelli nascosti. Questa classe di reti presenta, oltre un livello d'ingresso ed un livello d'uscita, uno o più livelli di neuroni nascosti. La funzione di questi livelli nascosti è quella di permettere alla rete di estrarre dal segnale in analisi statistiche di ordine maggiori rispetto a quanto sarebbe possibile con solo il livello d'ingresso e d'uscita [11]. Sebbene la presenza di un secondo livello nascosto aumenti le capacità della rete di estrarre caratteristiche particolari del segnale in analisi, l'aumento della qualità dei risultati si scontra con l'aumento della complessità computazionale annessa allo sviluppo e all'allenamento della rete stessa, tanto che l'uso di un secondo livello nascosto tipicamente viene sconsigliato a meno di esigenze particolari.

La rete di Figura 2.6 mostra lo schema di una MLP a tre strati a cui tipicamente si fa riferimento come rete 10 – 4 – 2, intendendo con questa dicitura che la rete utilizzata presenta dieci ingressi, uno strato nascosto di 4nel neuroni e uno strato d'uscita con 2.

2.2.2 Jump network

Una struttura alternativa al classico MLP è una rete neurale di classe feedforward in cui siano presenti dei "salti di livello", le cosiddette jump connection (Figura 2.7). Il nome di queste reti, Jump-net o reti jump, deriva proprio da questi salti di livello. La presenza di queste particolari connessioni fa sì che i diversi livelli che compongono la rete possono scambiarsi l'informazione senza necessariamente passare attraverso un livello nascosto intermedio. Il motivo d'interesse per questa classe di reti neurali risiede nella possibilità di valutare, nel medesimo istante, le caratteristiche lineari e non lineari del segnale in analisi [13]. Nello specifico:

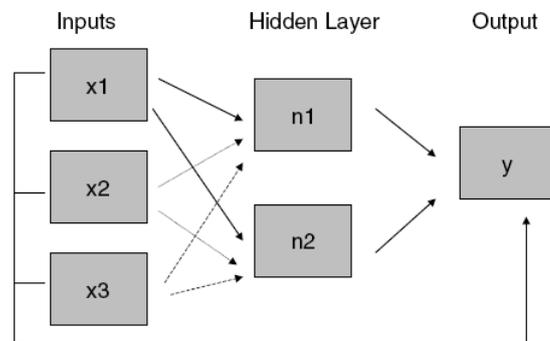


Figura 2.7: Rete neurale feedforward con connessioni jump (immagine tratta da [13])

1. Se la relazione tra gli ingressi e le uscite risulta puramente lineare, solo le connessioni jump presentano significatività;
2. se la relazione è puramente non lineare sono invece le sinapsi in uscita dal livello nascosto ad assumere maggiore significatività;
3. se la relazione ingresso-uscita presenta infine un comportamento misto, la significatività delle varie sinapsi assumerà un peso tale da tener conto in modo opportuno delle due componenti presenti nel segnale in analisi.

Attraverso un'unica rete neurale risulta quindi possibile, almeno in teoria, gestire serie temporali a comportamento misto lineare/non lineare senza doversi obbligatoriamente appoggiare a modelli esterni.

2.3 Accenno ad altre tipologie di reti neurali

2.3.1 Reti neurali ricorsive

Le reti neurali ricorsive (*recurrent neural net* - RNN) si differenziano da quelle feedforward per la presenza di almeno un collegamento ad anello feedback. La Figura 2.8 mostra ad esempio una rete neurale ricorsiva dotata di uno strato nascosto in cui oltre agli ingressi classici esterni alla rete sono presenti dei self-feedback loop, cioè delle sinapsi che permettono ai neuroni nascosti di "auto alimentarsi" con le proprie

uscite. La presenza dell'anello feedback ha un impatto profondo sulle performance della rete e sulla sua capacità di apprendimento. Dalla figura si può notare inoltre la presenza di particolari elementi di ritardo unitario, denotate dal simbolo z^{-1} . Come risultato dell'introduzione di questi elementi di ritardo, la rete presenta un comportamento fortemente non lineare [11]. Sebbene le FFNN siano in grado di modellare molte varietà di relazioni ingresso-uscita, esse mancano totalmente di memoria, con la conseguenza che se alimentate continuamente dallo stesso segnale la rete produrrà sempre la stessa uscita. Le reti ricorsive invece presentano una memoria teoricamente infinita proprio grazie alla presenza di feedback e degli elementi di ritardo, con la conseguenza che l'uscita della rete non dipende semplicemente dall'ingresso attuale con cui viene alimentata ma anche dal suo stato, permettendo alle RNN di rispondere in modo diverso ad uno stesso ingresso [13].

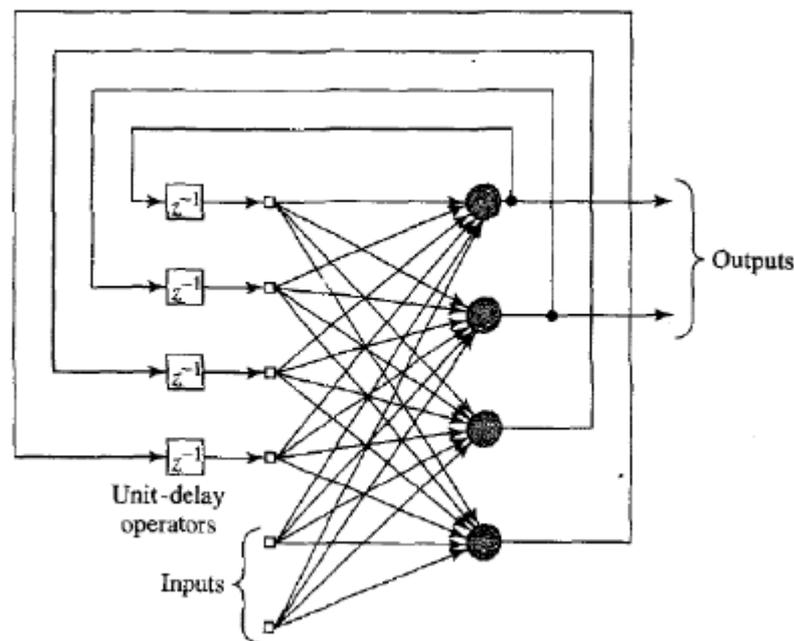


Figura 2.8: Rete neurale ricorsiva (immagine tratta da [11])

2.3.2 Reti neurali radial-basis function

Questa tipologia di reti differisce in maniera sostanziale dalle reti fin'ora considerate. Con questi modelli la costruzione della rete neurale viene vista come un problema di approssimazione di una curva in uno spazio a molte dimensioni. In tal senso l'apprendimento della rete corrisponde con il trovare una superficie in grado di aderire bene ai dati forniti in fase di training, valutando statisticamente la qualità del risultato. Allo stesso modo la capacità della rete di generalizzare bene il problema coincide con la capacità di questa superficie di seguire i dati del test set. In queste reti i neuroni forniscono un insieme di funzioni, note come funzioni radiali (o *radial-basis functions*) che definiscono una base per l'espansione dei vettori d'ingresso nello spazio degli strati nascosti [11].

I principali aspetti per cui queste reti si differenziano da quelle classiche sono:

1. l'ingresso di ogni neurone del primo strato nascosto e la distanza tra gli elementi del vettore in ingresso alla rete e i pesi che collegano tali elementi col neurone considerato, moltiplicata per il bias;

2. la funzione di trasferimento dei neuroni dello strato nascosto e una funzione radiale. La Figura 2.9 mostra ad esempio una funzione che assume valori crescenti man mano che la distanza tra i pesi e gli elementi del vettore d'ingresso si riduce, il cui valore massimo viene raggiunto qualora la distanza risulti nulla.

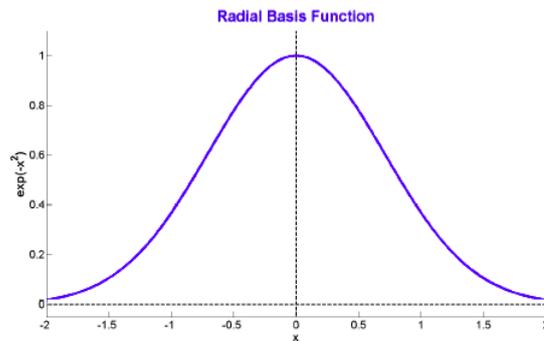


Figura 2.9: Esempio di funzione radiale

2.4 Apprendimento errore-correzione: l'algoritmo di *back-propagation*

Abbiamo già accennato in precedenza che nel caso di reti feedforward come il MLP l'apprendimento avviene attraverso la presentazione alla rete di un training set composto da diversi elementi, detti anche pattern, ingresso-uscita. Grazie all'allenamento la rete è in grado di generare una mappatura non lineare applicabile per l'approssimazione di funzioni, classificazione di pattern, analisi dei segnali, predizione etc.

Come vedremo più avanti, durante la fase d'apprendimento i pesi sinaptici della rete vengono adattati al fine di minimizzare la discrepanza tra un segnale target contenuto nel training set, indicato con t_k e il segnale in uscita dalla rete neurale, indicato a sua volta con y_k . L'algoritmo di *back-propagation*, o di retropropagazione, è uno dei metodi più utilizzati per l'allenamento delle reti il cui scopo è minimizzare una funzione d'errore opportunamente costruita sui segnali in uscita della rete. Nello specifico l'algoritmo di back-propagation è un algoritmo di ricerca di minimo non vincolato basato sul gradiente della funzione d'errore prima citata [12].

2.4.1 Apprendimento attraverso back-propagation

Una rete neurale i cui neuroni siano dotati di una funzione d'attivazione differenziabile produce un'uscita anch'essa differenziabile, sia secondo gli ingressi che secondo i pesi e i bias. Definendo dunque una funzione d'errore derivabile sulla base dell'uscita della rete stessa, risulterà anch'essa derivabile rispetto ai pesi sinaptici. Grazie a questa proprietà, le derivate parziali possono essere utilizzate al fine di individuare una legge di aggiornamento dei pesi tale da minimizzare, per mezzo di un algoritmo basato sul gradiente o un altro algoritmo maggiormente prestante, la funzione d'errore. Questo algoritmo che sfrutta la derivata parziale dell'errore calcolata secondo i pesi sinaptici prende il nome di algoritmo di *back-propagation (BP)*. Inizialmente sviluppato da diversi ricercatori, come Webos (1974) e Parker (1985), questa metodologia per l'allenamento delle reti neurali ha preso piede soprattutto dopo un articolo di Rumelhart, Hinton e Williams (1986), che l'hanno reso popolare [15].

Di seguito deriveremo l'algoritmo BP per una generica rete feedforward in cui la funzione d'attivazione non lineare per cui è stata scelta un'arbitraria funzione d'errore, con la specifica che tale funzione sia derivabile. Nelle reti feedforward multistrato, come il MLP, in ciascun neurone vengono elaborati gli ingressi secondo la formulazione

$$a_j = \sum_i w_{ji} z_i \quad (2.13)$$

in cui z_i rappresenta il segnale in uscita dal neurone i -esimo, o uno degli ingressi alla rete, che viene trasmesso attraverso una sinapsi avente peso sinaptico w_{ji} al neurone j -esimo. In particolare la sommatoria viene calcolata su tutti i neuroni collegati al neurone j . Il bias può essere gestito, come visto nel paragrafo 2.1.2, introducendo un ulteriore ingresso (o neurone) con potenziale costante $+1$. In tal modo non è più necessario considerare in maniera esplicita il bias. Il risultato della 2.13 viene quindi passata alla funzione d'attivazione non lineare $\varphi(\cdot)$ che produce il segnale in uscita dal neurone j esprimibile in questa forma

$$z_j = \varphi(a_j) \quad (2.14)$$

di cui è necessario fare una precisazione. E' da notare, infatti, che s'è usata una notazione unificata per indicare i diversi ingressi ai neuroni, anche se in realtà gli ingressi della rete dovrebbero essere indicati con x_i . In tal senso, anche le uscite presentate nella formula 2.14 comprendono anche le uscite della rete, esprimibili come y_k . Tali precisazioni sono necessarie per comprendere meglio i passaggi successivi.

Dobbiamo ora individuare una funzione d'errore opportuna tale da essere minimizzata per mezzo del valore dei pesi sinaptici della rete. Una possibilità è quella di usare una funzione ottenuta dalla combinazione lineare degli errori commessi dalla rete su tutti gli elementi del training set

$$E = \sum_n E^n \quad (2.15)$$

dove n sta ad indicare l' n -esimo pattern ingresso-uscita che compone il training set. Senza perdere di generalità si può supporre che E^n sia una funzione derivabile esprimibile secondo le uscite della rete

$$E^n = E^n(y_1, \dots, y_c) \quad (2.16)$$

L'obiettivo ora è quello di studiare una metodologia per calcolare la derivata parziale dell'errore rispetto ai pesi e ai bias della rete. Per semplicità prenderemo in considerazione un unico elemento del training set, riservandoci la possibilità di combinare i risultati di tutti gli elementi secondo la 2.15.

Per prima cosa il vettore d'ingresso viene fatto fluire attraverso la rete dagli ingressi alle uscite, supponendo di aver calcolato attraverso la 2.13 e la 2.14 i necessari segnali d'uscita dei neuroni nascosti e dei neuroni d'uscita. Questa prima fase dell'allenamento viene denominata *forward propagation*. Andiamo ora a stimare la derivata di E^n secondo un generico peso w_{ji} (per semplicità di notazione ometteremo gli apici n). In prima battuta è possibile notare come E^n dipenda da w_{ji} per via dei soli ingressi al neurone j . Attraverso derivazione a catena possiamo scrivere

$$\frac{\partial E^n}{\partial w_{ji}} = \frac{\partial E^n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} \quad (2.17)$$

Con l'uso della seguente notazione

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} \quad (2.18)$$

è possibile riformulare la 2.13 nel seguente modo

$$\frac{\partial a_j}{\partial w_{ji}} = z_i \quad (2.19)$$

e con le opportune sostituzioni, la 2.17 può dunque essere riscritta

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j z_i \quad (2.20)$$

Per mezzo di questa equazione la derivata richiesta risulta quindi esprimibile in termini di un δ_j che dovrà essere calcolato per tutti i neuroni della rete. Nel caso dello strato d'uscita il calcolo di δ_k è immediato. Grazie alla 2.18 infatti

$$\delta_k \equiv \frac{\partial E^n}{\partial a_k} = g'(a_k) \frac{\partial E^n}{\partial y_k} \quad (2.21)$$

dove si fa uso di y_k invece che di z_k . Per il calcolo della 2.21 è tuttavia necessario individuare un'espressione opportuna per $g'(a_k)$ e per $\frac{\partial E^n}{\partial y_k}$. Sfruttando nuovamente la derivazione a catena è possibile esprimere δ_j nel caso dei neuroni nascosti, ottenendo

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k} \frac{\partial a_k}{\partial a_j} \quad (2.22)$$

in cui la sommatoria dev'essere calcolata per tutti i neuroni k a cui è collegato il neurone j . Per un'esemplificazione del problema si può far riferimento alla Figura 2.10. Nella 2.22 variazioni di a_j vengono propagate alla funzione d'errore solo attraverso una variazione della variabile a_k . Applicando ora la 2.19 in 2.22 e prendendo in considerazione anche 2.13 e 2.14 si perviene alla formula del back-propagation

$$\delta_j = g'(a_j) \sum_k w_{kj} \delta_k \quad (2.23)$$

da cui si comprende bene che per il calcolo del δ nel caso di un neurone nascosto è necessario retro propagare quelli dei neuroni che lo succedono nella rete, come mostrato in Figura 2.10.

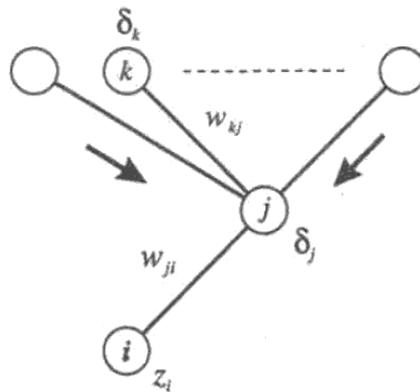


Figura 2.10: Esempio di calcolo del δ_j per il neurone nascosto a partire dai δ retropropagati dai neuroni k a cui il neurone j è connesso (immagine tratta da [15])

Dato che siamo in grado di calcolare il δ_k dei neuroni dello strato d'uscita, applicando ricorsivamente 2.23 è possibile ottenere tutti i δ di tutti i neuroni che compongono la rete feedforward, a prescindere

dalla topologia della rete stessa.

Possiamo dunque riassumere in quattro passi come è possibile calcolare la derivata parziale di E^n rispetto un generico peso della rete:

1. In principio forniamo alla rete l' n -esimo vettore d'ingresso del training set, facendolo propagare verso l'uscita sfruttando 2.13 e 2.14 per calcolare i segnali d'uscita di tutti i neuroni.
2. Per mezzo della 2.21 calcoliamo dunque δ_k per i neuroni dello strato d'uscita.
3. Otteniamo tutti i δ_j dei neuroni nascosti attraverso la retro-propagazione.
4. Attraverso la 2.20 infine possiamo calcolare le derivate richieste.

Applicando i precedenti passaggi per tutti i pattern che costituiscono il training set è possibile calcolare la derivata totale dell'errore, ottenendo così

$$\frac{\delta E}{\delta w_{ji}} = \sum_n \frac{\delta E^n}{\delta w_{ji}} \quad (2.24)$$

Sebbene per questa formulazione si sia supposto di utilizzare una funzione d'attivazione $g(\cdot)$ comune per tutti i neuroni, il problema è facilmente generalizzabile in modo da permettere l'uso di una funzione d'attivazione specifica per ciascun neurone.

Per completare il quadro è però necessario dare anche un'espressione alla derivata $g'(a)$. Per semplicità di calcolo analizzeremo il caso particolare di una funzione d'attivazione log-sigmoidale per i neuroni nascosti e di una funzione d'attivazione lineare per i neuroni d'uscita. Ricordiamo che la prima si presenta come

$$g(a) \equiv \frac{1}{1 + \exp(-a)} \quad (2.25)$$

ed ha la particolarità che la derivata può essere espressa facilmente come

$$g'(a) = g(a)[1 - g(a)] \quad (2.26)$$

Se la funzione d'errore è della tipologia ottenuta come somma quadratica degli errori, per l' n -esimo elemento del training set si ha

$$E^n = \frac{1}{2} \sum_{k=1}^c (y_k - t_k)^2 \quad (2.27)$$

Combinando la 2.26 e la 2.27 con la 2.23 possiamo ottenere i seguenti risultati per un neurone d'uscita

$$\delta_k = y_k - t_k \quad (2.28)$$

e per un neurone nascosto

$$\delta_j = z_j(1 - z_j) \sum_{k=1}^c w_{kj} \delta_k \quad (2.29)$$

dove la sommatoria viene effettuata su tutti i neuroni dello strato d'uscita. Ecco allora che si possono calcolare le derivate parziali della funzione d'errore per le due tipologie come

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j x_i \quad \frac{\partial E^n}{\partial w_{kj}} = \delta_k z_j \quad (2.30)$$

Tuttavia fin'ora ci siamo impegnati nel calcolo della derivate parziali della funzione d'errore, ma per ottenere un algoritmo d'apprendimento è necessario individuare un modo per aggiornare i pesi della rete in base a queste derivate. Tra le diverse strategie di ottimizzazione dei parametri, qui presentiamo quella in cui l'aggiornamento dei pesi avviene con un passo fissato μ . Esistono due diversi modi per allenare la rete, il primo in cui i pesi sinaptici vengono aggiornati ogni qualvolta viene presentato alla rete un nuovo pattern del training set (on-line learning), il secondo invece in cui il primo aggiornamento avviene solamente quando la rete è stata alimentata con tutti gli elementi del training set (batch learning). Nel primo caso l'aggiornamento dei pesi sinaptici risulta

$$\Delta w_{ji} = -\mu \delta_j x_i \quad (2.31)$$

mentre nel secondo si ha

$$\Delta w_{ji} = -\mu \sum_n \delta_j^n x_j^n \quad (2.32)$$

2.5 Uso di reti neurali nella predizione della glicemia futura

2.5.1 La rete di Perez-Gandia et al.

In [22], Pérez-Gandía et al. propongono per la predizione della glicemia futura una rete neurale alimentata in ingresso dalla serie glicemica raccolta a mezzo di sensore CGM negli ultimi 20 minuti di misura e la cui uscita è il valore della glicemia stimato per un determinato orizzonte di predizione (PH). La rete proposta presenta una struttura feedforward sviluppata su tre livelli, di cui i primi due caratterizzati rispettivamente da 10 e 5 neuroni aventi funzione d'attivazione a tangente sigmoidale, mentre il livello d'uscita è costituito da un singolo neurone con funzione d'attivazione lineare. In questo lavoro Pérez-Gandía e al. sfruttano un dataset ottenuto da due diversi sistemi CGM (9 soggetti monitorati con Medtronic Guardian e intervallo di campionamento di 5 minuti e 6 soggetti monitorati con l'Abbott Navigator e intervallo di campionamento di 1 minuto) e valutano le performance del predittore con tre diversi orizzonti temporali (15, 30 e 45 minuti) in termini di RMSE e ritardo nei fronti di salita e discesa della serie predetta.

2.5.2 La rete di Pappada et al.

In [23], Pappada et al. suggeriscono un modello di rete neurale dove, oltre ai dati ottenuti mediante sensore CGM, si utilizza la capacità intrinseca delle reti di gestire ingressi multipli ed eterogenei considerando i segnali relativi ai dosaggi insulinici, la misura in tempo reale della velocità di variazione della glicemia ed anche informazioni relative all'apporto nutrizionale, allo stile di vita e a fattori di tipo emozionale. In questo lavoro la rete neurale sviluppata presenta un primo livello con 11 neuroni (uno per ciascun ingresso utilizzato) e funzione d'attivazione a tangente sigmoidale, uno strato nascosto composto da 9 neuroni anch'essi a tangente sigmoidale, e uno strato d'uscita con un neurone avente funzione d'attivazione lineare. Per questo modello di rete è stato scelto come orizzonte temporale di predizione un valore di 75 minuti. I risultati dell'analisi svolta da Pappada et al. si mostrano interessanti come supporto al trattamento del diabete, anche se la quantificazione delle informazioni relative allo stile di vita e a fattori emozionali è difficile e altamente soggettiva. Tali segnali infatti non sembrano quantificabili in modo oggettivo.

2.5.3 La rete di Zecchin et al.

Zecchin et al. , in [24], propongono una nuova tipologia di rete in cui oltre ai dati ottenuti con campionamento a 1 minuti da sensore CGM si tiene conto dell'informazione sui pasti processata attraverso un modello fisiologico che genera la glucose rate of appearance (RA_G), ovvero un segnale proporzionale alla velocità con cui il glucosio entra effettivamente in circolo dopo un pasto, segnale che spiegheremo più nel dettaglio nel capitolo successivo. La rete neurale in questione utilizza inoltre le informazioni ottenute attraverso un predittore lineare del primo ordine, come quello proposto da [21].

L'orizzonte di predizione scelto è pari a 30 minuti. Nello specifico la rete di Zecchin et al. presenta una struttura avente uno strato nascosto di 8 neuroni con funzione d'attivazione a tangente sigmoideale e uno strato d'uscita a singolo neurone con funzione d'attivazione lineare. Gli ingressi alla rete sono i seguenti 6:

1. L'attuale valore di glicemia misurato attraverso sensore CGM;
2. Il trend relativo agli ultimi 15 minuti di monitoraggio;
3. L'errore del predittore poly(1) all'istante attuale

$$e(t) = \widehat{CGM}_I(t) - CGM(t)$$

dove $\widehat{CGM}_I(t)$ rappresenta la stima ottenuta attraverso modello polinomiale 30 minuti prima e relativo all'istante attuale, mentre $CGM(t)$ rappresenta l'attuale misura ottenuta dal sensore CGM

4. Il trend dell'errore relativo agli ultimi 15 minuti;
5. La stima della RA_G all'orizzonte di predizione
6. Tre stime del trend della RA_G intervallate di 10 minuti l'una dall'altra dall'istante attuale al PH.

La rete così strutturata non stima direttamente la glicemia all'orizzonte di predizione, bensì stima l'errore a PH del predittore lineare. Di conseguenza è necessario combinare l'informazione in uscita dalla rete neurale con la predizione della glicemia futura in uscita da poly(1) per ottenere la stima futura della glicemia. Questa rete dunque cerca di trattare in maniera separata le dinamiche lineari delle serie temporali ottenute da sensori CGM, elaborate per l'appunto dal predittore poly(1), da quelle non lineari elaborate per mezzo della rete neurale. Questa rete ha quindi bisogno di appoggiarsi ad un predittore esterno per trattare separatamente le diverse dinamiche insite nelle serie temporali glicemiche.

2.5.4 Potenziali margini di miglioramento

Da quanto visto risulta chiaro come le reti neurali siano in grado di trattare molto bene segnali aventi marcate dinamiche non lineari. Tuttavia, come anche messo in luce nel lavoro di Zecchin et al., alcuni segnali presentano anche delle componenti lineari che potrebbero essere trattate in maniera più opportuna attraverso modelli specificatamente studiati, come il modello polinomiale del primo ordine. Tuttavia, sarebbe preferibile poter utilizzare un unico predittore per gestire in modo semplice entrambe le dinamiche dei segnali in analisi. A tale proposito, come abbiamo già accennato nel Capitolo 2 e vedremo più dettagliatamente nel Capitolo 3, le reti jump con la loro particolare architettura, potrebbero essere in grado di combinare le caratteristiche di un predittore lineare con quelle di un predittore basato su rete neurale. Oggetto di questa tesi sarà proprio l'individuazione e la valutazione di alcune reti jump, confrontandole anche con i modelli stato dell'arte proposti da Sparacino et al. [21] e Zecchin et al. [24]

2.6 Conclusioni

In questo capitolo abbiamo parlato delle reti neurali, sviluppando il concetto a partire dalle analogie presenti tra le reti neurali biologiche, di cui è composto l'encefalo umano, e le reti neurali artificiali. Sono stati presentati dei modelli per i neuroni artificiali, accennando alle modalità con cui queste unità elementari combinano i segnali che gli pervengono in ingresso per mezzo di una somma pesata e li elaborino per mezzo di una funzione d'attivazione per fornire un potenziale d'uscita. Ci siamo soffermati anche su alcune possibili funzioni d'attivazione, spiegando in termini matematici il loro ruolo.

Abbiamo poi continuato presentando al lettore dei modelli di reti neurali noti come reti feedforward, in cui i segnali che alimentano la rete possono propagarsi dagli ingressi alle uscite in modo unidirezionale, attraversando uno o più strati di neuroni nascosti aventi la capacità di estrarre statistiche di ordine superiore per i segnali che la rete dovrebbe modellare. In questa classe di architetture abbiamo messo in luce le reti con connessioni jump, distinguendole dalle classiche feedforward per la capacità di trattare contemporaneamente le caratteristiche lineari e non lineari di un segnale senza la necessità di appoggiarsi a modelli esterni per valutare comportamenti lineari del segnale. Abbiamo inoltre accennato ad altre architetture neurali, come le reti ricorsive o basate su funzioni radiali, illustrandone brevemente il funzionamento.

Abbiamo presentato la teoria relativa all'algoritmo di back-propagation, spiegandone il principio di funzionamento e presentando un modello per l'apprendimento basato sull'aggiornamento dei pesi sinaptici secondo un parametro d'apprendimento costante. Infine, abbiamo descritto tre applicazioni fatte in letteratura sull'uso delle reti neurali nella predizione della glicemia futura, evidenziando come valga la pena di trovare un modello che, senza necessariamente appoggiarsi a modelli esterni, riesca a trattare contemporaneamente le dinamiche lineari e non lineari insite nel segnale glicemico.

Nel prossimo capitolo ci occuperemo più nello specifico delle jump net e spiegheremo come tale architettura verrà impiegata in questa tesi per la predizione della glicemia futura sfruttando informazioni provenienti da domini diversi, come per esempio quelli ottenuti da sensori CGM e quelli ottenuti per mezzo di modelli fisiologici dell'assorbimento del glucosio dopo pasto.

Capitolo 3

Approcci alla predizione della glicemia basati su jump-net

Come già anticipato nel Capitolo 2, le reti neurali possono essere un valido strumento per l'analisi e la predizione delle serie temporali. I dati di una serie temporale sono tipicamente ottenuti per campionamento di una funzione particolare, di cui molto spesso non si conosce una descrizione matematica completa. Le reti neurali, grazie all'apprendimento basato su training set e l'algoritmo di back-propagation, riescono a generare una funzione in grado di approssimare in maniera sufficientemente corretta i dati a disposizione. Questo risultato viene anche confermato dal teorema dell'approssimazione universale, applicabile alle reti neurali feedforward. Il teorema, tuttavia, non fornisce alcuna indicazione sulla topologia o l'architettura della rete adatta ad approssimare al meglio la serie temporale, ma è un buon punto di partenza per affrontare il problema della predizione. Abbiamo specificato nel Capitolo 2 che le reti neurali feedforward, come il MLP, presentano uno o più strati nascosti che permettono l'estrazione delle caratteristiche non lineari presenti nel segnale in analisi. S'è detto inoltre che l'architettura jump, in cui gli ingressi si collegano anche direttamente all'uscita, oltre che al primo strato nascosto, ha la capacità di estrarre, grazie ad un'unica rete, anche le componenti lineari della funzione valutata. In questo capitolo analizzeremo con maggiore dettaglio le reti jump, con particolare attenzione per quel che riguarda la scelta delle architetture che verranno utilizzate per la predizione della glicemia futura.

3.1 Architetture jump testate: tipologie "A" e "B"

Le jump net fanno parte delle reti feed forward, con la particolarità che le connessioni sinaptiche non si estendono solamente tra strati successivi di neuroni.

In Figura 3.1 viene presentata una generica rete feedforward realizzata attraverso il Neural Networks Toolbox fornito da Matlab. La rete qui illustrata presenta un unico strato nascosto di 8 neuroni aventi funzione d'attivazione a tangente sigmoideale e uno strato d'uscita con un unico neurone con funzione d'attivazione lineare. Come precedentemente indicato nel Paragrafo 2.2.1, la rete può essere quindi riassunta a mezzo della dicitura 8-1. Gli ingressi che la alimentano sono tre mentre l'uscita della rete è unica. Specifichiamo però che, avendo l'immagine la sola intenzione di fornire un'esemplificazione delle diverse architetture neurali, sia gli ingressi che le uscite sono intenzionalmente non definiti. Per concludere, dato che i neuroni sono tra di loro completamente connessi, la rete viene detta totalmente connessa.

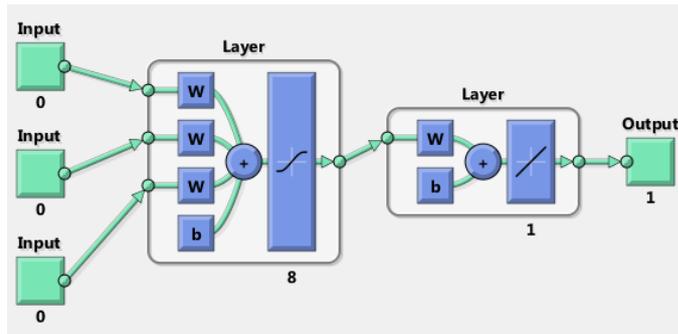


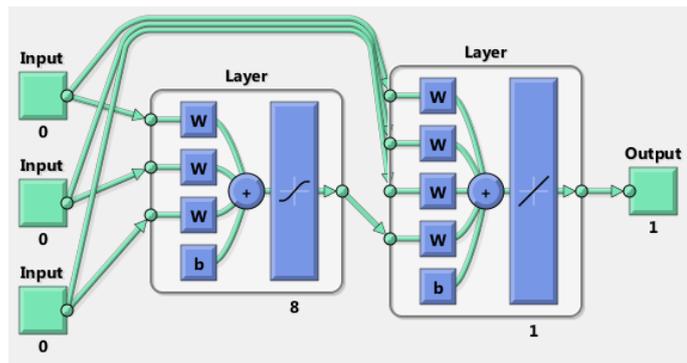
Figura 3.1: Esempio di generica rete feedforward

In Figura 3.2a presentiamo invece una prima classe delle reti jump, quella che definiremo jump-net di tipo A o più semplicemente jump-net A. In questo caso, come si evince dalla figura, gli ingressi, oltre che essere collegati come nel caso precedente allo strato nascosto, sono collegati direttamente allo strato d'uscita. La possibilità di queste architetture di estrarre sia le caratteristiche non lineari che lineari della serie temporale risiede nella diversità delle funzioni d'attivazione presenti nei due strati. Le componenti non lineari delle serie temporali infatti, come detto nel Paragrafo 2.2.2, vengono estratte dallo strato nascosto grazie anche alla presenza della funzione d'attivazione non lineare a tangente sigmoideale, mentre le componenti lineari del segnale vengono gestite grazie allo strato d'uscita che presenta funzione d'attivazione lineare. Questa classe di jump-net è quella presentata in [13].

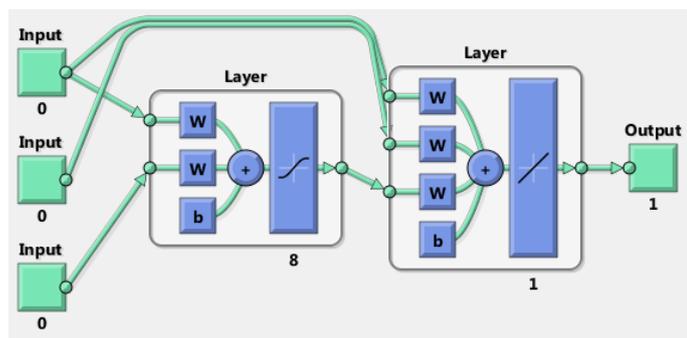
In questa tesi, tuttavia, viene fatto un passo ulteriore prendendo in considerazione una diversa tipologia di architettura, che noi chiameremo jump-net di tipo B, o più semplicemente jump-net B. In queste reti la topologia cambia in quanto gli ingressi non sono collegati necessariamente sia al primo strato nascosto sia all'uscita. Dalla Figura 3.2b infatti è possibile notare che:

1. il primo ingresso presenta delle connessioni sia allo strato nascosto che allo strato d'uscita
2. il secondo ingresso non presenta più il collegamento allo strato nascosto, che viene completamente bypassato, mentre la sinapsi diretta al livello d'uscita viene mantenuta
3. il terzo ingresso, infine, non presenta più la sinapsi jump ma solamente quella diretta allo strato nascosto.

Successivamente spiegheremo nel dettaglio il metodo che abbiamo adottato per la scelta di una tipologia piuttosto che di un'altra. Per ora possiamo comunque dire che la decisione di collegare un ingresso al solo strato d'uscita nasce dall'osservazione, o dalla conoscenza a priori, che tale ingresso contribuisce all'uscita della rete solo con termini lineari.



(a) Jump-net A



(b) Jump-net B

Figura 3.2: Un esempio di possibili architetture jump, nelle tipologie da noi etichettate come A e B

3.2 Possibili ingressi utilizzabili nella predizione della glicemia mediante reti neurali

Le reti neurali, come anticipato nel Capitolo 2, hanno il vantaggio di poter integrare all'interno dello stesso modello ingressi e uscite di domini diversi. Nel caso della predizione della glicemia futura questo aspetto risulta particolarmente interessante, in quanto sarebbe possibile sfruttare informazioni che non derivano direttamente dal sensore CGM, ma anche ulteriori conoscenze, come quelle relative ai pasti. In questa tesi sono stati presi in considerazione diverse tipologie d'ingresso, che andiamo ora ad presentare.

3.2.1 Informazioni sulla glicemia attuale, derivata ed integrale.

Il segnale di glicemia ottenuto attraverso sensore CGM è il principale segnale considerato per reti studiate nel presente elaborato. Si presenta come una serie temporale di misure ottenute con passo di campionamento costante. In questa tesi sono stati utilizzati sia dati ottenuti da simulatore che dati reali raccolti con sensore Dexcom[®] 7-Plus con passo di campionamento pari a 5 minuti. In [24] la serie temporale era campionata a 1 minuto mediante sensore FreeStyle Navigator[™]. Oltre al dato relativo all'attuale misura di glicemia, tra gli ingressi candidati è stata considerata anche la serie delle differenze prime effettuata sui 15 minuti precedenti l'istante di misura, e la serie integrale della glicemia, anch'essa valutata sui 15 minuti precedenti l'istante di misura. In particolare, per le queste ultime due serie temporali è stata utilizzata questa formulazione

$$dCGM(t) = CGM(t) - CGM(t - 3Ts) \quad (3.1)$$

$$iCGM(t) = \sum_{k=1}^3 CGM(t - kTs) \quad (3.2)$$

dove con Ts si intende il periodo di campionamento.

In figura 3.3 possiamo vedere un esempio di questi tre ingressi per il caso di una serie glicemica, in questo caso simulata, con campionamento a 5 minuti

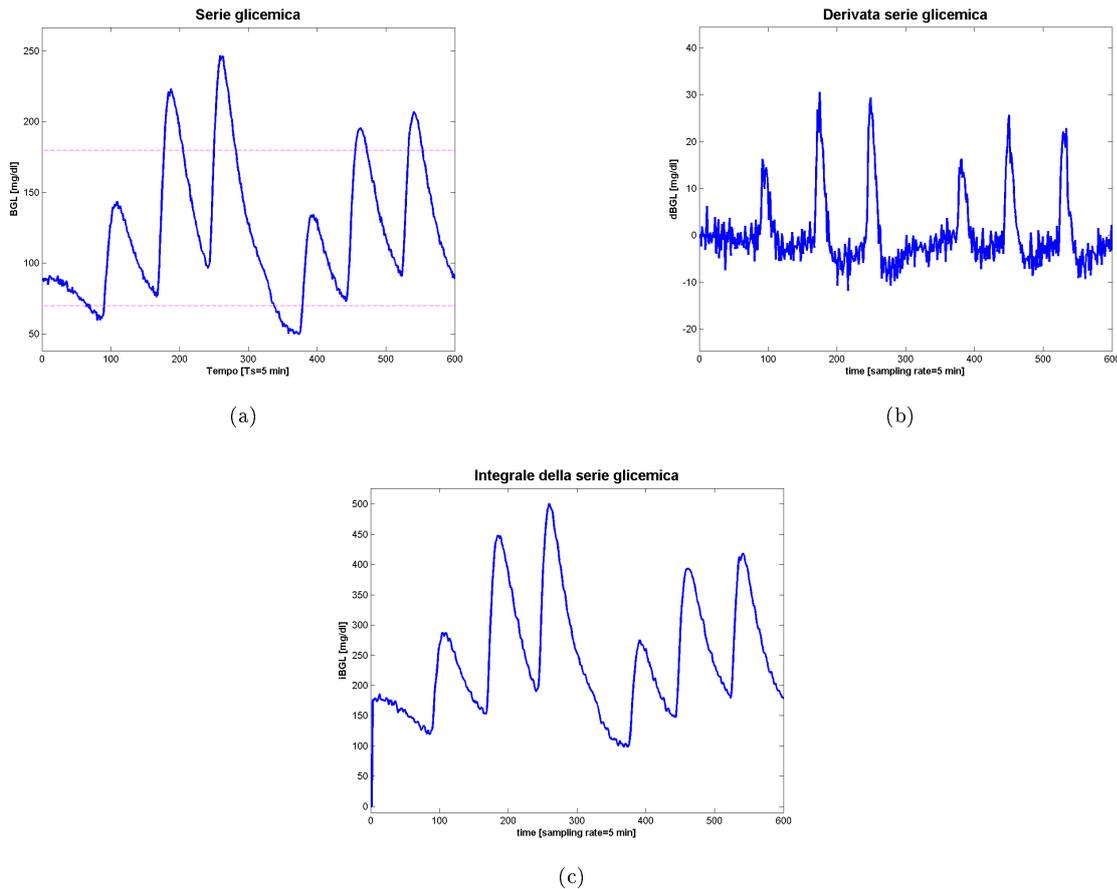


Figura 3.3: Serie temporali simulate relative alla glicemia; (a) la serie glicemica campionata a 5 minuti; (b) la serie delle differenze prime; (c) la serie integrata.

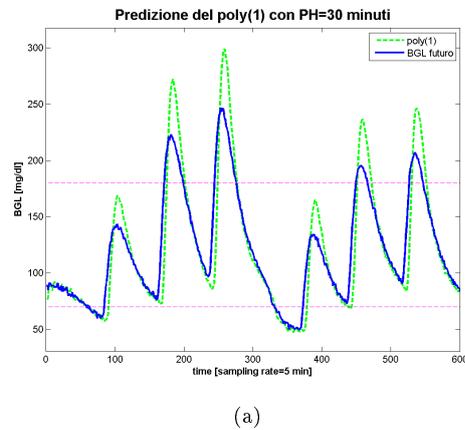
3.2.2 Predizione della glicemia futura mediante modello poly(1)

Sebbene le jump-net siano in grado di estrarre oltre che le caratteristiche non lineari del segnale in analisi quelle lineari, tra gli ingressi candidati sono stati presi in considerazione anche dati ottenuti per mezzo di modello poly(1) proposto in [21]. Per la predizione con modello polinomiale del primo ordine è stato utilizzato un coefficiente d'oblio μ tale da minimizzare l'indice J, definito in [26]. L'orizzonte di predizione è stato scelto pari a 30 minuti, in accordo con i risultati di [24]. Oltre alla predizione della glicemia futura ottenuta con poly(1), come possibile segnale d'ingresso per la rete è stato scelto anche l'errore commesso dal polinomiale all'istante attuale, ottenuto secondo la formulazione

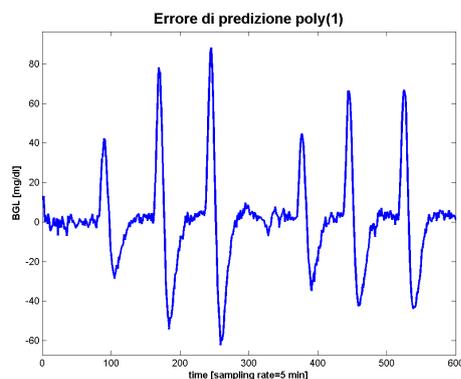
$$e(t) = \widehat{CGM}_t(t) - CGM(t) \quad (3.3)$$

$$de(t) = e(t) - e(t - 3Ts) \quad (3.4)$$

in cui $\widehat{CGM}_t(t)$ sta ad indicare la stima ottenuta attraverso modello polinomiale 30 minuti prima e relativo all'istante attuale, mentre $CGM(t)$ indica l'attuale misura del sensore CGM.



(a)



(b)

Figura 3.4: Serie temporali ottenute dal predittore lineare; (a) Predizione del modello polinomiale ottenuta con orizzonte di predizione PH=30 minuti (linea tratteggiata) e serie glicemica simulata target (linea continua); (b) Errore del predittore poly(1) calcolato secondo la 3.3.

3.2.3 Modello per l'assorbimento del glucosio dopo pasto: la rate of appearance

Dalla fisiologia è ben noto che la quantità di carboidrati ingeriti durante un pasto è direttamente correlata con la glicemia. Si è ritenuto quindi opportuno considerare anche questa informazione come ingresso utile per la rete neurale. Si potrebbe pensare, erroneamente, che un'informazione impulsiva sulla quantità di carboidrati ingeriti durante il pasto sia un segnale sufficiente per la rete, ma così non è. Bisogna ricordare infatti che una rete feedforward non conserva memoria dello stato passato del sistema, quindi un semplice ingresso impulsivo in corrispondenza dei pasti, proporzionale alla quantità di carboidrati ingeriti, risulterebbe inefficace. L'effetto che i carboidrati hanno sul metabolismo non è immediato e nemmeno si esaurisce immediatamente dopo il pasto. Per superare queste difficoltà s'è deciso quindi di utilizzare la rate of appearance (ra) dei carboidrati ingeriti, ricavata grazie ad un modello fisiologico [27] con parametri di popolazione stimati in [25] al quale vengono fornite in ingresso informazioni sulla

quantità di carboidrati ingeriti e sul peso del paziente. La rate of appearance dei carboidrati può quindi essere generata on-line semplicemente fornendo alla funzione la quantità di carboidrati assunti. Da qui nasce la possibilità di sfruttare sia il valore attuale della rate of appearance dei carboidrati, che il valore futuro. Oltre alla serie temporale della rate of appearance, sono state considerate anche la serie delle differenze prime calcolata sui 15 minuti precedenti al valore di ra considerato (sia esso quello attuale o quello futuro) e la serie integrale, anch'essa calcolata sui 15 minuti che precedono l'istante temporale in uso. In Figura 3.5 vediamo un esempio della serie temporale, della serie delle differenze prime e della serie integrale. Nello specifico le due serie derivate dalla rate of appearance vengono così stimate:

$$dra_G = ra_G(t) - ra_G(t - 3Ts) \quad (3.5)$$

$$ira_G = \sum_{k=1}^3 ra_g(t - kTs) \quad (3.6)$$

con Ts che ancora una volta rappresenta il periodo di campionamento.

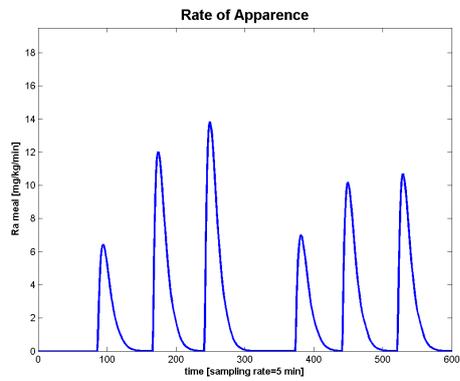
3.3 Jump-net A: modello base

Le reti jump utilizzate in questa tesi presentano un'architettura pressoché simile. Ruolo importante nella differenziazione lo giocano il numero e la tipologia degli ingressi, oltre al numero di neuroni degli strati nascosti. I modelli sono invece caratterizzati da una struttura comune composta di un solo strato nascosto, i cui neuroni avranno sempre funzione d'attivazione a tangente sigmoidale, e un unico strato d'uscita, avente un neurone con funzione d'attivazione lineare. Per quanto concerne le sinapsi, essendo i modelli candidati delle jump-net A gli ingressi sono connessi direttamente sia allo strato nascosto che allo strato d'uscita. Infine per tutte le reti è stato scelto un comune segnale target rappresentato dal valore glicemico ad un dato orizzonte temporale (*Prediction Horizon*, PH) futuro, che in questa tesi è stato scelto di 30 minuti, in accordo con quanto presentato in [24]. La scelta del numero e della tipologia degli ingressi, nonché del numero di neuroni dello strato nascosto, sono stati ottenuti valutando 4 parametri (RMSE, J, ESOD, TG) che spiegheremo successivamente nel Paragrafo 3.6.

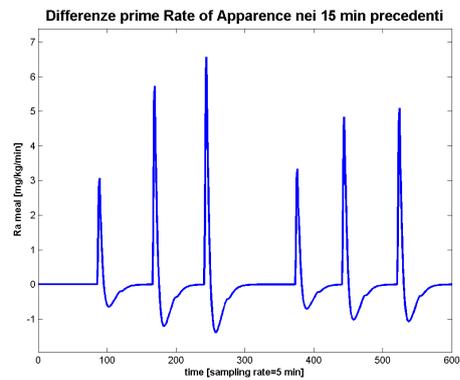
3.4 Algoritmo di *cross-validation* per la scelta della struttura ottimale

Per scegliere la struttura delle reti e degli ingressi sono state valutate le performance di vari modelli candidati. Data la natura stocastica delle reti neurali, è stato utilizzato un algoritmo di *k-fold cross validation* per l'analisi delle performance. In questo algoritmo il training set è stato diviso in \mathbf{k} sottoinsiemi disgiunti, contenenti ognuno lo stesso numero di coppie ingresso-uscita.

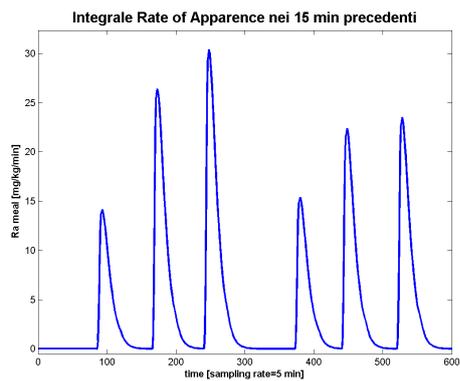
Nel nostro caso sono stati utilizzati 5 sottoinsiemi, ciascuno dei quali presenta 6 serie temporali della durata di 3 giorni (1728 campioni). La rete viene quindi allenata su $\mathbf{k}-1$ sottoinsiemi e testata sul sottoinsieme rimanente. Tale procedura viene quindi ripetuta per \mathbf{k} volte, utilizzando ogni volta come nuovo test set il sottoinsieme escluso dall'allenamento. Per ciascuna rete, questo procedimento viene ripetuto per 8 volte e al termine le performance globali della rete in analisi vengono calcolate come media delle performance ottenute nei \mathbf{k} test set. Si è scelto di testare ciascun modello candidato incrementando linearmente il numero di neuroni dell'unico strato nascosto a partire da 1 fino ad arrivare a 12 neuroni.



(a)



(b)



(c)

Figura 3.5: Serie temporali relative alla rate of appearance; (a) la serie relativa RA_G ; (b) la serie delle sue differenze prime; (c) la serie della rate of appearance integrata.

3.5 Metriche di valutazione delle performance della rete

Le performance delle reti candidate alla predizione della glicemia futura sono state valutate in questo elaborato per mezzo principalmente di 4 indici: il *Root Mean Square Error* (RMSE), il J [26], il guadagno temporale (*Time Gain* o TG) e l'energia associata alle derivate seconde del profilo predetto (*Energy of Second Order Derivative* o ESOD). Ciascun parametro presenta dei vantaggi, ma anche dei limiti. Per poter quindi fornire un giudizio finale soddisfacente è necessario considerare i parametri contemporaneamente, non prescindendo però da un'analisi visiva del profilo predetto per validarne l'effettiva bontà.

3.5.1 Definizione dell' RMSE

Il RMSE è l'indice maggiormente utilizzato in letteratura per valutare la bontà del profilo predetto, ma nel caso delle reti neurali è anche l'indice che tipicamente viene minimizzato durante l'allenamento. L'indice esprime la distanza tra il segnale predetto e il segnale target ed è definito

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}^{(i)} - y_{\text{CGM}}^{(i)})^2}$$

dove $\hat{y}^{(i)}$ indica l' i -esimo campione del segnale predetto, mentre $y_{\text{CGM}}^{(i)}$ è l' i -esimo campione del segnale target e N la lunghezza della serie temporale (cioè il numero totale di campioni). Come nel caso di J, anche qui più l'indice è prossimo allo 0 più la predizione segue il segnale target. Un problema legato a questo indice è che non tiene conto della regolarità del profilo predetto, per cui segnali anche molto rumorosi ma prossimi al target presentano RMSE minore di segnali perfettamente aderenti al target ma affetti da un piccolo bias o da un ritardo temporale ininfluenza

3.5.2 Guadagno temporale medio

Il TG medio misura l'anticipo medio con cui riusciamo a predire il segnale target ed è definito come

$$\text{TG} = \text{PH} - \text{ritardo}_{\text{medio}}$$

dove con $\text{ritardo}_{\text{medio}}$ indichiamo uno shift del profilo predetto tale da minimizzare il RMSE. Il principale vantaggio di questo parametro risiede nella semplicità della sua formulazione. Tuttavia presenta dei limiti, alcuni dei quali sono gli stessi presentati per il RMSE. Per il calcolo del guadagno temporale viene effettuato un pre-filtraggio per limitare le oscillazioni ad alta frequenza poiché, talvolta, la presenza di tale rumore rende poco veritiero il guadagno temporale.

3.5.3 Energia delle derivate seconde:

L' $\text{ESOD}_{\text{norm}}$ è definito come [28]

$$\text{ESOD}_{\text{norm}} = \frac{\text{ESOD}_{\text{pred}}}{\text{ESOD}_{\text{target}}}$$

dove con $\text{ESOD}_{\text{pred}}$ indichiamo l'energia del profilo predetto, mentre con $\text{ESOD}_{\text{target}}$ l'energia del profilo target, entrambi ottenuti mediante

$$\text{ESOD} = \frac{1}{N} \sum_{t=1}^N (d_{2t})^2$$

dove con d_{2t} si indica la derivata seconda del segnale

$$d_{2t} = \frac{y_t - 2y_{t-T_s} + y_{t-2T_s}}{T_s^2}$$

L'energia delle derivate seconde è un indice è in grado di misurare la regolarità di un segnale. Nel nostro caso, tuttavia, non siamo interessati direttamente all'energia del segnale predetto, quanto piuttosto che il segnale predetto il target presentino lo stesso grado di regolarità. In tal senso, quanto più $ESOD_{norm}$ è prossima a 1, tanto più la predizione è buona e regolare quanto il target. Se $ESOD_{norm} < 1$, la predizione risulta sovra regolarizzata, mentre se $ESOD_{norm} > 1$ la predizione contiene delle oscillazioni spurie non presenti nel segnale target.

3.5.4 Regolarità del profilo predetto: J

L'indice J è definito come [26]

$$J = \frac{ESOD_{norm}}{TG_{norm}^2}$$

Dove $ESOD_{norm}$ è il rapporto tra l'energia delle derivate seconde del profilo predetto e l'energia delle derivate seconde del segnale target mentre TG_{norm} indica il guadagno temporale normalizzato, cioè il rapporto tra il guadagno temporale medio e l'orizzonte di predizione. Il profilo predetto, come spiegato in [28, 26], per essere soddisfacente deve presentare una buona regolarità e un buon guadagno temporale. Come visto in precedenza, però, per ottenere tale risultato è necessario trovare un compromesso tra l' $ESOD_{norm}$ e il TG_{norm} , in quanto questi due indici risultano tipicamente inversamente correlati, cioè quando aumenta il TG diminuisce la regolarità del profilo predetto, e viceversa.. L'indice J è stato studiato proprio per cercare di considerare allo stesso tempo entrambi i parametri e valutare la regolarità del profilo predetto attraverso un unico indice. Nello specifico, "l'utilità clinica" del segnale predetto è tanto più alta quanto più l'indice J è prossimo allo 0.

3.6 Le reti candidate valutate

La fase preliminare di questo lavoro di tesi ha visto la creazione di 14 reti neurali aventi tutte l'architettura precedentemente descritta, ma diverse combinazioni d'ingressi. Una prima scrematura delle reti candidate è stata fatta tenendo fisso il numero di neuroni dello strato nascosto. Nello specifico il numero di neuroni nascosti è 8, per adeguarsi alla rete feedforward presentata in [16]. Questa prima fase ha portato alla scelta di 6 jump-net A, di cui riportiamo i risultati in termini di RMSE e TG nella tabella [rif. tabella]. Queste reti sono state allenare e testate utilizzando un dataset simulato che descriveremo nel dettaglio nel capitolo successivo. Successivamente, per ognuna di queste reti è stato applicato l'algoritmo di k-fold cross validation, con l'obiettivo di individuare il numero di neuroni nascosti ottimale (ovvero tale da realizzare un equo compromesso tra gli indici di valutazione RMSE, J, ESOD, TG). Nelle Figure {3.6 – 3.11}

possiamo vedere i grafici che illustrano i risultati per ciascuna rete, mentre in tabella

3.1 vengono presentati il numero di neuroni nascosti selezionato per ciascuna delle reti candidate.

In alcuni casi, come per esempio quello della NNJA5 (Figura 3.7) , NNJA8 (Figura 3.8) , NNJA13 (Figura 3.11), non tutti gli indici valutati raggiungono il valore ottimo in corrispondenza dello stesso numero di neuroni nascosti. In questa situazione abbiamo deciso di scegliere il numero di neuroni nascosti minimo che fornisce un buon compromesso per i vari indici. Questo nel tentativo di ridurre la complessità delle

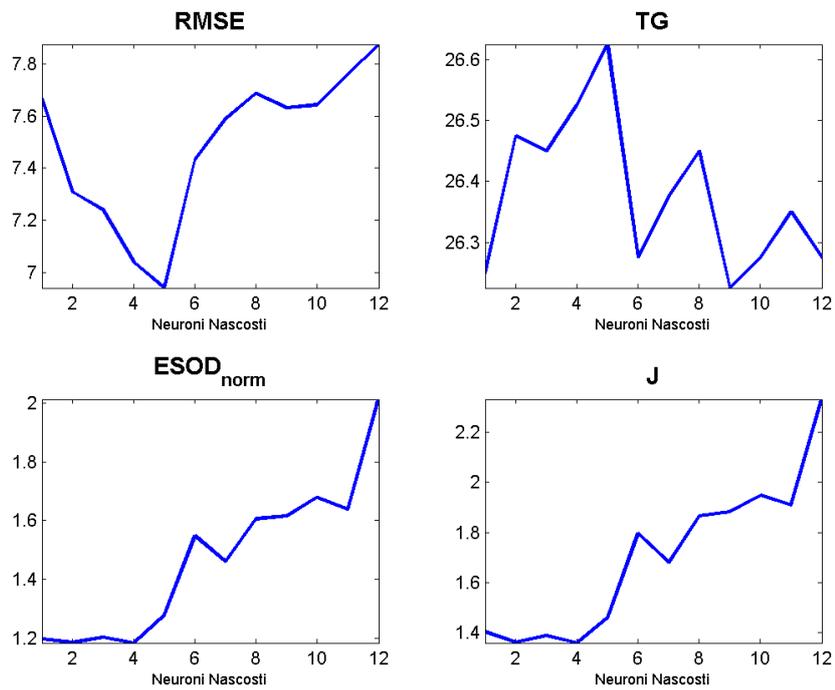


Figura 3.6: Risultati della cross validazione: rete NNJA2. (a) RMSE; (b) TG; (c) ESOD_{norm}; (d) J

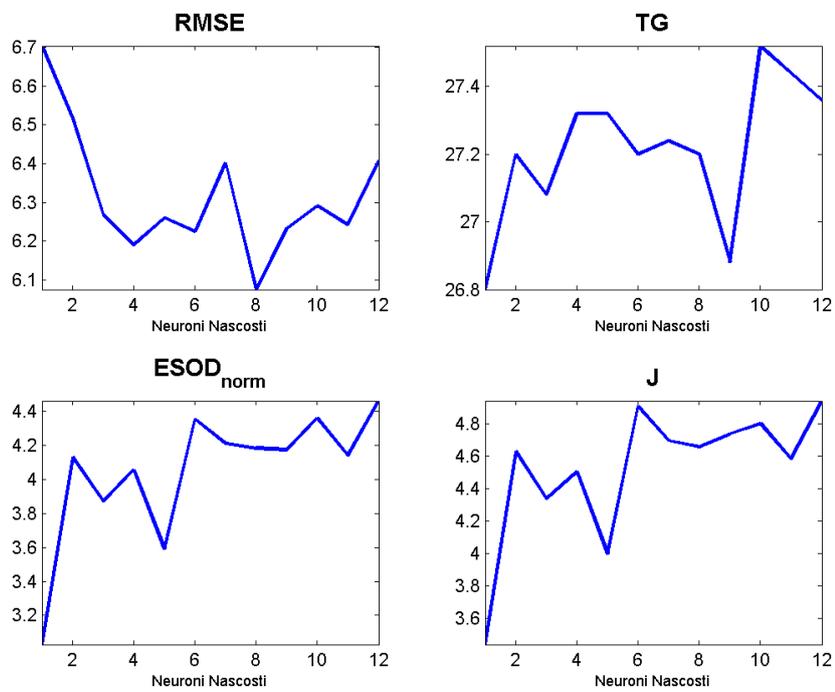


Figura 3.7: Risultati della cross validazione: rete NNJA5. (a) RMSE; (b) TG; (c) ESOD_{norm}; (d) J

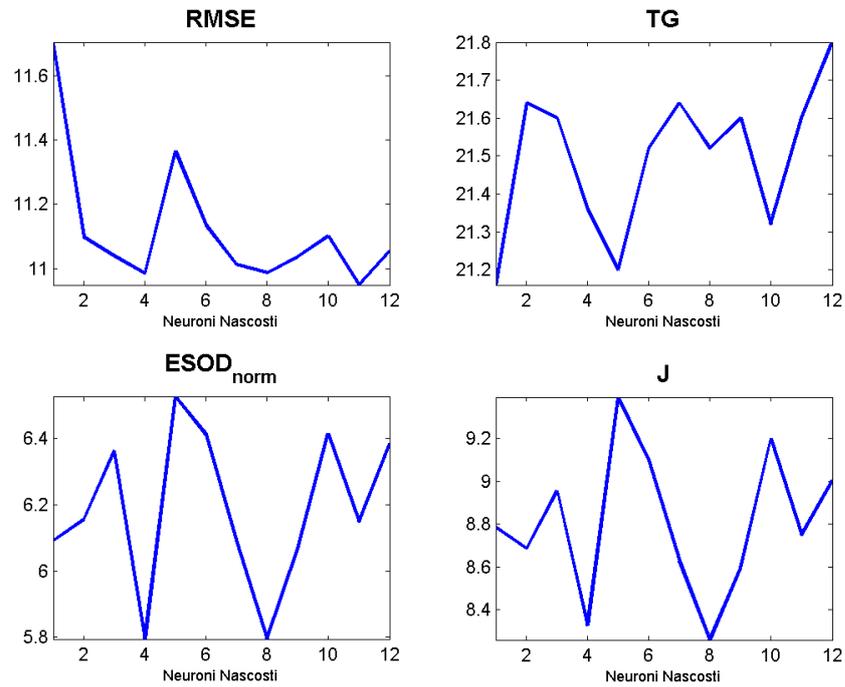


Figura 3.8: Risultati della cross validazione: rete NNJA8. (a) RMSE; (b) TG; (c) ESOD_{norm}; (d) J

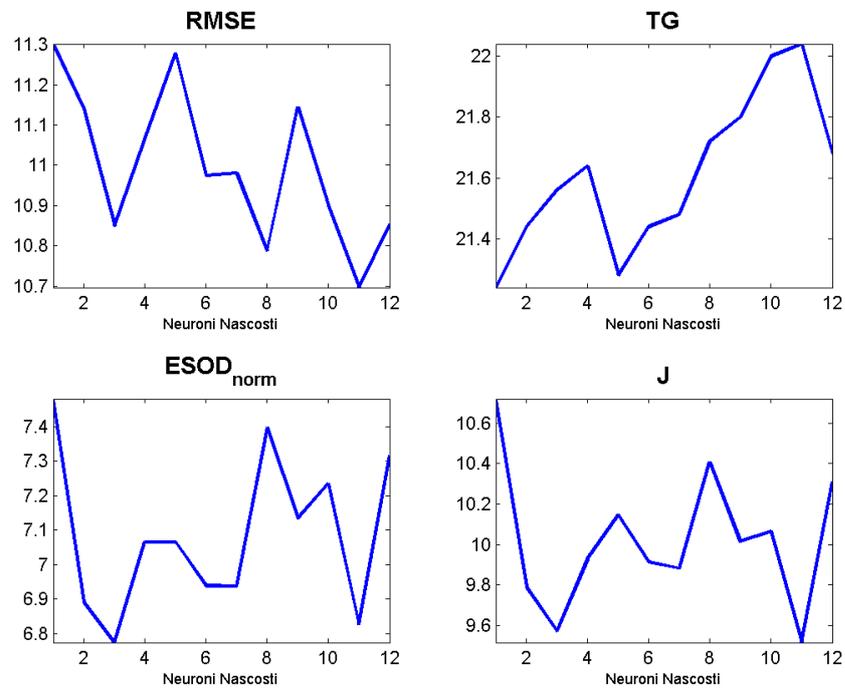


Figura 3.9: Risultati della cross validazione: rete NNJA12. (a) RMSE; (b) TG; (c) ESOD_{norm}; (d) J

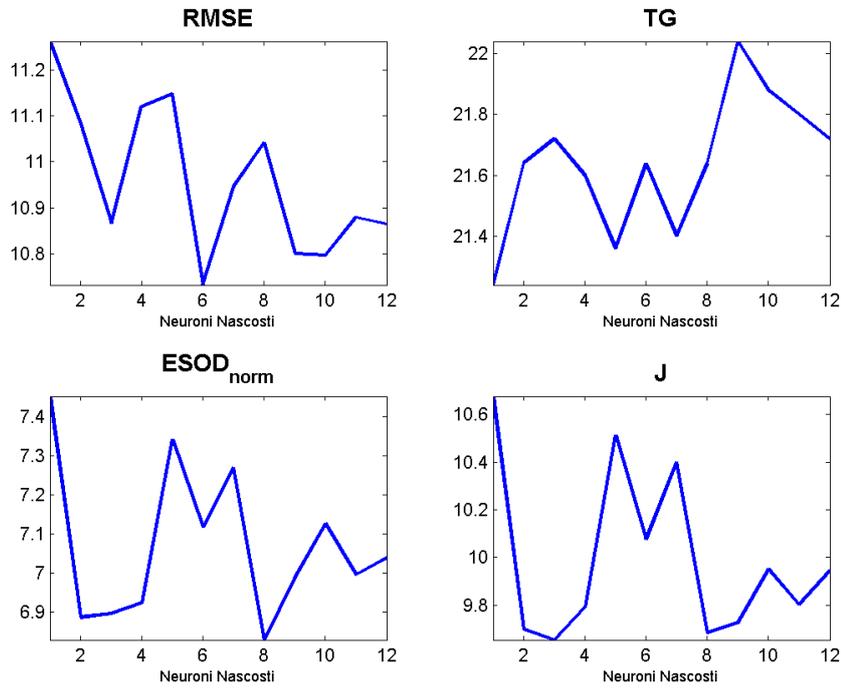


Figura 3.10: Cross Validazione: rete NNJA14

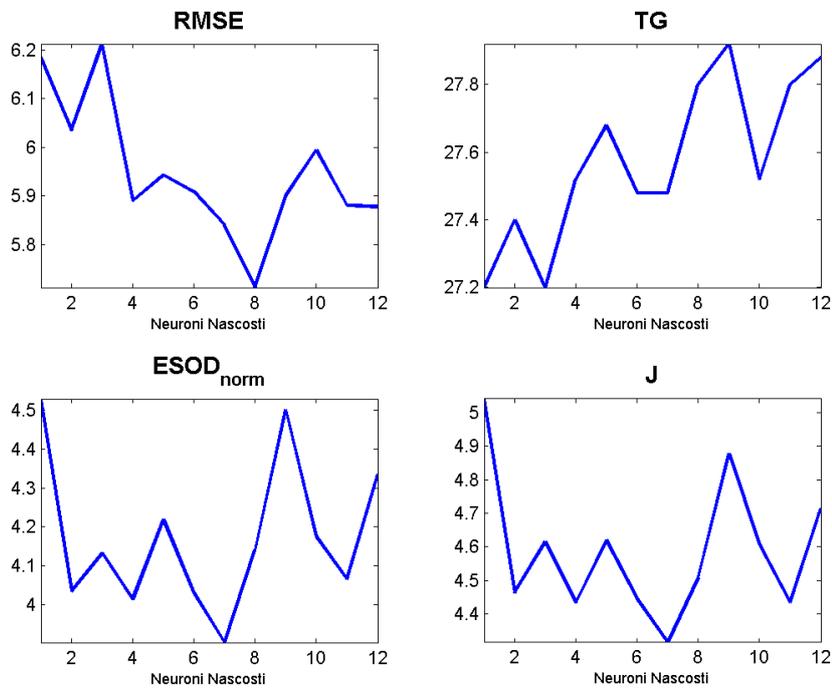


Figura 3.11: Risultati della cross validazione: rete NNJA13. (a) RMSE; (b) TG; (c) ESOD_{norm}; (d) J

| Rete | n° ottimo di neuroni nascosti |
|--------|-------------------------------|
| NNJA2 | 5 |
| NNJA5 | 5 |
| NNJA8 | 8 |
| NNJA12 | 11 |
| NNJA14 | 9 |
| NNJA13 | 8 |

Tabella 3.1: Numero di neuroni definitivi scelti per lo strato nascosto per le reti candidate

reti senza inficiare in modo determinante i risultati ottenuti. È necessario fare un'ulteriore precisazione riguardo le reti NNJA8, NNJA12, NNJA14. Queste, infatti, sono state scartate dopo l'analisi di k-fold cross validazione in quanto, in corrispondenza del numero ottimo di neuroni nascosti, presentavano dei risultati significativamente peggiori rispetto a quelli ottenuti dalle altre reti candidate, testate anch'esse con numero di neuroni nascosti ottimo.

3.7 Da jump-net A a jump-net B: un possibile metodo di scelta

Una volta individuate le reti candidate, che verranno impiegate successivamente per la predizione della glicemia futura in ambiente simulato e con misure ottenute da pazienti reali, è necessario fornire un metodo per trasformare queste jump-net A nelle corrispondenti jump-net B, che non presentano tutte le connessioni sinaptiche verso lo strato nascosto o lo strato d'uscita. Il metodo da noi impiegato per compiere la scelta s'è basato sull'analisi della matrice dei pesi sinaptici, rappresentativa della "forza" di ciascuna connessione sinaptica. Quanto più il valore associato ad una sinapsi è prossimo allo 0, tanto minore è il peso di quella sinapsi nel determinare l'uscita della rete neurale. Attraverso un'analisi visiva delle matrici dei pesi sinaptici siamo stati quindi in grado di formulare delle nuove jump-net, questa volta della tipologia B. Nello specifico il procedimento da noi adottato segue questa procedura:

1. Per ciascuna rete interessata viene effettuata un'analisi qualitativa/quantitativa della matrice dei pesi associati a ciascuna sinapsi neurale;
2. si individuano i pesi che, in media, risultano meno significativi dal punto di vista quantitativo;
3. iterativamente, ciascun peso candidato viene posto pari a zero e contemporaneamente la rete jump così modificata viene testata senza che essere riallenata. I pesi successivamente vengono posti a zero in combinazioni diverse o tutti assieme, e la rete viene nuovamente testata senza riallenamento;
4. per ciascuna rete viene effettuato uno studio visivo dei profili predetti ed una valutazione delle qualità medie della rete in termini di RMSE, Time Gain, ESOD_{norm} e J;
5. le reti che forniscono durante la fase precedente i migliori risultati vengono scelte come reti jump B candidate. Le reti vengono riallenate dopo aver fisicamente eliminato la connessione sinaptica "superflua" e testate nuovamente secondo gli indici qualitativi e un'ispezione visiva dei profili al fine di individuare differenze con la tipologia A.

3.8 Le jump-net candidate

Di tutte le reti che sono state valutate nelle fasi preliminari illustrate in questo capitolo, abbiamo selezionato tre reti che si sono distinte per le loro prestazioni in termini di RMSE, Time Gain, Esod e J. Andiamo ora a presentarle, indicando il numero di neuroni nascosti impiegati e il numero/tipologia degli ingressi che alimentano le reti stesse

NNJA2

Si tratta della rete più semplice tra quelle selezionate. Gli ingressi alla rete sono tre:

1. La glicemia ottenuta all'istante attuale dalla serie temporale simulata;
2. Il valore futuro della rate of appearance del glucosio calcolato all'orizzonte temporale di 30 minuti; bisogna notare che questo dato è ottenibile proprio grazie al modello indicato in 3.2.3 a partire dalla sola quantità di carboidrati ingeriti durante il pasto;
3. La serie delle differenze prime relative alla rate of appearance futura, calcolata come in 3.5 dove invece di considerare come t l'istante attuale, viene considerato il valore della rate of Appearance a $t=30$.

La rete presenta uno strato nascosto con 5 neuroni, una buona soluzione di compromesso tra l'aderenza ai dati, la regolarità del profilo predetto e il guadagno temporale ottenibile.

NNJA5

Quattro ingressi alimentano questa rete, ma le differenze con la NNJA2 risultano minime. I primi tre ingressi coincidono infatti per le due reti, mentre il quarto è costituito dalla serie delle differenze prime relative alla glicemia attuale, calcolata come in 3.1.

Anche questa rete è composta di 5 neuroni nascosti, come indicato in tabella 3.1;

NNJA13

L'ultima rete proposta nasce dall'idea di disporre di un'architettura in qualche modo simile a quella proposta in [24]. Ecco allora che in questo caso la rete presenta 6 ingressi:

1. La serie temporale dell'errore commesso dal modello polinomiale come indicato nell'equazione 3.3;
2. La serie delle differenze prime relative all'errore del polinomiale, come in 3.4;
3. Il valore futuro della rate of appearance del glucosio, calcolato nuovamente con l'orizzonte temporale fissato a 30 min;
4. La serie delle differenze prime della rate of appearance futura, calcolata come 3.5;
5. La misura all'istante attuale della glicemia prelevata dalla serie temporale simulata;
6. La serie delle differenze prime della glicemia misurata all'istante attuale, ottenuta per mezzo di 3.1.

La rete NNJA13 utilizza un numero di neuroni nascosti pari a 8, in linea con quanto verificato in [24]. È utile ricordare che, come già anticipato in precedenza, tutte le reti predicono il valore della glicemia futura con un'orizzonte temporale di 30 min. La rete NNJA13 presenta quindi degli ingressi simili alla rete proposta da Zecchin et al. in [24] ma predice un target completamente diverso, per cui non possiamo di fatto considerarla come la "versione jump" della rete studiata in [24].

3.9 Conclusioni

In questo capitolo abbiamo concentrato la nostra attenzione sulle jump-net, architettura neurale di tipo feedforward caratterizzata dalla presenza di connessioni sinaptiche non soltanto tra gli ingressi e lo strato di neuroni nascosti, ma anche tra gli ingressi e il livello d'uscita. Abbiamo visto due diverse tipologie di reti jump, da noi denominate jump A (reti completamente connesse) e jump B (reti solo parzialmente connesse). Sono stati poi presentati alcuni ingressi utilizzabili nella predizione della glicemia attraverso reti neural jump, tra cui abbiamo presentato anche il segnale relativo alla glucose rate of appearance, velocità con cui il glucosio compare nel plasma. Si è passati poi alla scelta delle reti jump candidate attraverso l'impiego di un algoritmo di k-fold cross validation che, data la natura stocastica delle reti, c'ha permesso di valutarne le performance in termini di RMSE, TG, ESOD_{norm} e J, presentati anch'essi in questo capitolo. Di tutte le reti testate, tre hanno superato questa fase preliminare e sono state utilizzate per il test su data set simulato e su data set reale: NNJA2, NNJA5, NNJA13. Infine, abbiamo messo in luce un possibile metodo per passare dall'architettura A all'architettura B, attraverso un'analisi qualitativa della matrice dei pesi sinaptici e l'eliminazione dei pesi mediamente meno significativi.

Nel prossimo capitolo, le reti NNJA2, NNJA5 e NNJA13 verranno allenate e testate su data set simulato con campionamento a 5 minuti e confrontate con altri due modelli stato dell'arte: poly(1) proposto da Sparacino et al. [21] e NN-LPA, rete neurale proposta da Zecchin et al. in [24].

Capitolo 4

Data Sets

4.1 Creazione del data-set simulato

Come vedremo nel Capitolo 5, le reti jump candidate per la predizione della glicemia futura sono state inizialmente allenate e testate su un data base simulato, costruito per mezzo del simulatore UVA/PD proposto da Dalla Man et al. in [25] e approvato dall’FDA nel 2008. 66 soggetti sono stati simulati per 7 giorni consecutivi, con misure glicemiche ogni 5 minuti. Il protocollo di simulazione ha previsto 3 pasti giornalieri con variabilità random per dosi e orari, al fine di rendere tali simulazioni il più realistiche possibili. La colazione avviene nell’intervallo compreso tra le 6 e le 8 di mattina e fornisce una quantità di carboidrati pari a 45 ± 10 g. Il pranzo avviene nell’intervallo compreso tra le 12 e le 14 e consiste di 75 ± 15 g di carboidrati. Infine, la cena avviene nella fascia oraria tra le 19 e le 21 e l’apporto di carboidrati è pari a 85 ± 15 g. Il simulatore sfrutta anche le informazioni relative alla terapia insulinica che, nel nostro caso, risulta essere un trattamento con infusione basale d’insulina a cui si affiancano dei boli in concomitanza dei pasti, boli il cui dosaggio è studiato in modo da controbilanciare l’effetto dei carboidrati ingeriti. Occasionalmente, la quantità di insulina iniettata è stata aumentata o diminuita in maniera realistica, per riprodurre gli errori che possono accadere normalmente nella vita quotidiana.

Il Data Set così costituito è stato poi suddiviso in due ulteriori gruppi utilizzati separatamente per l’allenamento e il test delle reti jump. Nello specifico, delle 66 serie temporali simulate per 7 giorni consecutivi, 30 sono state utilizzate per costruire un training set specifico per l’allenamento mentre le restanti 36 sono state impiegate nella creazione di un test set. Durante la fase di allenamento, il training set è stato ulteriormente suddiviso tramite il Neural Network Toolbox di Matlab al fine di ottenere dei buoni risultati di cross validazione: il 70% dei dati sono stati impiegati per il training mentre il restante 30% per la validazione delle reti allenate. Il *validation set* viene utilizzato in fase di training per evitare di incorrere nel problema dell’overfitting: quando la qualità dei risultati sul validation set diminuisce, infatti, l’allenamento della rete viene interrotto; in questo modo si cerca di garantire una migliore capacità di generalizzazione dei risultati, senza che la rete si adatti troppo ai dati presentati, e al rumore ad essi sovrapposto. In Figura 4.1 possiamo vedere un esempio di profilo utilizzato

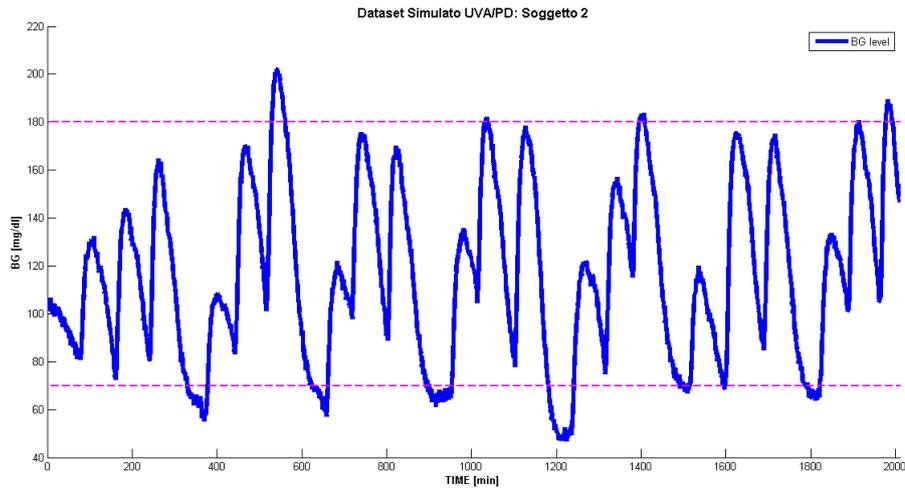


Figura 4.1: Esempio di profilo glicemico costruito con simulatore UVA/PD [25], paziente virtuale n° 2; linea continua blu: profilo glicemico; linea tratteggiata magenta: soglie di ipo/iperglicemia

4.2 Data Sets reali

4.2.1 Data Set reale Dexcom® 7-Plus (campionamento 5 min)

Nel Capitolo 6 verrà presentata l'analisi delle reti jump effettuata su un primo data set reale, ottenuto dal monitoraggio di pazienti diabetici effettuato per mezzo di sensore Dexcom® 7-Plus con campionamento a 5 minuti. Nello specifico, il monitoraggio ha interessato 9 soggetti per 5 giorni consecutivi, durante i quali il paziente ha annotato informazioni relative al proprio stile di vita, in particolare orario e quantità di carboidrati ingeriti e dettagli sulla terapia insulinica. È quindi stato possibile, grazie al simulatore presentato in sezione 4.1, calcolare il valore futuro della rate of appearance dei carboidrati nel plasmatico, ingresso necessario a tutte le reti jump.

Per massimizzare le capacità di generalizzazione e predizione delle reti, a fronte di un numero ridotto di dati a disposizione è stato deciso di suddividere ciascuna delle 5 serie temporali in un *training-set* della durata di 3 giorni e in un *test-set* della durata di 2 giorni. In questo modo è stato possibile allenare e testare le reti su tutti i pazienti a disposizione.

Sebbene nel Capitolo 6 si potrà trovare un'analisi completa delle performance delle reti allenare e testate su questo primo data set reale, sentiamo fin da ora la necessità di dire il numero ridotto di dati a disposizione ottenuti da soggetti reali, soprattutto se confrontato con le informazioni che a disposizione con il dataset simulato, non facilita soprattutto l'allenamento delle reti neurali, che per loro natura necessiterebbero di una grande quantità di dati per riuscire a ottenere delle prestazioni ottimali. In Figura 4.2 possiamo osservare un esempio di un profilo glicemico raccolto in questo dataset.

4.2.2 Data Set reale Abbott Navigator™ (campionamento 1 min)

Nel Capitolo 7, le reti jump candidate verranno infine analizzate nelle performance utilizzando un secondo data set reale, ottenuto in questo caso attraverso sensore Abbott Navigator™, in grado di fornire una misura di glicemia al minuto. Questa scelta è stata fatta al fine di completare l'analisi prestazionale delle reti candidate, utilizzando un dataset già presentato in letteratura in [24]. In Figura 4.3 viene presentato un esempio di profilo ottenuto con sensore Abbott Navigator™ con campionamento ad 1 minuto.

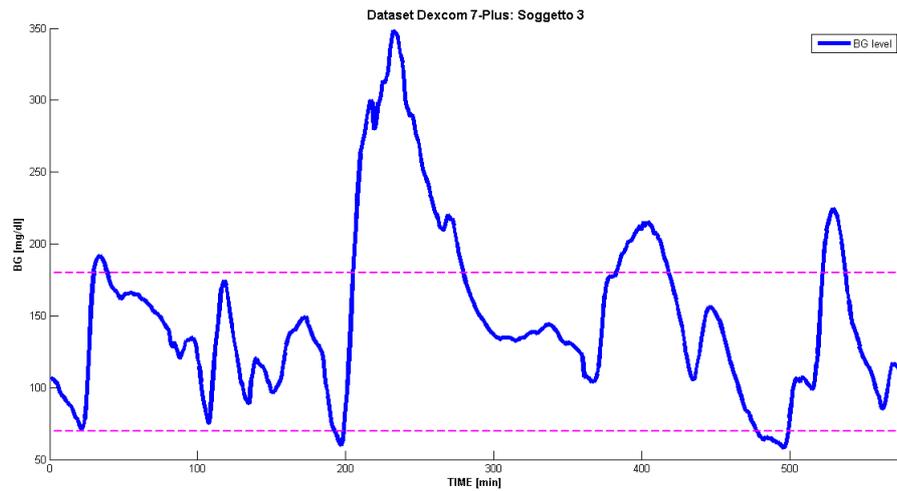


Figura 4.2: Esempio di profilo glicemico misurato con sensore Dexcom® 7-Plus, paziente n° 3; vedi sopra legenda

Sebbene consci del fatto che l'uso delle reti sviluppate per i dati a 5 minuti possa non risultare ottima, abbiamo deciso comunque di tentare il confronto con i risultati già presentati in letteratura, con l'idea di lasciare ad un'eventuale studio futuro il compito d'individuare delle nuove architetture propriamente ottimizzate per le serie temporali ad 1 minuto.

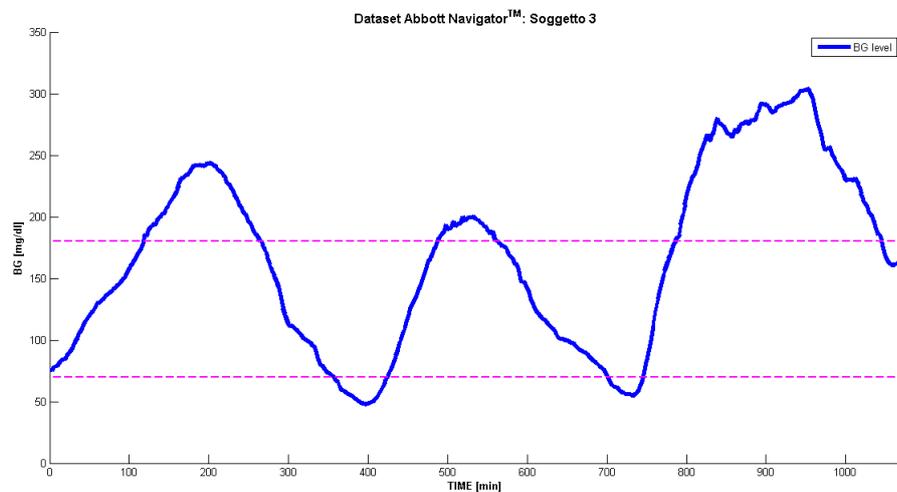


Figura 4.3: Esempio di profilo glicemico misurato con sensore Abbott Navigator™, paziente n° 3; vedi sopra legenda

4.3 Conclusioni

In questo capitolo abbiamo presentato i diversi data base che verranno utilizzati successivamente nel corso di questa tesi per la valutazione delle reti neurali jump, spiegandone brevemente le caratteristiche e i periodi di campionamento utilizzati. Il primo data set, di tipo simulato (Sezione 4.1), verrà ampiamente

utilizzato nel Capitolo 5, mentre i due data set reali, presentati rispettivamente nelle Sezioni 4.2.1 e 4.2.2, verranno utilizzati invece per i test discussi nei Capitoli 6 e 7.

Capitolo 5

Risultati su data set simulato

5.1 Predizione del profilo glicemico

Prima di entrare nel dettaglio dei grafici, è necessario fare una precisazione sulla nomenclatura utilizzata in legenda per indicare i diversi profili.

NNJA #: rappresenta la rete jump, in una delle versioni candidate. Ricordiamo che durante questa fase stiamo ancora trattando le reti nella tipologia A, cioè quelle in cui gli ingressi sono connessi sia allo strato nascosto che allo strato d'uscita (vedi Sezione 3.1);

CGM_target: rappresenta invece il profilo di glicemia target estratto dalla serie temporale ottenuta per mezzo del simulatore [25];

POLY(1): è il profilo ottenuto per mezzo del modello polinomiale di ordine uno.

NN-LPA: è la rete proposta da Zecchin et al. in [24], che troveremo però solo nella successiva valutazione dei parametri.

Per facilitare la lettura dei risultati abbiamo preferito presentare le diverse reti concorrenti in grafici separati. Le Figure 5.1-5.3 mettono in luce il comportamento dei modelli predittivi per due soggetti simulati (Soggetto n° 2 e Soggetto n° 15) di cui abbiamo estratto una porzione esemplificativa della durata di circa 2 giorni (600 campioni) dalla serie temporale originale (7 giorni di monitoraggio). Nella figura troviamo in blu il segnale glicemico utilizzato come target, in verde tratteggiato la predizione ottenuta per mezzo del modello polinomiale di ordine 1 e in rosso la predizione ottenuta con la rete jump.

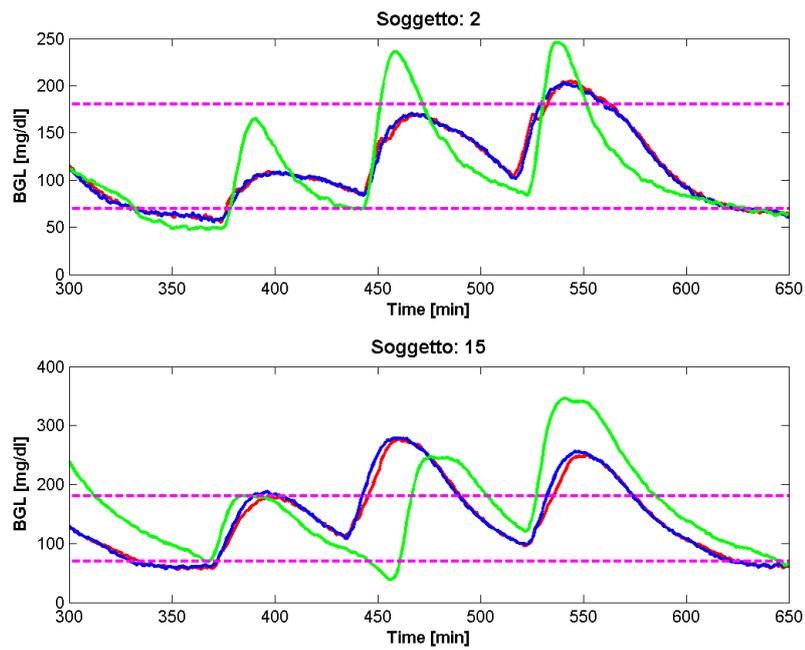


Figura 5.1: Predizione della glicemia ottenuta con NNJA2: (a) Soggetto 2; (b) Soggetto 15; (In blu viene indicato il segnale glicemico target; in verde tratteggiato la predizione ottenuta con il modello poly(1); in rosso la predizione ottenuta con la rete jump; in magenta tratteggiato vengono indicate le soglie di ipo/iperglicemia)

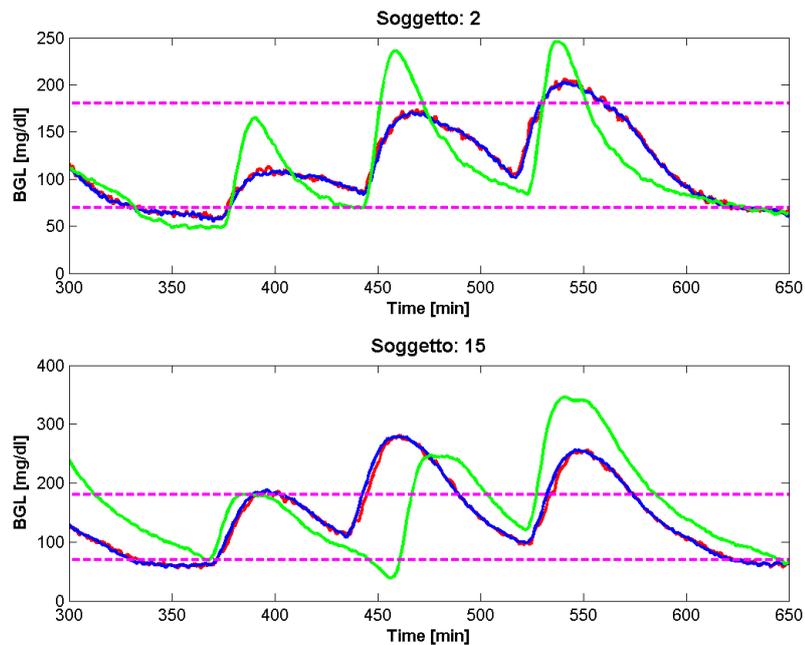


Figura 5.2: Predizione della glicemia ottenuta con NNJA5: (a) Soggetto 2; (b) Soggetto 15; (v. descrizione Figura 5.1)

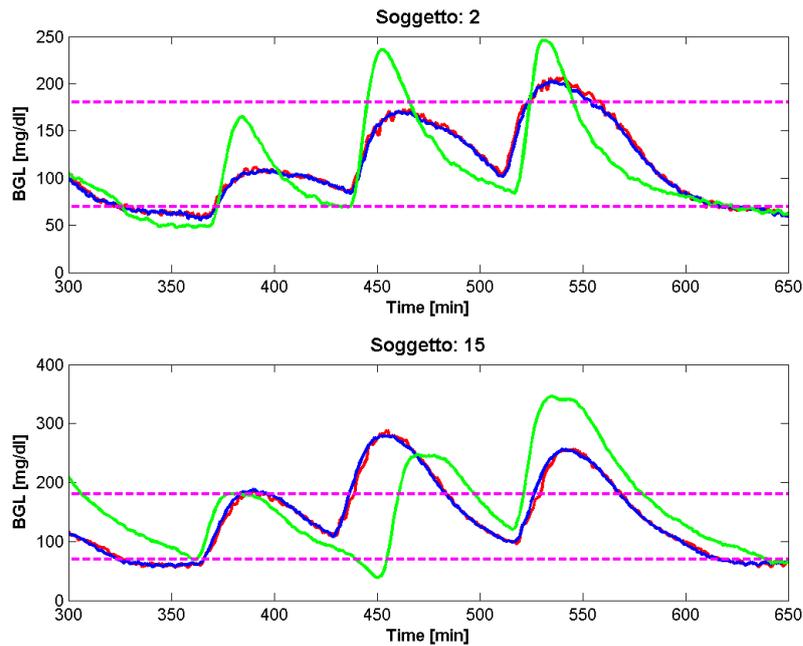


Figura 5.3: Predizione della glicemia ottenuta con NNJA13: (a) Soggetto 2; (b) Soggetto 15; (v. descrizione Figura 5.1)

Da tali grafici appare immediatamente chiaro come il segnale predetto a mezzo delle reti jump risulti molto aderente al target. Il guadagno temporale ottenuto da questi modelli, come sarà confermato più tardi dalla valutazione separata dei diversi indici di qualità, risulta molto elevato, tanto da essere prossimo all'orizzonte di predizione di 30 minuti. La predizione presenta inoltre un andamento confrontabile con quello del target anche se, come verrà sotto evidenziato nell'analisi del parametro $ESOD_{norm}$, all'aumentare della complessità della rete si ha un aumento marginale dell'irregolarità del profilo prodotto. Le reti neurali proposte riescono a cogliere bene anche i diversi trend di salita e discesa del segnale glicemico, individuando con buona precisione sia l'attraversamento delle soglie di iperglicemia che quello delle soglie ipoglicemiche. Parlando del modello polinomiale del primo ordine, si nota subito come esso sia in grado di generare un segnale molto regolare, aspetto di particolare importanza soprattutto per quanto riguarda la generazione degli allarmi in concomitanza all'attraversamento delle soglie critiche. Rispetto alle reti jump il segnale ottenuto per mezzo del modello polinomiale presenta un guadagno temporale più marginale (risultato che verrà confermato anche dal boxplot di Figura 5.4). Sebbene il modello polinomiale sia in grado di seguire abbastanza bene il segnale target durante i fronti di salita e di discesa, in più di un'occasione esso rischia di generare degli allarmi non necessari, sia in ipoglicemia che in iperglicemia. Questo fenomeno si verifica soprattutto quando il segnale target è prossimo alle soglie ipo e iperglicemiche. Vedremo ora più nel dettaglio le caratteristiche delle reti in termini di aderenza al profilo target (RMSE), guadagno temporale ottenuto (TG), e regolarità del profilo predetto ($ESOD_{norm}$ e J), presentati sia numericamente in Tabella 5.2 che graficamente a mezzo di boxplots (Figure 5.4-5.7).

RMSE

In figura 5.4 presentiamo i risultati delle reti neurali (Jump e NN-LPA) e del modello polinomiale in termini di RMSE. Si può subito notare come le reti neurali presentino un RMSE inferiore rispetto a quanto ottenuto con il modello polinomiale. Questo risultato però non sorprende, in quanto ricordiamo

che l'allenamento delle reti avviene proprio minimizzando questo indice. Un'ulteriore considerazione riguarda il fatto che all'aumentare della complessità delle reti jump, in termini di numero d'ingressi e di neuroni nascosti, si riduce il valore del RMSE. Tale andamento è legato al fatto che aumentando la complessità della rete si aumenta di conseguenza la sua capacità di carpire meglio le dinamiche insite nel segnale analizzato. Le reti jump presentate, infine, raggiungono prestazioni confrontabili con quelle della rete NN-LPA. Alcuni test statistici¹ hanno dimostrato che vi sono differenze significative (significatività del 5%) nei risultati ottenuti dalle tre reti jump per l'indice RMSE.

Ricordiamo però che il RMSE non è sufficiente a valutare la bontà della predizione perché non tiene conto di informazioni quali la regolarità del profilo predetto, caratteristica importante per il segnale glicemico, o l'anticipo della predizione ottenuta.

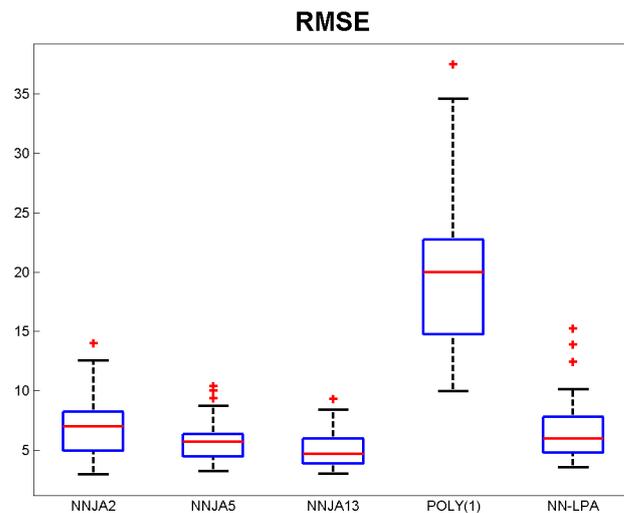


Figura 5.4: Aderenza al profilo target dei modelli in competizione, espressa in termini di RMSE; (tratto rosso: valore mediano; bordi del box: 25mo e 75mo percentile del data set analizzato; tratti neri esterni: estremi del data set analizzato, non considerabili come outliers; croci rosse: outliers)

TG

In figura 5.5 vengono illustrati in modo riassuntivo i guadagni temporali di tutti i modelli testati. Anche qui le reti risultano più performanti del modello polinomiale, con guadagni temporali superiori ai 25 minuti, contro i 7-10 minuti ottenuti con il poly(1). Le reti jump inoltre presentano dei risultati pressoché identici a NN-LPA. È di particolare interesse notare come reti jump molto semplici, come la NNJA2 o la NNJA5, siano in grado di ottenere dei guadagni temporale medi confrontabile con reti più complesse, siano esse jump o feedforward. I test statistici usati (vedi Nota 1) hanno in questo caso dimostrato che non vi sono differenze significative tra le performance dei tre modelli candidati (significatività del 5%). Al fine di un'eventuale implementazione on-chip, dunque, l'uso di uno di questi modelli predittivi più semplici potrebbe garantire un buon compromesso tra qualità dei risultati e complessità del predittore.

¹Come test di significatività sono stati usati il T-Test e il Mann-Whitney Test, dopo aver verificato per mezzo di test Lilliefors se i risultati fossero distribuiti in maniera gaussiana o meno

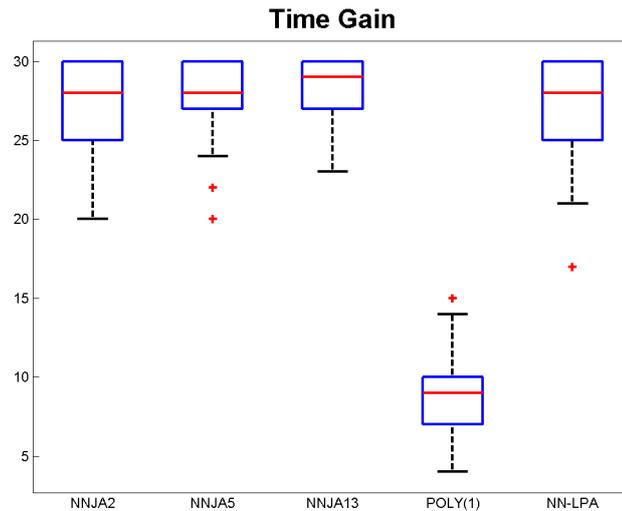


Figura 5.5: Guadagni temporali ottenuti con i modelli in competizione, espressi in termini di indice TG; (descrizione come in Figura 5.4)

ESOD_{norm}

In figura 5.6 mostriamo i risultati riassuntivi in termini di regolarità della predizione, calcolata come rapporto tra l'energia delle derivate seconde del segnale predetto e quella del target. Questa caratteristica è molto importante per la glicemia, soprattutto per la generazione di allarmi all'attraversamento delle soglie critiche. Un segnale che nei pressi di tali soglie presenti delle oscillazioni ad alta frequenza potrebbe infatti dare origine a continui allarmi di fatto non fisiologici. Maggiore sarà la regolarità del segnale, maggiore sarà la probabilità che l'attraversamento delle soglie avvenga in maniera univoca e che gli allarmi vengano prodotti in maniera efficiente.

La predizione ottenuta con le reti neurali presenta una regolarità inferiore a quella del target, mentre il modello polinomiale tende a produrre dei profili sovra-regolarizzati. Tale fatto non stupisce, soprattutto se relazionato con i risultati dell'indice RMSE. Tipicamente, infatti, una maggiore aderenza della predizione al target comporta ulteriormente l'aderenza al rumore sovrapposto al target stesso. Ne consegue quindi che buone prestazioni in termini di RMSE sono solitamente legate ad una ridotta regolarità del profilo. I risultati delle reti neurali rimangono tuttavia molto soddisfacenti, con un profilo che anche visivamente risulta sufficientemente regolare. I test statistici di significatività (vedi Nota 1) hanno infine dimostrato che i risultati prodotti dalle reti jump candidate sono effettivamente diversi.

Questo fatto porta nuovamente la nostra attenzione sulla rete NNJA2 che, in media, riesce a produrre un segnale di energia comparabile a quella del target. Questa caratteristica, abbinata alla ridotta complessità del modello, fa di NNJA2 un modello interessante per il motivo precedentemente espresso di voler trovare una soluzione implementabile on-chip che sia di compromesso tra l'aderenza al target, un buon guadagno temporale e una sufficiente regolarità.

J

L'indice J è stato studiato per includere in una sola variabile le capacità del modello in analisi di anticipare la predizione del segnale target, mantenendo contemporaneamente una buona regolarità del profilo predetto. Ricordiamo qui che quanto più J si avvicina a 0, tanto migliore risulta la predizione ottenuta.

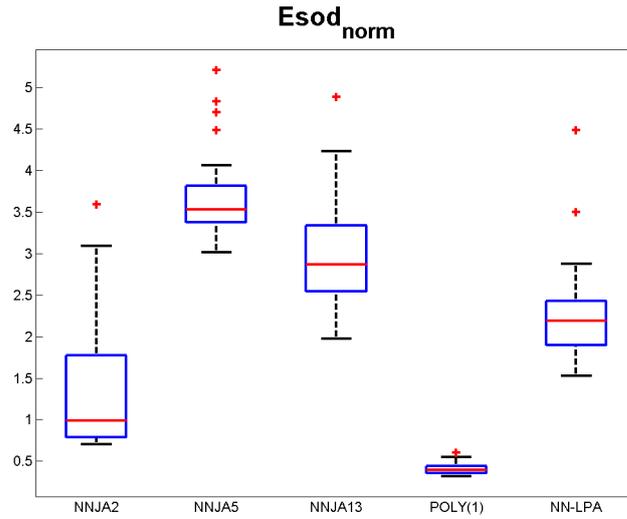


Figura 5.6: Regolarità dei profili predetti dai modelli in competizione, espressa in termini di indice $ESOD_{norm}$; (descrizione come in Figura 5.4)

In figura 5.7 riportiamo i risultati per J . Confrontando questa figura con la Figura 5.6 notiamo che non vi sono grandi differenze. Nello specifico, per il modello polinomiale J risulta prossimo allo 0 proprio in seguito al segnale molto regolare che questo predittore è in grado di generare. Per le reti neurali, sia nel caso jump che nel caso della rete NN-LPA, i valori di J sono leggermente più alti, ma comunque ridotti rispetto all'indice $ESOD_{norm}$, proprio grazie all'anticipo con cui questi modelli riescono a predire il segnale target. Ancora una volta, grazie alla buona regolarità del profilo e all'elevato TG, la rete NNJA2 presenta un'indice J più vicino allo 0 delle altre reti proposte e, grazie nuovamente ai test statistici (vedi Nota 1), abbiamo potuto appurare che i risultati ottenuti dai tre modelli jump A sono significativamente diversi.

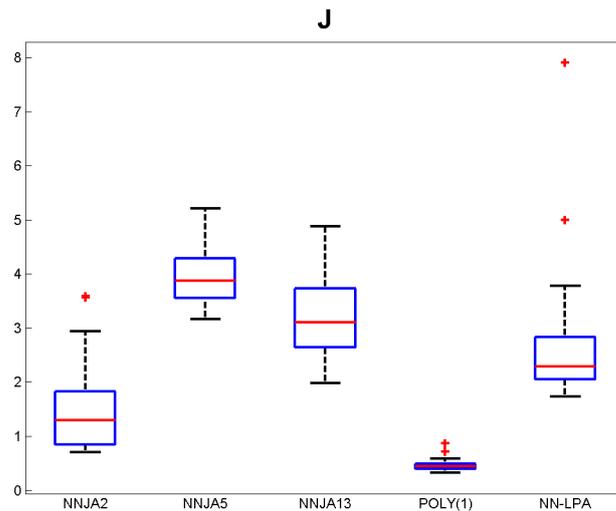


Figura 5.7: Utilità clinica dei profili predetti dai modelli in competizione, espressa in termini di indice J ; (descrizione come in Figura 5.4)

CG-EGA

In Tabella 5.1 riportiamo i valori percentuale relativi all'accuratezza clinica delle predizioni ottenute con le reti jump A qui analizzate, ottenuti per mezzo di test CG-EGA. Dalla tabella possiamo vedere come tutte le reti siano in grado di predire in maniera accurata il valore futuro della glicemia plasmatica. Di particolare interesse sono i valori ottenuti per la zona ipoglicemica, maggiormente pericolosa e di difficile previsione in quanto, eccettuate le informazioni derivanti dallo storico del segnale glicemico, non a disposizione al momento segnali correlati in qualche modo con l'insorgere di tali episodi. Le reti jump, dunque, offrono delle predizioni accurate dal punto di vista clinico in tutto il range glicemico.

| | Ipoglicemia | | Euglicemia | | Iperglicemia | |
|--------|-------------|---------|------------|---------|--------------|---------|
| | Accurate | Erronee | Accurate | Erronee | Accurate | Erronee |
| NNJA2 | 99.91% | 0.09% | 99.6% | 0.4% | 99.3% | 0.7% |
| NNJA5 | 99.7% | 0.3% | 99.57% | 0.43% | 99.24% | 0.76% |
| NNJA13 | 99.89% | 0.11% | 99.62% | 0.38% | 98.78% | 1.22% |

Tabella 5.1: Risultati medi percentuali da analisi CG-EGA per le tre reti jump A proposte, valutati su data set simulato

In Tabella 5.2 vengono infine presentati i risultati medi per tutti i modelli analizzati

L'analisi della tabella conferma quanto precedentemente detto: i profili glicemici predetti con le reti jump A risultano mediamente molto più aderenti al target di quelli ottenuti con modello polinomiale di ordine 1, e confrontabili con quelli ottenuti con la rete NN-LPA. Il guadagno temporale risulta nettamente superiore a quello ottenuto con modello poly(1) e anche la regolarità, espressa sia in termini di energia delle derivate seconde che di indice J, risulta molto buona.

5.2 Reti jump di tipologia B

5.2.1 Strutturazione delle 3 reti candidate

Come precedentemente indicato nel capitolo 3, il passaggio da reti jump A a reti jump B viene fatto applicando la seguente procedura d'analisi:

1. Per ciascuna rete viene effettuata un'analisi qualitativa della matrice dei pesi associati a ciascuna sinapsi neurale;
2. si individuano i pesi che, in media, risultano meno significativi dal punto di vista quantitativo;
3. iterativamente ciascun peso candidato viene posto pari a zero e contemporaneamente viene testata la rete jump così modificata senza che venga riallenata. I pesi successivamente vengono posti a zero in combinazioni diverse o tutti assieme, e la rete viene nuovamente testata senza essere riallenata;
4. per ciascuna rete viene effettuato uno studio visivo dei profili predetti ed una valutazione delle qualità medie della rete in termini di RMSE, Time Gain, ESOD_{norm} e J;
5. le reti che forniscono durante la fase precedente i migliori risultati vengono scelte come reti jump B candidate. Le reti vengono riallenate dopo aver fisicamente eliminato la connessione sinaptica "superflua" e testate nuovamente secondo gli indici qualitativi e un'ispezione visiva dei profili al fine di individuare differenze con la tipologia A.

L'analisi ha portato alla formulazione di queste tre reti jump B

| | NNJA2 | | NNJA5 | | NNJA13 | | POLY(1) | | NN-LPA | |
|-----------|---------|--------|---------|--------|---------|--------|---------|--------|--------|--------|
| | media | sd | media | sd | media | sd | media | sd | media | sd |
| RMSE | 7.2008 | 2.8624 | 5.7572 | 1.8610 | 5.0249 | 1.5000 | 20.0683 | 6.7362 | 6.8858 | 2.5660 |
| Time Gain | 27.2778 | 2.8447 | 26.7778 | 2.7683 | 28.2778 | 2.0648 | 9.0000 | 2.6403 | 27.38 | 3.0897 |
| FSOD norm | 1.0722 | 0.4555 | 3.8041 | 0.4989 | 2.9800 | 0.6447 | 0.4081 | 0.0660 | 2.2432 | 0.4801 |
| J | 1.1900 | 0.4945 | 4.2941 | 0.6233 | 3.1800 | 0.7215 | 0.4561 | 0.1078 | 2.5464 | 0.9617 |

Tabella 5.2: Risultati (media e standard deviation) valutati su data set simulato per i modelli jump, per il predittore polinomiale e per la rete NN-LPA

NNJB2:

In questa rete è stata eliminata la sinapsi che collegava l'ingresso relativo alla serie delle differenze prime della rate of appearance direttamente allo strato d'uscita.

NNJB5

Dalla rete NNJA5 è stata eliminata la sinapsi che portava l'ingresso relativo alla serie delle differenze prime del segnale glicemico direttamente allo strato d'uscita.

NNJB13

In questo caso diversi modelli erano stati presi in considerazione, eliminando di volta i volta connessioni diverse. L'analisi visiva dei profili di predizione ci ha però spinto a scegliere tra le reti possibili quella in cui viene eliminata la sinapsi tra l'ingresso relativo all'errore commesso dal polinomiale e lo strato nascosto. Sebbene tale scelta in linea di principio possa sembrare controintuitiva, l'analisi visiva dei profili e gli indici di qualità hanno confermato che questa rete si comporta meglio delle alternative sottoposte ad analisi.

5.2.2 Risultati

Facendo riferimento alle Figure (5.8-5.10) possiamo notare facilmente che le reti jump B forniscono dei risultati paragonabili a quelli delle reti jump A mostrati in Sezione 5.1.

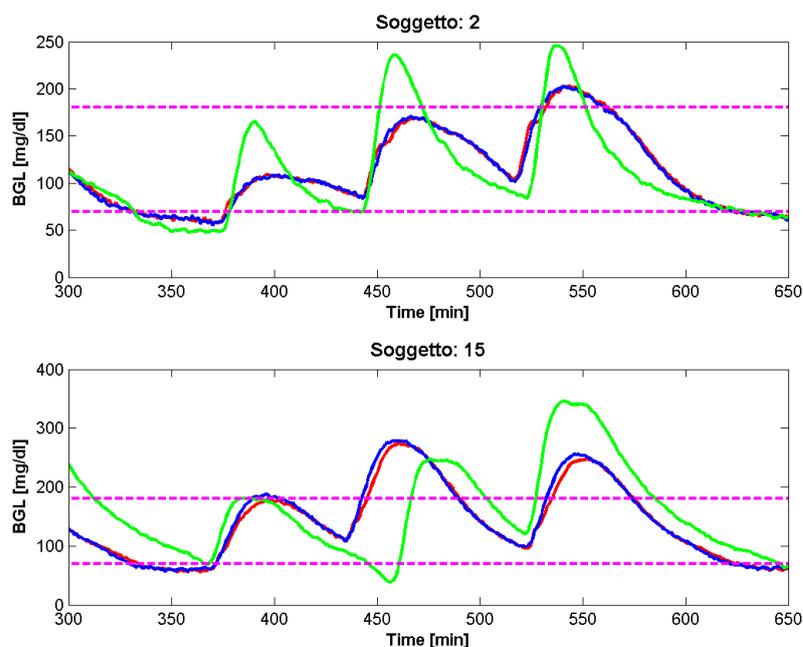


Figura 5.8: Predizione della glicemia ottenuta con NNJB2: (a) Soggetto 2; (b) Soggetto 15; (v. descrizione Figura 5.1, stessi dati utilizzati)

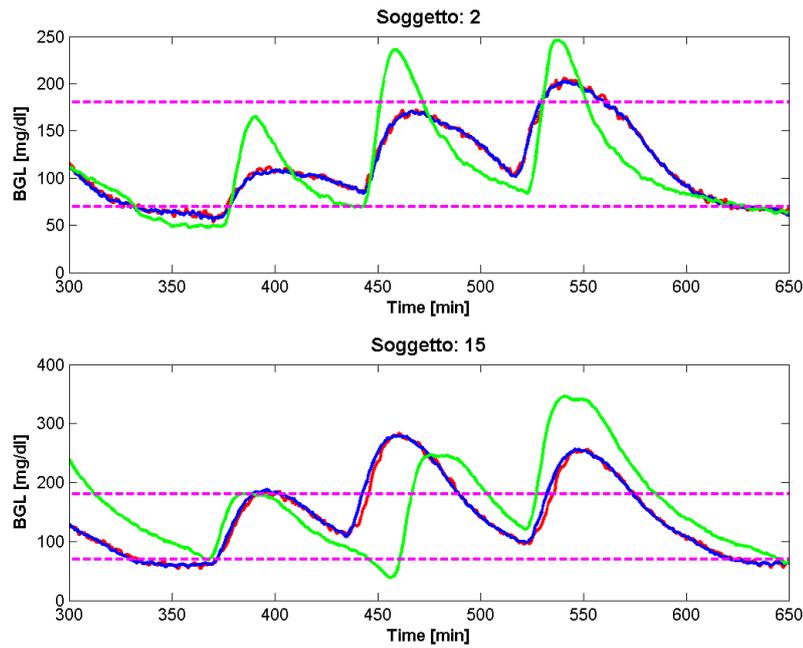


Figura 5.9: Predizione della glicemia ottenuta con NNJB5: (a) Soggetto 2; (b) Soggetto 15; (v. descrizione Figura 5.1, stessi dati utilizzati)

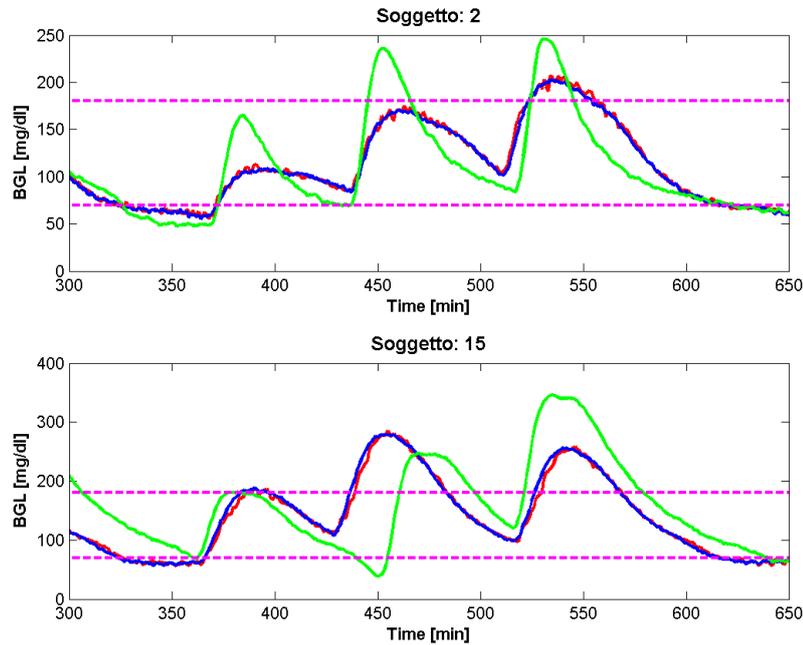


Figura 5.10: Predizione della glicemia ottenuta con NNJB13: (a) Soggetto 2; (b) Soggetto 15; (v. descrizione Figura 5.1, stessi dati utilizzati)

I profili molto spesso coincidono e, da un'analisi quantitativa degli indici (Tabella 5.3) possiamo vedere che non vi sono variazioni significative.

| | NNJB2 | | NNJB5 | | NNJB13 | |
|-----------|---------|--------|---------|--------|--------|--------|
| | media | sd | media | sd | media | sd |
| RMSE | 7.1845 | 2.7441 | 5.6339 | 1.8498 | 5.1039 | 1.5289 |
| Time Gain | 27.4167 | 2.9605 | 27.3056 | 2.7549 | 28.25 | 2.0476 |
| ESOD norm | 1.1650 | 0.791 | 3.4387 | 0.4592 | 3.1703 | 0.7296 |
| J | 1.2757 | 0.8119 | 3.8068 | 0.5798 | 3.3814 | 0.7898 |

Table 5.3: Risultati (media e standard deviation) ottenuti con le reti jump B per gl'indici qualitativi proposti

La Figura 5.11 illustra invece tali risultati riassumendoli graficamente in termini di Boxplots. Valgono quindi le stesse considerazioni fatte per le reti della tipologia A. Per avvalorare questa ipotesi sono stati nuovamente effettuati dei test statistici, prima solo sui risultati delle reti jump B e successivamente confrontando le reti jump A con le rispettive reti jump B.

I primi test hanno messo in luce che per quanto riguarda il guadagno temporale le tre reti B riescono ad ottenere dei risultati statisticamente non diversi, mentre i restanti tre indici i risultati statistici mostrano le reti come diverse. Nuovamente, dunque, la rete NNJB2 sembra essere un buon compromesso tra complessità architetturale e prestazioni generali, con un buon guadagno temporale, un profilo predetto sufficientemente aderente al target e una buona regolarità del segnale ottenuto.

I test che hanno messo a confronto le reti jump A con le rispettive jump B hanno dimostrato infine che il passaggio alla tipologia B non comporta differenze significative nei risultati. Volendo quindi ridurre ulteriormente la complessità dei modelli proposti, non appare azzardato il passaggio dalla tipologia A alla tipologia B. Tale passaggio permetterebbe infatti di ridurre il numero di parametri che è necessario calcolare in fase di training delle reti.

Per quanto compete la valutazione “clinica” delle performance delle reti nella loro versione B, la tabella 5.4 riporta la percentuale di misure accurate ed erronee ottenute dal test CG-EGA. Risulta chiaro come le reti jump B, come già visto per le reti jump A, riescano a fornire delle misure clinicamente valide, con percentuali d'accuratezza sempre superiori al 99% in tutto il range glicemico.

| | Ipoglicemia | | Euglicemia | | Iperglicemia | |
|--------|-------------|---------|------------|---------|--------------|---------|
| | Accurate | Erronee | Accurate | Erronee | Accurate | Erronee |
| NNJB2 | 99.96% | 0.04% | 99.73% | 0.27% | 99.3% | 0.7% |
| NNJB5 | 99.8% | 0.2% | 99.63% | 0.37% | 99.21% | 0.79% |
| NNJB13 | 99.9% | 0.1% | 99.6% | 0.4% | 98.77% | 1.23% |

Table 5.4: Risultati medi percentuali da analisi CG-EGA per le tre reti jump B proposte, valutati su data set simulato

5.3 Considerazioni riassuntive

In questo capitolo abbiamo testato le reti jump con un data-set simulato per valutarne le prestazioni sia in termini di indici di qualità (RMSE, Time Gain, ESOD_{norm} e J) che in termini d'ispezione visiva dei profili predetti. I risultati sono stati confrontati con i modelli stato dell'arte poly(1) [21] e NN-LPA [24].

I risultati ottenuti nella Sezione 5.1 con i modelli jump A sono paragonabili a quelli ottenuti con NN-LPA e superiori a quelli del modello polinomiale di primo ordine. Le reti jump sembrano quindi capaci di apprendere e seguire l'andamento della glicemia futura fornendo delle predizioni che, oltre che

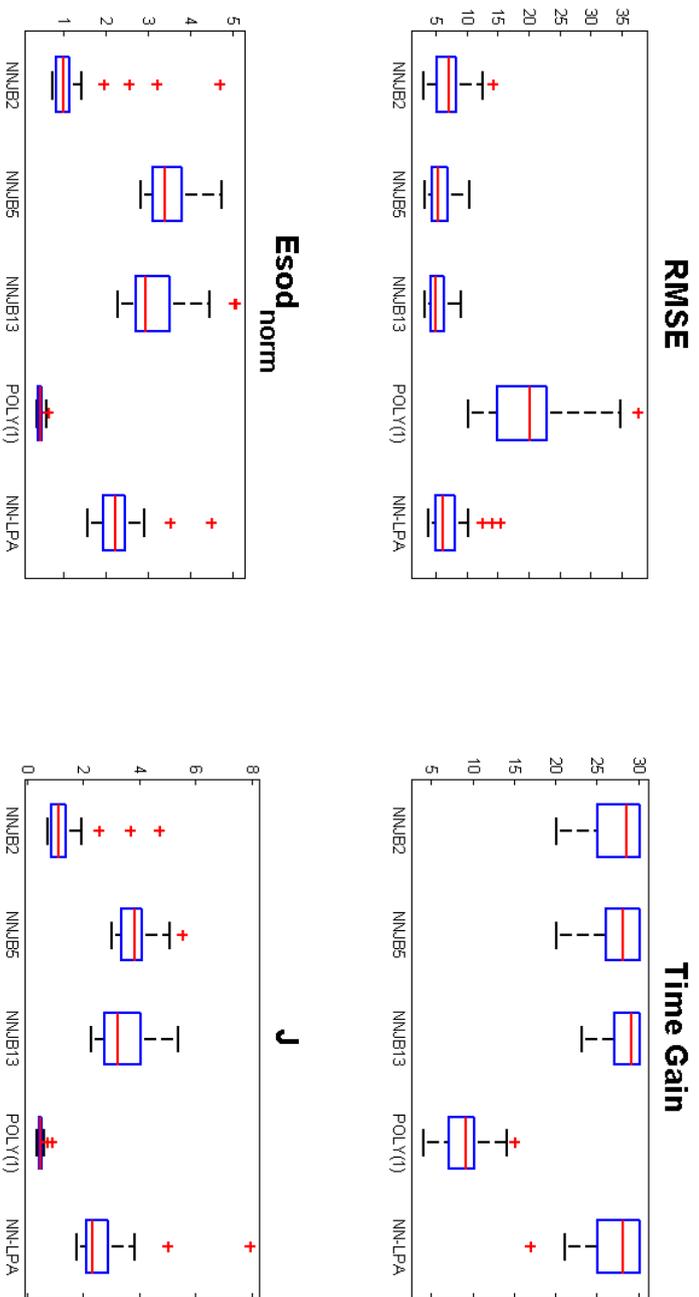


Figura 5.11: Boxplot dei risultati per i quattro indici di qualità calcolati per tutti i modelli in competizione; (a) RMSE; (b) Time Gain; (c) $ESOD_{norm}$; (d) J ; v. descrizione Figura 5.4

buone in termini di indici di qualità, risultano anche clinicamente accurate, come dimostrano i risultati ottenuti per mezzo del test CG-EGA. Le reti riescono a cogliere la stragrande maggioranza degli episodi ipoglicemici e iperglicemici, garantendo una buona aderenza al profilo target e mantenendo una regolarità molto buona, limitando quindi la presenza di oscillazioni spurie ad alta frequenza.

Nella Sezione 5.2, attraverso un'analisi dei pesi sinaptici, abbiamo successivamente individuato dei candidati per la tipologia B delle reti jump. Per ciascuna rete jump A in particolare è stata individuata un'alternativa B, che è stata riallenata e testata sui medesimi training-set e test-set utilizzati per le reti jump A. Tali reti si sono dimostrate tanto prestanti quanto quelle della tipologia A, con dei risultati che, ancora una volta, sono molto buoni, sia in termini visivi che in termini di indici di qualità. Test statistici hanno inoltre dimostrato che il passaggio dalla tipologia A alla tipologia B non ha portato a differenze significative nei risultati, garantendo quindi che, con il passaggio da A a B, non si verifica una degradazione dei risultati ottenuti.

Come nota finale vogliamo dire che di particolare interesse è risultata la rete NNJA2/NNJB2. Sebbene queste reti siano molto semplici in termini di numero di ingressi e complessità dell'architettura neurale, esse riescono a predire il segnale target con un anticipo statisticamente equivalente a quello di reti più complesse, come la NNJA5 o la NNJA13, mantenendo un profilo sufficientemente aderente al target, sia in termini di regolarità che in termini di residui. Dovendo quindi fare una scelta di compromesso tra qualità dei risultati e complessità del modello utilizzato, le reti NNJA2/NNJB2 sembrano essere le più adatte a tale scopo.

Nel prossimo capitolo andremo a testare le reti jump anche su dataset reale e verificheremo se le reti NNJA2/NNJB2 siano ancora in grado di produrre i buoni risultati ottenuti in sede di simulazione.

Capitolo 6

Risultati su data set reale Dexcom® 7-Plus

6.1 Predizione del profilo glicemico

È stata qui utilizzata la medesima nomenclatura già presentata per il Capitolo 5; abbiamo nuovamente scelto di prediligere una maggiore chiarezza nella presentazione dei profili di predizione, distinguendo i grafici ottenuti per ciascuna delle reti jump candidate.

I profili predetti dalle reti jump A, Figure (6.1 - 6.3) (Soggetti n°3 e n°7 data set reale Dexcom® 7-Plus) presentano una variabilità molto più alta di quanto riscontrato durante la fase di simulazione. Le reti riescono a cogliere abbastanza bene le dinamiche glicemiche dei soggetti analizzati, anche se in alcuni casi (vedi soggetto 7) le reti jump producono delle oscillazioni ad alta frequenza attorno al segnale target. Questi risultati non devono tuttavia sorprendere, per due motivi principalmente: in prima battuta è necessario ricordare che durante la simulazioni le medesime reti erano state allenate con un dataset molto più ampio di quello a disposizione per il caso reale; in secondo luogo i dati reali, per loro natura, sono comunque molto più difficili da predire, a causa delle dinamiche intrinsecamente meno regolari di quelle ottenute in simulazione. Sebbene però il numero di pazienti e di campioni a disposizione siano molto inferiori, e il segnale da predire sia molto più complesso, le reti neurali riescono comunque a fornire informazioni utili sull'evoluzione della glicemia plasmatica nel prossimo futuro; . Ancora una volta quindi non sorprende che le reti generino delle oscillazioni spurie, essendo di fatto anche il segnale target molto più irregolare.

Per quanto concerne l'attraversamento delle soglie critiche di ipo e iperglicemia, le reti jump A si comportano abbastanza bene, anche se nuovamente la presenza di oscillazioni spurie porta alla generazione di allarmi per eventi glicemici critici effettivamente non presenti. Una nota particolare riguarda la rete NNJA2: essa, che durante la precedente fase di test su database simulato era risultata la migliore in termini di compromesso tra complessità architetturale e bontà dei risultati, in questo caso, probabilmente a causa della ridotta complessità dell'architettura e dell'esiguità del dataset a disposizione, presenta un ritardo non trascurabile nella predizione. Tuttavia, per quanto concerne la regolarità del profilo predetto, tale rete si mostra ancora attraente, con una quasi totale assenza di oscillazioni ad alta frequenza nel segnale predetto.

Andiamo ora a valutare nello specifico i risultati ottenuti dalle reti per ogni singolo indice di qualità, con i risultati medi raccolti in tabella 6.2.

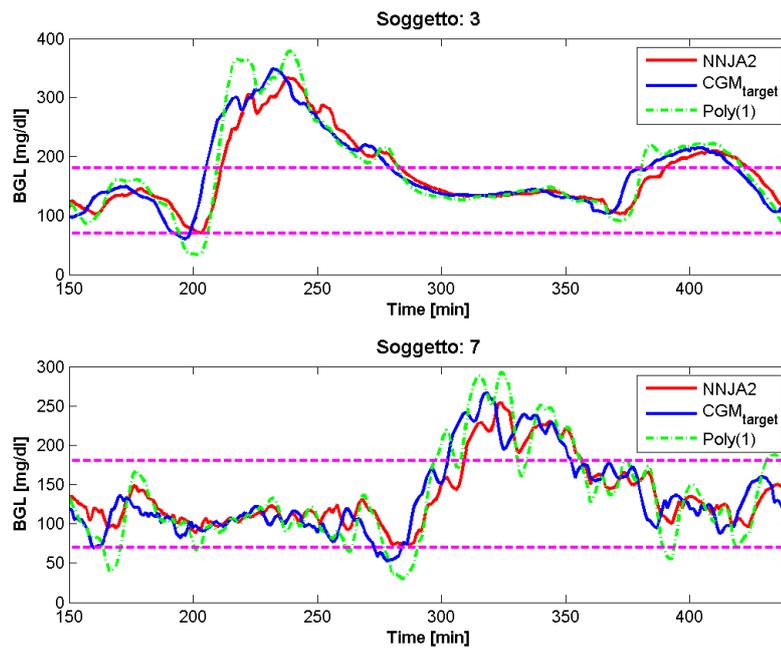


Figura 6.1: Predizione della glicemia ottenuta con NNJA2: (a) Soggetto 3; (b) Soggetto 7; (In blu viene indicato il segnale glicemico target; in verde tratteggiato la predizione ottenuta con il modello poly(1); in rosso la predizione ottenuta con la rete jump; in magenta tratteggiato vengono indicate le soglie di ipo/iperglicemia)

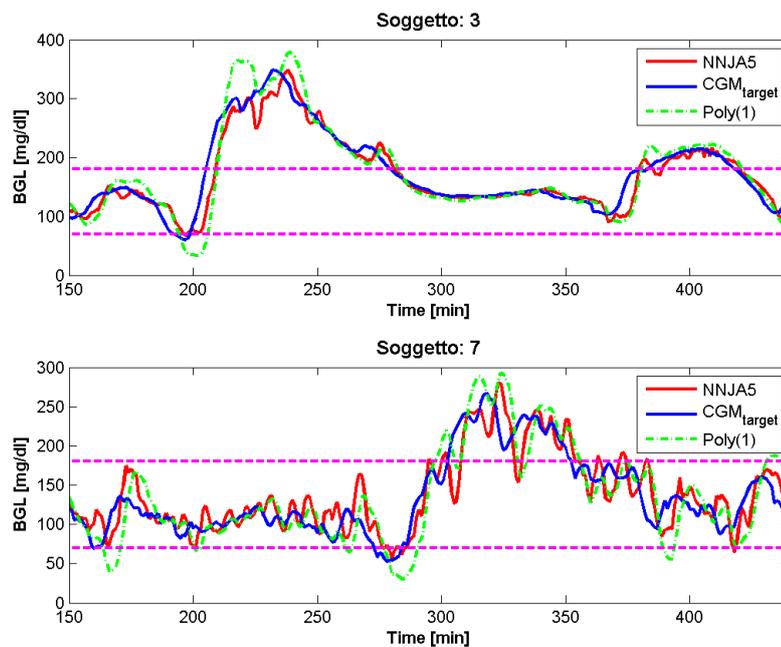


Figura 6.2: Predizione della glicemia ottenuta con NNJA5: (a) Soggetto 3; (b) Soggetto 7; (v. descrizione Figura 6.1)

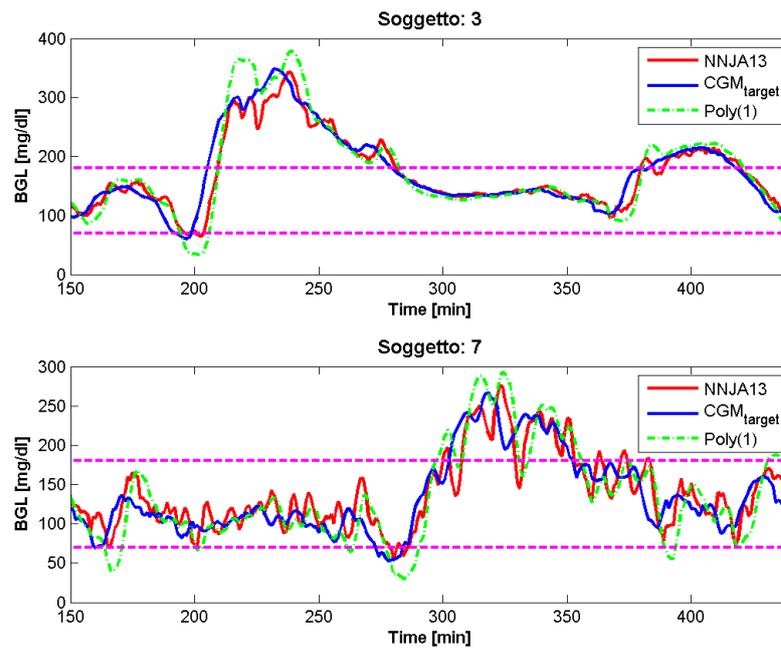


Figura 6.3: Predizione della glicemia ottenuta con NNJA13: (a) Soggetto 3; (b) Soggetto 7; (v. descrizione Figura 6.1)

RMSE

In Figura 6.4 (a) presentiamo i risultati riassuntivi in termini di RMSE ottenuti dalle reti neurali e dal modello polinomiale del primo ordine.

E' possibile vedere come, ancora una volta, le reti neurali dimostrino in media un RMSE inferiore a quello del modello polinomiale. Nuovamente questo risultato è riconducibile alle modalità con cui le reti vengono allenate, cioè minimizzando proprio il RMSE. Ricordiamo però che questo indice è parzialmente indicativo della bontà del segnale prodotto in quanto in grado di fornire solo un'informazione relativa all'aderenza al profilo target, trascurando altri fattori importanti nel definire la qualità complessiva del predittore. Una nota particolare va fatta alla rete NNJA2: sebbene, ribadiamo, essa presenti un'architettura relativamente semplice e l'analisi visiva dei profili abbia messo in luce alcuni limiti di tale modello, in termini quantitativi il RMSE ottenuto con la NNJA2 risulta ancora inferiore a quello misurato per il modello polinomiale, ma significativamente superiore di quello delle reti NNJA5 e NNJA13. In questo caso dunque, al fine di realizzare un compromesso tra qualità della predizione e complessità della rete neurale, la rete NNJA5 è in grado di produrre dei buoni risultati, con un'architettura più semplice di quella della rete NNJA13.

Time Gain

Per quanto concerne il guadagno temporale, nel boxplot di Figura 6.4 (b) possiamo osservare come si distribuiscono i risultati per i modelli candidati. In prima istanza risulta chiaro, come precedentemente accennato durante l'analisi visiva dei profili, come la rete NNJA2 non sia in grado di fornire un guadagno temporale sufficiente. La predizione, infatti, risulta sempre troppo ritardata, a causa probabilmente sia della semplicità della rete che dell'esiguo numero di dati presenti per il dataset reale.

Le ulteriori reti proposte mostrano invece un guadagno temporale confrontabile con quello della rete stato dell'arte NN-LPA, con un anticipo nella predizione che si assesta intorno agli 8 minuti. Il modello polinomiale, infine, presenta un TG leggermente inferiore a quello degli altri modelli ma ancora buono. I test di significatività mostrano, com'era intuibile, che le prestazioni ottenute con la rete NNJA2 sono statisticamente diverse da quelle ottenute con le altre due reti candidate, mentre queste ultime risultano tra di loro non significativamente diversi. Se, anche qui, volessimo operare una scelta di modello basata sul compromesso tra prestazioni ottenute e complessità dell'architettura, la rete NNJA5 sembra rispettare tali requisiti per l'uso con soggetti reali.

ESOD_{norm}

I risultati ottenuti per la regolarità dei profili predetti sono presentati in Figura 6.4 (c). La rete NNJA2, che abbiamo precedentemente detto mostrare un'aderenza al profilo target minore di quella delle altre reti jump, raggiunge qui un punteggio rappresentativo di una buona regolarità del segnale predetto. Le ulteriori reti jump, invece, presentano un profilo la cui energia associata alle derivate seconde è circa 4 volte quella del segnale target. Sebbene generino dei profili meno regolari rispetto agli altri modelli, le reti NNJA5 e NNJA13 producono dei risultati comunque accettabili in termini di ESOD_{norm}. Test statistici hanno dimostrato la non significativa diversità di queste reti per quanto compete questo indice, mentre i risultati ottenuti per mezzo della rete NNJA2 appaiono statisticamente diversi. Anche in questo caso la rete NNJA5 risponde alla richiesta di compromesso tra complessità e qualità dei risultati.

J

In Figura 6.4 (d) vediamo, infine, il boxplot dei valori ottenuti dai modelli candidati per l'indice J. Visti i risultati ottenuti dalla rete NNJA2 per l'indice TG, non sorprende affatto che l'indice J assuma valori fuori scala. Un TG prossimo a zero, infatti, porta l'indice J ad assumere valori molto elevati che, come vediamo anche dalla tabella 6.2, non avrebbe nemmeno senso misurare. Le reti NNJA5 e NNJA13 presentano invece dei valori di J maggiori di quelli ottenuti con il modello polinomiale o la rete NN-LPA. Ciò può essere ricondotto al fatto che, sebbene il guadagno temporale ottenuto dalle reti jump sia in linea con quello ottenuto con la rete NN-LPA, un indice ESOD_{norm} maggiore influenza la bontà complessiva della predizione facendo crescere J, che ricordiamo essere proporzionalmente collegato a quest'ultimo. È stato verificato che i risultati ottenuti con la NNJA5 non sono statisticamente diversi da quelli della rete NNJA13, mentre quelli della NNJA2 lo sono se confrontati con quelli di entrambe le alternative jump. Col fine ancora una volta d'individuare un buon compromesso tra qualità dei risultati e complessità dell'architettura neurale, la rete NNJA5 sembra nuovamente essere un buon candidato per tale scopo.

CG-EGA

In tabella 6.1 sono riportati i valori ottenuti dall'analisi CG-EGA. A differenza di quanto ottenuto in simulazione (Sezione 5.1), qui le reti jump presentano delle difficoltà nel predire correttamente i fenomeni ipoglicemici, con la sola rete NNJA13 in grado di individuarne correttamente più del 50%. Tale deficit è imputabile al numero ridotto di dati a disposizione per il training: non disponendo di un ampio dataset d'allenamento, le reti non riescono a riprodurre i risultati ottenuti durante la precedente fase di simulazione. Fisiologicamente, inoltre, il numero di eventi ipoglicemici presenti nelle serie temporali plasmatiche è minore di quello degli episodi eu- e iperglicemici. È facilmente intuibile quindi come le reti

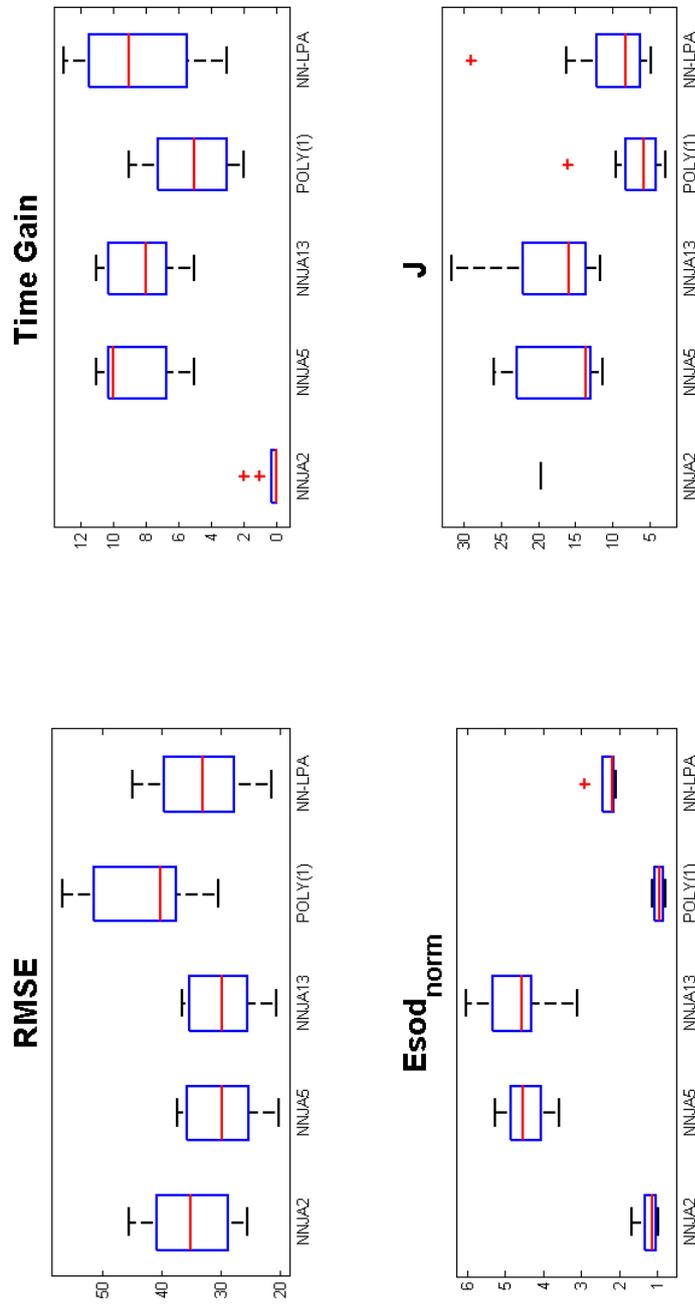


Figura 6.4: Boxplot dei risultati per i quattro indici di qualità calcolati per tutti i modelli in competizione; (a) RMSE; (b) Time Gain; (c) $ESOD_{norm}$; (d) J ; ((tratto rosso: valore mediano; bordi del box: 25mo e 75mo percentile del data set analizzato; tratti neri esterni: estremi del data set analizzato, non considerabili come outliers; croci rosse: outliers))

abbiano a disposizione un minor numero di informazioni utili alla predizione accurata dell'insorgenza di un evento ipoglicemico futuro.

Per quanto riguarda invece i risultati per le fasce di euglicemia e di iperglicemia, tutte e tre le reti svolgono un buon lavoro di predizione, con risultati in media superiori all'88% di predizioni accurate. Tale risultato è imputabile sia al maggior numero di campioni disponibili nelle rispettive fasce d'interesse sia all'assenza, tra gli ingressi della rete, d un segnale direttamente correlato agli episodi ipoglicemici.

| | Ipoglicemia | | Euglicemia | | Iperglicemia | |
|--------|-------------|---------|------------|---------|--------------|---------|
| | Accurate | Erronee | Accurate | Erronee | Accurate | Erronee |
| NNJA2 | 31.7% | 68.3% | 92.1% | 7.9% | 85.2% | 14.8% |
| NNJA5 | 40.2% | 59.8% | 89.7% | 10.3% | 84.6% | 15.4% |
| NNJA13 | 56.3% | 43.7% | 88.9% | 11.1% | 85.3% | 14.7% |

Table 6.1: Risultati medi percentuali da analisi CG-EGA per le tre reti jump A proposte, valutati su data set reale Dexcom® 7-Plus

In Tabella 6.2 vengono infine presentati i risultati medi per tutti i modelli analizzati

L'analisi della tabella conferma quanto precedentemente detto: ad eccezione della rete NNJA2, i profili glicemici predetti con le reti jump NNJA5 e NNJA13 risultano mediamente più aderenti al target di quelli ottenuti con modello polinomiale di ordine 1, e confrontabili con quelli ottenuti con la rete NN-LPA. Il guadagno temporale risulta superiore a quello ottenuto con modello poly(1) e anche la regolarità, espressa sia in termini di energia delle derivate seconde che di indice J, risulta ancora buona.

6.2 Reti jump di tipologia B

6.2.1 Strutturazione delle 2 reti candidate

Come precedentemente indicato nei capitoli 3 e 5, il passaggio da reti jump A a reti jump B viene fatto applicando la seguente procedura d'analisi:

1. Per ciascuna rete viene effettuata un'analisi qualitativa della matrice dei pesi associati a ciascuna sinapsi neurale;
2. si individuano i pesi che, in media, risultano meno significativi dal punto di vista quantitativo;
3. iterativamente ciascun peso candidato viene posto pari a zero e contemporaneamente viene testata la rete jump così modificata senza che venga riallenata. I pesi successivamente vengono posti a zero in combinazioni diverse o tutti assieme, e la rete viene nuovamente testata senza essere riallenata;
4. per ciascuna rete viene effettuato uno studio visivo dei profili predetti ed una valutazione delle qualità medie della rete in termini di RMSE, Time Gain, ESOD_{norm} e J;
5. le reti che forniscono durante la fase precedente i migliori risultati vengono scelte come reti jump B candidate. Le reti vengono riallenate dopo aver fisicamente eliminato la connessione sinaptica "superflua" e testate nuovamente secondo gli indici qualitativi e un'ispezione visiva dei profili al fine di individuare differenze con la tipologia A.

Nel caso di dataset reale abbiamo riscontrato delle difficoltà a formulare un'alternativa B per la rete NNJA2. Infatti il passaggio ad un possibile candidato B della rete non ha migliorato i risultati ottenuti nella

| | NNJA2 | | NNJA5 | | NNJA13 | | POLY(1) | | NNdeLin | |
|-----------|----------|------------|---------|--------|---------|--------|---------|--------|---------|--------|
| | media | sd | media | sd | media | sd | media | sd | media | sd |
| RMSE | 34.8527 | 6.9131 | 29.6472 | 5.9851 | 29.4257 | 5.8523 | 43.3148 | 9.3669 | 33.1206 | 7.5570 |
| Time Gain | 0.3333 | 0.7071 | 8.6667 | 2.2361 | 8.4444 | 2.1858 | 5.3333 | 2.3979 | 8.3333 | 3.6401 |
| ESOD norm | 1.1933 | 0.2177 | 4.4706 | 0.5578 | 4.7372 | 0.8604 | 0.9570 | 0.1274 | 2.2940 | 0.2664 |
| J | ∞ | <i>NaN</i> | 16.6955 | 5.7724 | 17.9671 | 6.3725 | 6.7438 | 4.0137 | 10.7742 | 7.6980 |

Table 6.2: Risultati (media e standard deviation) valutati su data set Dexcom® 7-Plus per i modelli jump, per il predittore polinomiale e per la rete NN-LPA

fase precedente, anzi ha portato a risultati qualitativamente inferiori. Ciò sembra essere effettivamente legato al ridotto numero di dati a disposizione che, combinati con la semplicità della rete, non hanno fornito risultati utili.

Le reti jump B candidate per la successiva fase d'analisi sono quindi 2: NNJB5 e NNJB13.

NNJB5

Di tutte le alternative possibili come alternativa B della rete NNJA5, quella risultata mediamente migliore è stata ottenuta eliminando la connessione diretta al livello d'uscita del ingresso relativo alla serie delle differenze prime del segnale glicemico.

NNJB13

Diverse reti sono risultate interessanti come candidate B della rete NNJA13. L'alternativa da noi proposta è una rete in cui è stata eliminata la connessione diretta al livello d'uscita del segnale associato alla serie delle differenze prime del segnale glicemico.

6.2.2 Risultati

Le Figure 6.5 e 6.6 illustrano i profili predetti dalle due reti jump B candidate. Un'ispezione visiva mette in luce che non vi sono differenze sostanziali tra le versioni jump B qui presentate e i risultati precedentemente raccolti per le reti jump A (figure 6.2-6.3). Valgono di conseguenza le medesime considerazioni.

Per quanto riguarda invece gli indici qualitativi RMSE, Time Gain, $ESOD_{norm}$ e J, la Figura 6.7 riassume graficamente la distribuzioni di tali indici per i quattro modelli a confronto. Possiamo notare un leggero incremento nel guadagno temporale ottenuto con le reti NNJB5 e NNJB13, fatto confermato anche dai risultati numerici presentati in Tabella 6.3, mentre non variano di molto i risultati per gli altri tre indici.

| | NNJB5 | | NNJB13 | |
|-----------|---------|--------|---------|--------|
| | media | sd | media | sd |
| RMSE | 29.5248 | 5.7277 | 28.9064 | 6.0816 |
| Time Gain | 9.5556 | 2.6034 | 9.7778 | 2.7285 |
| ESOD norm | 4.1739 | 0.9021 | 5.2030 | 1.0749 |
| J | 14.2775 | 6.0928 | 17.2496 | 6.5501 |

Table 6.3: Risultati medi valutati su dataset reale per le reti di tipo jump B.

I test statistici (T-Test e Mann-Whitney Test) sono stati effettuati per valutare se le due reti presentassero differenze significative. Da questi test è emerso che le reti non sono tra loro statisticamente diverse per quanto concerne il guadagno temporale, mentre lo sono per gli altri indici d'interesse. Dato quindi che la rete NNJB5 è in grado di generare un profilo di predizione leggermente più regolare di quello ottenuto a mezzo della rete NNJB13, come dimostrano i valori in Tabella 6.3 (vedi $ESOD_{norm}$ e J), essa risponde meglio alla richiesta di compromesso tra qualità dei risultati e complessità dell'architettura.

Ulteriori test sono stati infine effettuati per verificare la presenza di differenze significative tra i modelli jump A e B. Mentre per le reti NNJA5 e NNJB5 è stato verificato non esservi differenze statisticamente significative, per le reti NNJA13 e NNJB13 tali differenze sono state verificate per gli indici RMSE, Time Gain e $ESOD_{norm}$, con la versione B che presenta dei risultati leggermente migliori rispetto alla versione

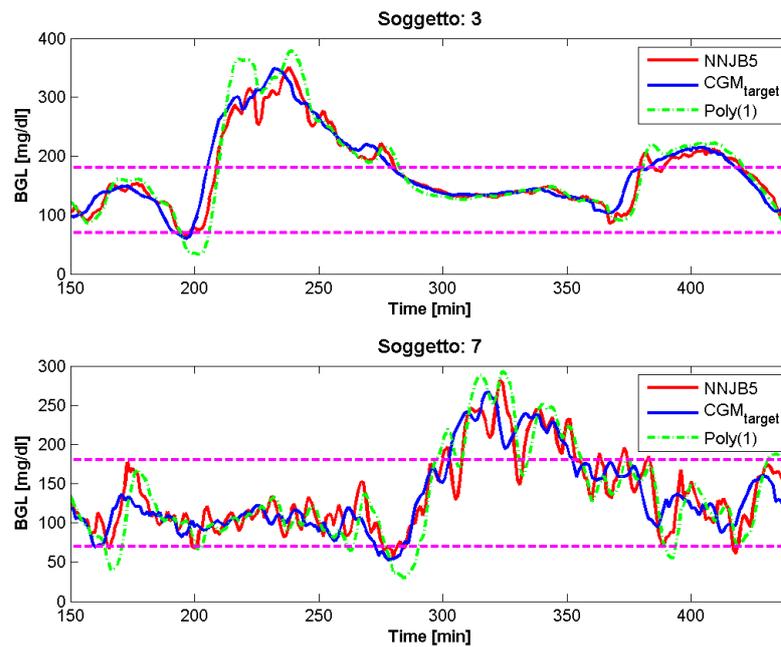


Figura 6.5: Predizione della glicemia ottenuta con NNJB5: (a) Soggetto 3; (b) Soggetto 7; (v. descrizione Figura 6.1, stessi dati utilizzati)

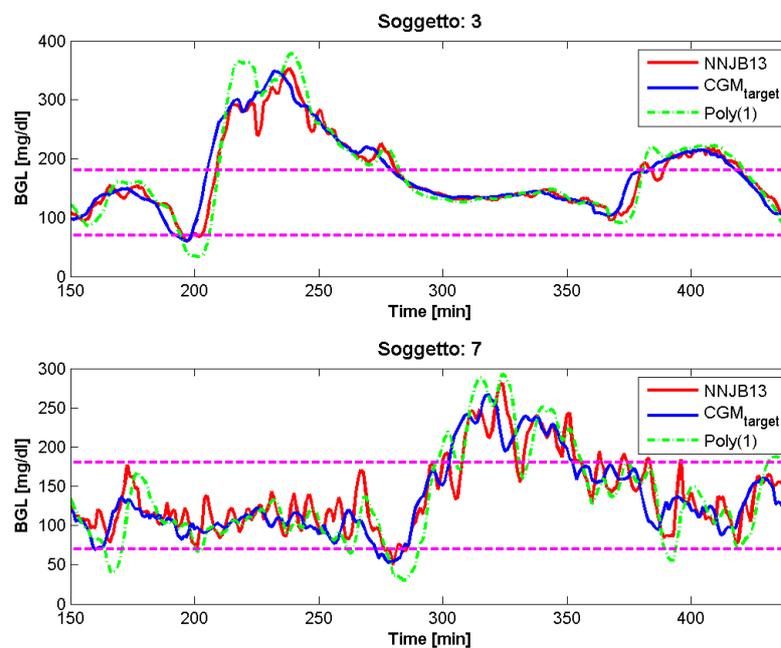


Figura 6.6: Predizione della glicemia ottenuta con NNJB13: (a) Soggetto 3; (b) Soggetto 7; (v. descrizione Figura 6.1, stessi dati utilizzati)

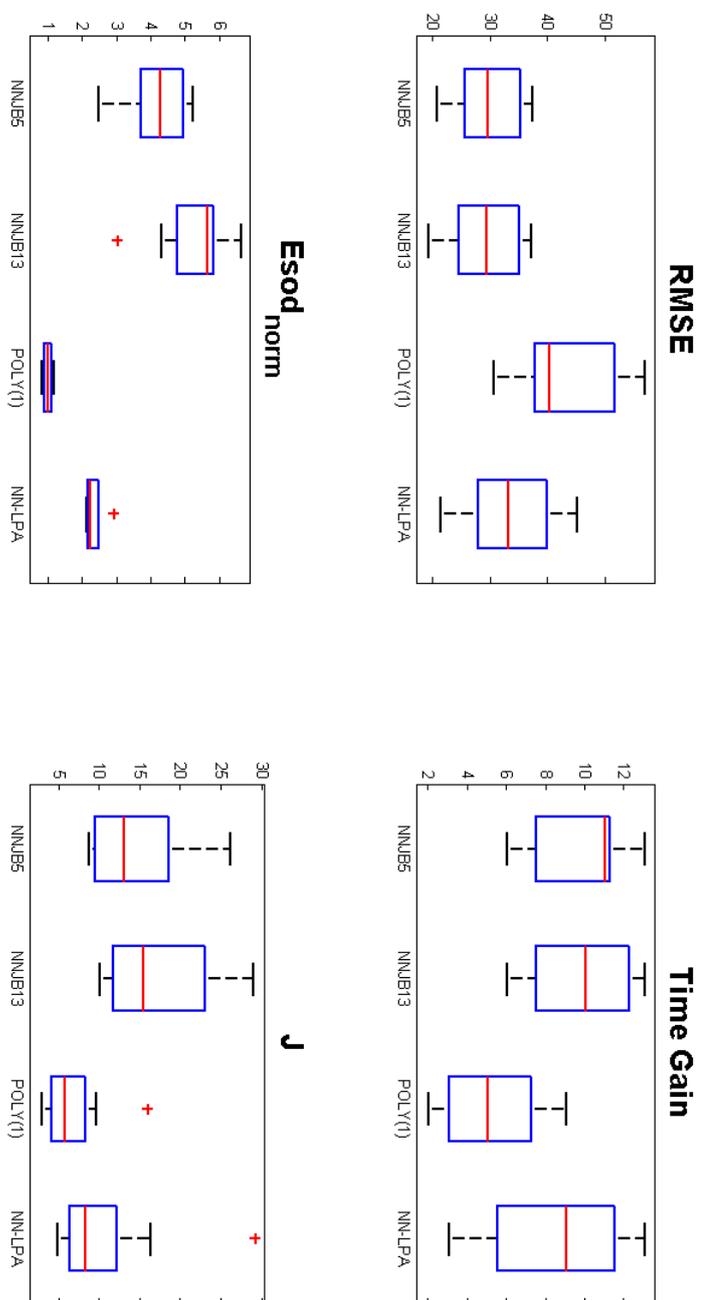


Figura 6.7: Boxplot dei risultati per i quattro indici di qualità calcolati per tutti i modelli in competizione; (a) RMSE; (b) Time Gain; (c) $ESOD_{norm}$; (d) J ; (v. descrizione Figura ??)

| | Ipoglicemia | | Euglicemia | | Iperglicemia | |
|---------|-------------|---------|------------|---------|--------------|---------|
| | Accurate | Erronee | Accurate | Erronee | Accurate | Erronee |
| NNJB5R | 28.5% | 71.5% | 89.3% | 10.7% | 84.4% | 15.6% |
| NNJB13R | 52.45% | 47.55% | 89.3% | 10.7% | 84.5% | 15.5% |

Table 6.4: Risultati medi percentuali da analisi CG-EGA per le reti jump B proposte, valutati su data set reale Dexcom[®] 7-Plus

A. Nel caso dell'indice J, invece, tali differenze non sono state individuate. Questo ulteriore test ci porta a concludere che la rete NNJB5, versione ridotta della NNJA5, sembra bilanciare bene la capacità di produrre risultati soddisfacenti con una architettura neurale di complessità ridotta rispetto alle altre reti jump candidate

Infine, anche per le reti jump B è stata effettuata un'analisi CG-EGA, i cui risultati vengono presentati in Tabella 6.4. Essi non si discostano molto da quelli ottenuti precedentemente, con un leggero calo d'accuratezza per la rete NNJB5 nella predizione di episodi ipoglicemici. Valgono tuttavia le stesse considerazioni fatte prima relativamente al numero di soggetti a disposizione e alla complessità intrinseca delle dinamiche glicemiche reali.

6.3 Considerazioni riassuntive

In questo capitolo abbiamo inizialmente utilizzato per l'analisi delle reti jump candidate un database ottenuto da soggetti reali monitorati per 5 giorni consecutivi con campionamento a 5 minuti. In Sezione 6.1 abbiamo valutato le prestazioni di questi modelli sia in termini di indici di qualità (RMSE, Time Gain, ESOD_{norm} e J) che per mezzo di un'analisi visiva dei profili predetti. I risultati sono stati confrontati inoltre con quelli ottenuti dai modelli stato dell'arte poly(1) [21] e NN-LPA [24].

Le reti NNJA5 e NNJA13 hanno ottenuto dei risultati allineati a quelli della rete NN-LPA, con profilo predetto sufficientemente aderente al target ed un guadagno temporale che si assesta intorno agli 8 minuti, leggermente migliore di quanto ottenuto con il modello polinomiale del primo ordine. Le due reti, tuttavia, producono dei profili leggermente più irregolari della rete proposta da Zecchin et al., con delle oscillazioni ad alta frequenza che occasionalmente generano dei falsi allarmi in concomitanza delle soglie di ipo- e iperglicemia. Una nota a parte va fatta per la rete NNJA2, la quale non s'è dimostrata adeguata in questa fase di test a predire la glicemia futura in modo soddisfacente. Questo fatto è probabilmente riconducibile sia alla ridotta complessità della rete, che di conseguenza non è in grado di apprendere appieno le complesse dinamiche che interessano le serie temporali ottenute da soggetti reali, sia all'esiguo numero di dati a disposizione.

I risultati ottenuti per la rete NNJA2 ci hanno indotto (Sezione 6.2) a non proporre un'alternativa B della rete, in quanto le alternative testate non fornivano nessun guadagno effettivo in termini di prestazioni. Per le reti NNJA5 e NNJA13, invece, un'analisi dei pesi sinaptici ha permesso di formulare due nuovi modelli jump B, uno per ciascuna rete. Queste reti, nella tipologia B, sono state successivamente riallenate e testate sul medesimo data set utilizzati per le reti A. Abbiamo visto che il passaggio alla tipologia B non ha comportato un peggioramento nelle prestazioni delle reti. Nel caso della rete NNJB13, anzi, le performance sono leggermente migliorate.

Di particolare interesse si sono verificate essere le reti NNJA5/NNJB5. Queste infatti, sebbene ancora relativamente semplici in termini di numero di ingressi e complessità dell'architettura neurale, riescono a predire il segnale target con un anticipo statisticamente non dissimile a quello ottenuto con le più complesse NNJA13/NNJB13, mantenendo un profilo sufficientemente aderente al target e abbastanza

buono in termini di regolarità. Dovendo quindi operare una scelta di compromesso tra qualità dei risultati e complessità del modello utilizzato, nel caso di dataset reale la coppia di reti NNJA5/NNJB5 sembra essere la più adatta.

Capitolo 7

Risultati su data set reale Abbott NavigatorTM

7.1 Predizione del profilo glicemico

Al fine di svolgere un'analisi più completa delle prestazioni ottenute dalle reti jump, abbiamo deciso di svolgere dei test anche sul dataset DIAdvisor con campionamento ad 1 minuto utilizzato già in [24]. Sebbene consci del fatto che il passaggio a serie temporali ad 1 minuto richiederebbe di sviluppare da principio tutta l'analisi svolta con le serie a 5 minuti per individuare le architetture neurali migliori, abbiamo comunque deciso di tentare un confronto utilizzando le reti candidate proposte per il dataset a 5 minuti, lasciando ad un eventuale lavoro futuro lo studio di architetture diverse ottimizzate per le serie temporali ad un minuto.

Le Figure (7.1-7.3) (Soggetti n°1 e n°3 data set reale Abbott NavigatorTM) riportano i profili predetti dalle tre reti jump candidate. Possiamo vedere delle differenze sostanziali tra i risultati ottenuti per il soggetto 1 e quelli ottenuti per il soggetto 3. Sono presenti, nello specifico, nel soggetto 1 alcune oscillazioni a livello delle soglie che possono dare origine a falsi segnali d'allarme. Nel complesso, le reti sembrano comportarsi abbastanza bene nella predizione del segnale target.

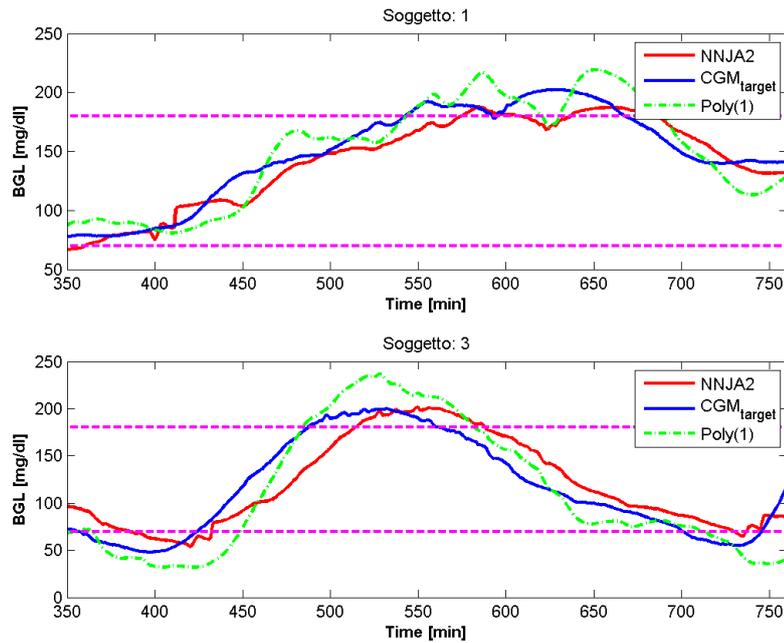


Figura 7.1: Predizione della glicemia ottenuta con NNJA2: (a) Soggetto 1; (b) Soggetto 7; (In blu viene indicato il segnale glicemico target; in verde tratteggiato la predizione ottenuta con il modello poly(1); in rosso la predizione ottenuta con la rete jump; in magenta tratteggiato vengono indicate le soglie di ipo/iperglicemia)

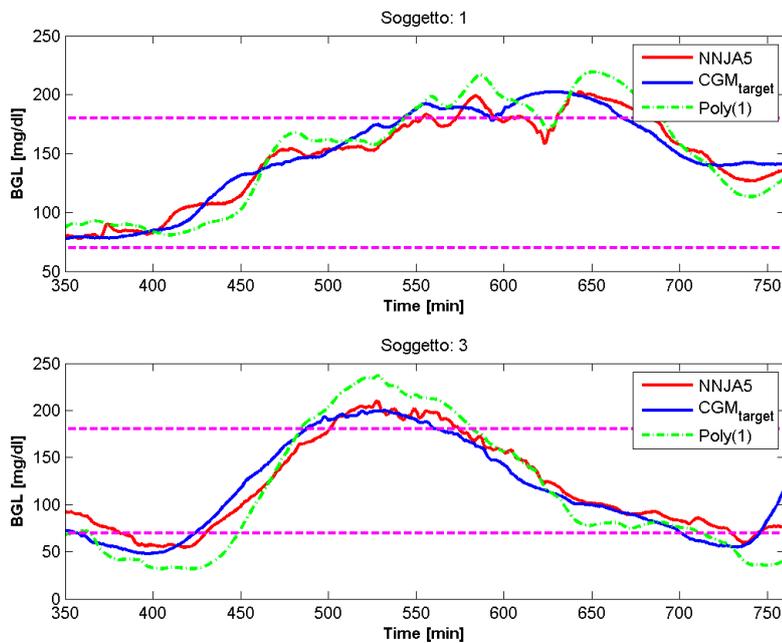


Figura 7.2: Predizione della glicemia ottenuta con NNJA5: (a) Soggetto 1; (b) Soggetto 7; (v. sopra descrizione)

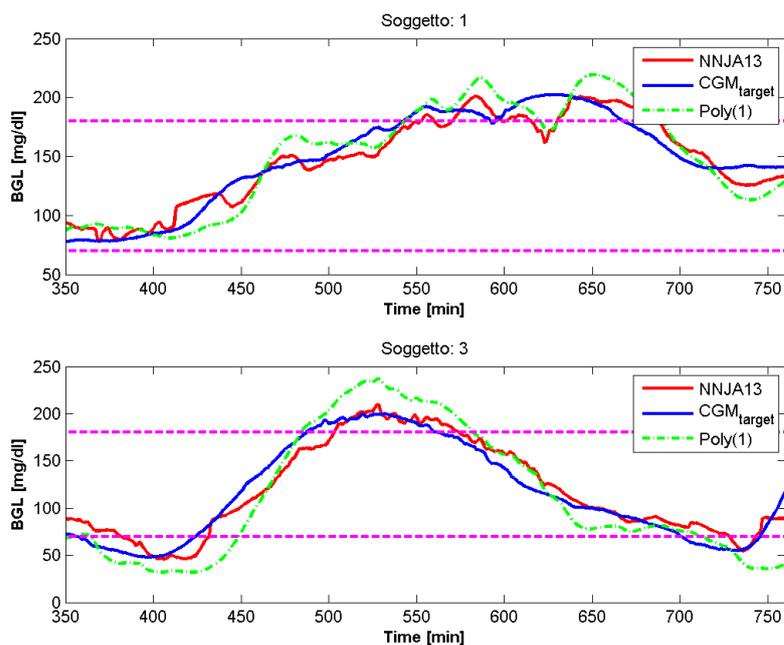


Figura 7.3: Predizione della glicemia ottenuta con NNJA13: (a) Soggetto 1; (b) Soggetto 7; (v. sopra descrizione)

Questo viene confermato anche dai risultati raccolti in tabella 7.1, nella quale si riportano i valori ottenuti dai modelli analizzati per i quattro indici di qualità. Si può vedere come le reti jump si allineino, in termini di performance, con gli altri modelli stato dell'arte: l'aderenza al target risulta buona, come anche l'anticipo ottenuto sulla predizione e la regolarità del profilo predetto. Fa in lieve misura eccezione la rete NNJA2, il cui guadagno temporale è ridotto rispetto a quello dei concorrenti NNJA5 e NNJA13. Tali risultati sono individuabili graficamente anche grazie al boxplot in Figura 7.4.

Per tutti gli indici sono stati effettuati test statistici di significatività. Per quanto riguarda l'aderenza al target, NNJA2/NNJA5 e NNJA5/NNJA13 sono risultate statisticamente simili, a differenza invece della coppia NNJA2/NNJA13. Come precedentemente illustrato, per quanto riguarda il guadagno temporale i test statistici hanno confermato la similitudine dei risultati delle reti NNJA5 e NNJA13, mentre non è stato possibile riscontrare ciò confrontando le due reti con la più semplice NNJA2. In termini di regolarità del profilo predetto, le tre reti non risultano statisticamente confrontabili. Infine, l'analisi statistica relativa all'indice J ci porta a concludere che per questo indice le coppie NNJA2/NNJA5 e NNJA2/NNJA13 risultano statisticamente simili, mentre non possiamo giungere a tale risultato per quanto riguarda il confronto tra NNJA5 e NNJA13.

Anche in questo caso è stata effettuata un'analisi dell'accuratezza della predizione ottenuta con le reti candidate per mezzo del test CG-EGA. In tabella 7.2 vengono riportate le percentuali di misure accurate ottenute dalle tre reti. È evidente come le reti jump siano in grado di cogliere la maggioranza degli eventi in tutti i range d'interesse. In particolare si vuol porre l'attenzione sui risultati ottenuti in ipoglicemia, dove più del 70% degli eventi vengono colti, sebbene tali eventi risultino sempre i più difficili da analizzare.

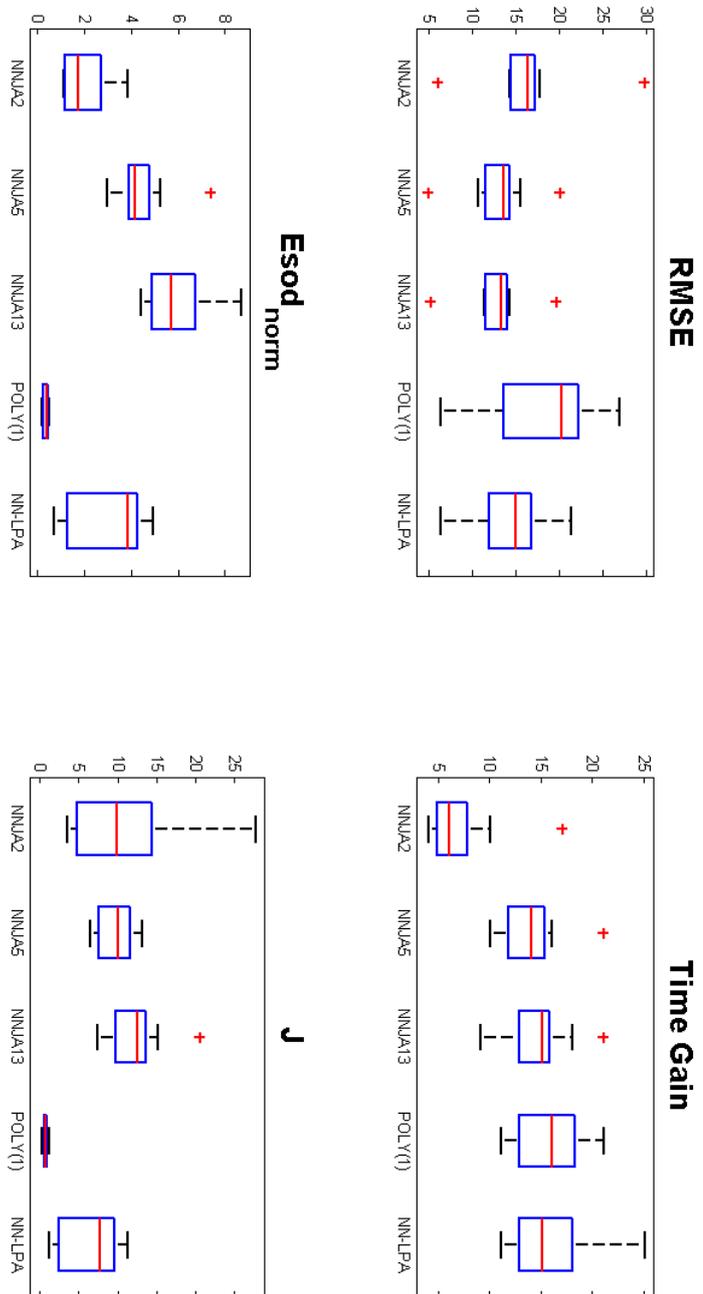


Figura 7.4: Boxplot dei risultati per i quattro indici di qualità calcolati per tutti i modelli in competizione; (a) RMSE; (b) Time Gain; (c) ESOD_{norm}; (d) J; (tratto rosso: valore mediano; bordi del box: 25mo e 75mo percentile dei data set analizzato; tratti neri esterni: estremi del dataset analizzato, non considerabili come outliers; croci rosse: outliers)

| | NNJA2 | | NNJA5 | | NNJA13 | | POLY(1) | | NN-LPA | |
|-----------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|
| | media | sd |
| RMSE | 16.2201 | 6.1044 | 12.8888 | 3.9998 | 12.7345 | 3.7413 | 18.1094 | 6.2673 | 14.2404 | 4.2505 |
| Time Gain | 7.1111 | 4.1366 | 14.1111 | 3.2189 | 14.7778 | 3.4197 | 15.6667 | 3.3166 | 15.8889 | 4.1966 |
| ESOD norm | 2.0412 | 1.0443 | 4.4365 | 1.2491 | 5.9168 | 1.4309 | 0.3033 | 0.1148 | 2.9973 | 1.6312 |
| J | 10.0952 | 8.0343 | 9.6163 | 2.3645 | 12.4973 | 3.7853 | 0.6116 | 0.2664 | 6.1431 | 3.8880 |

Table 7.1: Risultati (media e standard deviation) valutati su dataset Abbott Navigator™ per i modelli jump, per il predittore polinomiale e per la rete NN-LPA

| | Ipoglicemia | | Euglicemia | | Iperglicemia | |
|--------|--------------------|---------|-------------------|---------|---------------------|---------|
| | Accurate | Erronee | Accurate | Erronee | Accurate | Erronee |
| NNJA2 | 72.1% | 27.9% | 98.4% | 1.6% | 94.6% | 5.4% |
| NNJA5 | 75.1% | 24.9% | 97.6% | 2.4% | 92.5% | 7.5% |
| NNJA13 | 77.5% | 22.5% | 97.1% | 2.9% | 92.4% | 7.6% |

Table 7.2: Risultati percentuali CG-EGA per le tre reti jump A proposte, valutati su data set Abbott Navigator™

7.2 Reti jump di tipologia B

7.2.1 Strutturazione delle 3 reti candidate

Seguendo l'iter già presentato in Sezione 3.7 e 6.2.1, abbiamo individuato tre reti jump B con le seguenti caratteristiche:

NNJB2

Viene eliminata la sinapsi che collega l'ingresso relativo alla serie delle differenze prime della rate of appearance al livello d'uscita, lasciando inalterata la connessione allo strato nascosto.

NNJB5

Anche in questo caso l'architettura della rete viene modificata attraverso l'eliminazione della connessione sinaptica tra l'ingresso relativo alla serie delle differenze prime della rate of appearance e il livello d'uscita.

NNJB13

Un'analisi dei pesi ci ha portato all'eliminazione della sinapsi che collega il livello d'uscita con l'ingresso relativo alla serie temporale della rate of appearance del glucosio.

7.2.2 Risultati

Come è possibile notare dai grafici riportati nelle Figure (7.5 - 7.7), i profili ottenuti a mezzo delle reti jump B non si discostano molto da quelli prodotti dalle reti jump A. Valgono quindi anche qui i commenti fatti precedentemente.

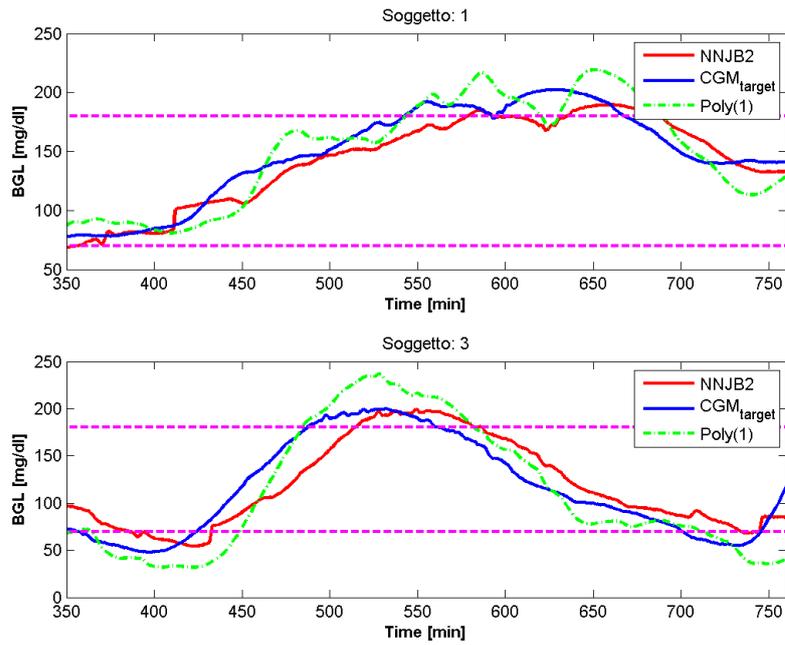


Figura 7.5: Predizione della glicemia ottenuta con NNJB2: (a) Soggetto 1; (b) Soggetto 7; (v. sopra descrizione)

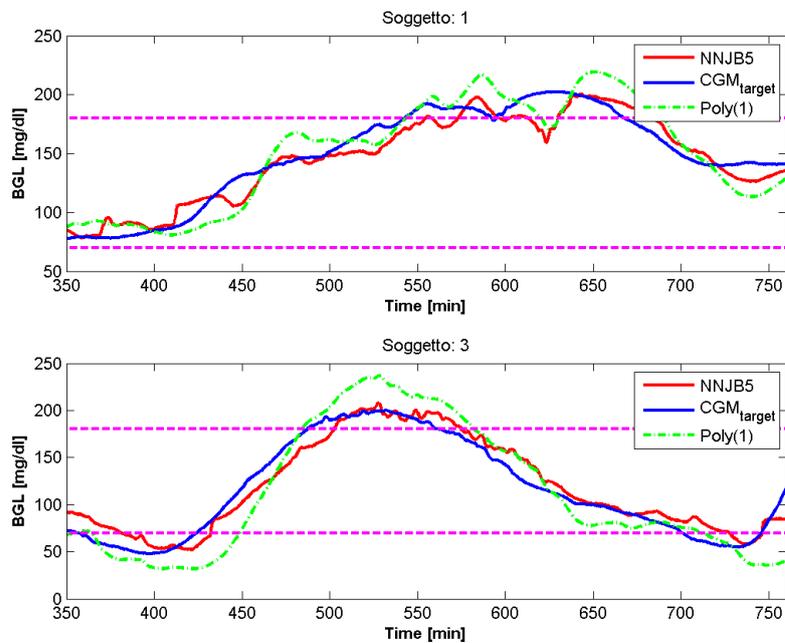


Figura 7.6: Predizione della glicemia ottenuta con NNJB5: (a) Soggetto 1; (b) Soggetto 7; (v. sopra descrizione)

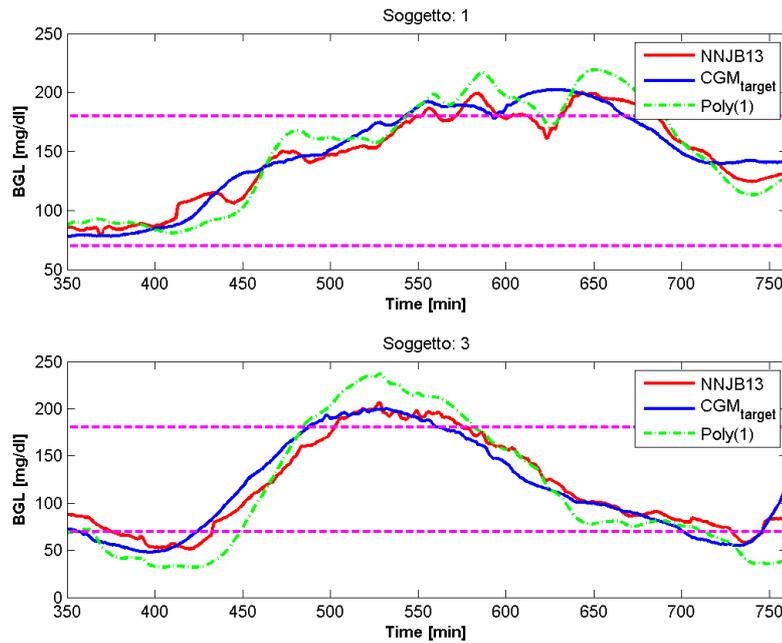


Figura 7.7: Predizione della glicemia ottenuta con NNJB13: (a) Soggetto 1; (b) Soggetto 7; (v. sopra descrizione)

Un'analisi degli indici di qualità (Figura 7.8 e Tabella 7.3) evidenzia come le reti riescano a generare dei profili aderenti al target e con un guadagno temporale confrontabile con quello dei due predittori stato dell'arte. In termini di indici $ESOD_{norm}$ e J , nuovamente le reti jump non si discostano di molto da quanto è in grado di fare la rete NN-LPA, sebbene tutte e quattro le reti neurali presentino un profilo meno regolare di quello ottenuto con il modello polinomiale, che tende alla sovraregolarizzazione.

| | NNJB2 | | NNJB5 | | NNJB13 | |
|-----------|---------|---------|---------|--------|---------|--------|
| | media | sd | media | sd | media | sd |
| RMSE | 16.2757 | 6.0609 | 12.9271 | 3.6981 | 12.9223 | 3.9632 |
| Time Gain | 6.1111 | 3.9511 | 13.8889 | 2.5221 | 14.4444 | 3.2447 |
| ESOD norm | 2.3147 | 1.1316 | 5.5515 | 1.4641 | 4.8811 | 1.2280 |
| J | 15.3974 | 11.5007 | 12.2054 | 3.2276 | 10.4253 | 2.7452 |

Table 7.3: Risultati medi valutati su dataset reale con campionamento ad 1 minuto per le reti di tipo jump B, per il modello polinomiale di ordine 1 e per la rete proposta da Zecchin et al.

Test statistici di significatività hanno dimostrato che le reti NNJB2/NNJB5 e NNJB2/NNJB13 presentano risultati statisticamente non simili per quanto riguarda gli indici TG e $ESOD_{norm}$, mentre sono simili in riferimento al RMSE e a J . Le reti NNJB5/NNJB13 invece sono simili rispetto sia all'aderenza al target che al guadagno temporale, mentre non risultano statisticamente simili in termini di regolarità.

Per verificare se il passaggio dalla tipologia A alla tipologia B avesse portato a delle differenze statisticamente rilevanti, anche i risultati ottenuti con i due modelli sono stati confrontati. Ciò che s'è visto è che in generale il passaggio alla tipologia B non ha portato differenze significative nei risultati relativi all'aderenza e al guadagno temporale, mentre per gli indici di regolarità le varie reti risultano statisticamente non simili.

Infine, un'analisi di accuratezza delle predizioni ottenute con le reti jump B è stata svolta, i cui risultati sono raccolti in Tabella 7.4. Risulta chiaro come il passaggio alla tipologia B non infici più di molto i

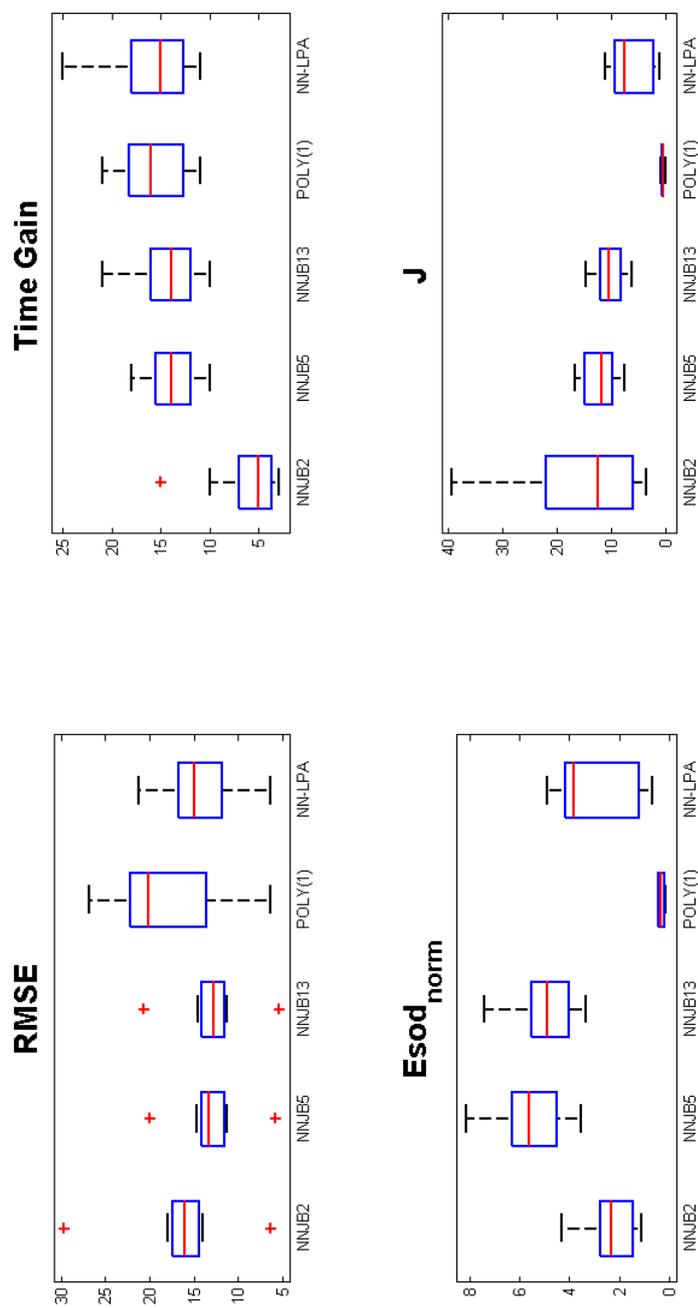


Figura 7.8: Boxplot dei risultati per i quattro indici di qualità calcolati per tutti i modelli in competizione; (a) RMSE; (b) Time Gain; (c) $ESOD_{norm}$; (d) J ; (v. sopra descrizione)

risultati ottenuti in predizione, predizione che riesce a cogliere la maggioranza degli eventi in tutti i range d'interesse.

| | Ipoglicemia | | Euglicemia | | Iperglicemia | |
|--------|--------------------|---------|-------------------|---------|---------------------|---------|
| | Accurate | Erronee | Accurate | Erronee | Accurate | Erronee |
| NNJB2 | 70.50% | 29.50% | 98.73% | 1.27% | 94.97% | 5.03% |
| NNJB5 | 75.92% | 24.08% | 97.09% | 2.91% | 92.50% | 7.50% |
| NNJB13 | 77.73% | 22.27% | 97.28% | 2.72% | 92.54% | 7.46% |

Table 7.4: Risultati percentuali CG-EGA per le tre reti jump B proposte, valutati su data set Abbott Navigator™

7.3 Considerazioni riassuntive

Per completare il quadro rappresentativo delle reti jump in ambito reale, le tre candidate sono state testate anche su data set Abbott Navigator™ con campionamento ad 1 minuto già utilizzato in [24]. In Sezione 7.1, abbiamo visto come, mediamente, le reti producano dei risultati equiparabili a quelli ottenuti con i modelli stato dell'arte poly(1) e NN-LPA, in termini di aderenza al profilo target e guadagno temporale della predizione. In termini di regolarità invece le reti jump producono dei profili leggermente meno regolari degli altri modelli candidati. Successivamente in Sezione 7.2, l'analisi dei pesi ci ha permesso di formulare tre reti jump B che, una volta allenate e testate, non hanno evidenziato sostanziali cambiamenti nelle performance rispetto a quanto ottenuto con le reti jump A. Molto incoraggianti sono stati i risultati ottenuti in termini di accuratezza delle predizioni, con le tre reti candidate in grado di cogliere la maggioranza degli eventi glicemici, siano essi critici o meno.

Nonostante i risultati soddisfacenti ottenuti in questa fase di test, ci teniamo a sottolineare che l'analisi effettuata con il data set Abbott Navigator™ risulta solamente preliminare, in quanto sarebbe necessario studiare delle architetture ad-hoc per trattare adeguatamente queste serie temporali, che per quanto poco sono comunque diverse da quelle elaborate nel caso di data set simulato e data set reale Dexcom® 7-Plus.

Capitolo 8

Conclusioni e sviluppi futuri

8.1 Riassunto del lavoro svolto

Scopo di questo elaborato è stato la valutazione di nuove architetture neurali per la predizione della glicemia futura mediante sensori CGM abbinati a modelli dell'assorbimento del glucosio dopo pasto. Nel Capitolo 3 è stato proposto un diverso modello di rete neurale, denominato jump-net, che si differenzia dalle normali reti feedforward per il fatto che le connessioni sinaptiche non sono presenti solo tra gli ingressi e lo strato di neuroni nascosti ma anche tra gli ingressi e lo strato d'uscita. Come ci si attendeva, questa particolare architettura è in grado di cogliere contemporaneamente sia le dinamiche non lineari, carpite dallo strato di neuroni nascosti, sia le dinamiche puramente lineari, elaborate invece attraverso lo strato d'uscita avente funzione d'attivazione lineare.

Abbiamo sviluppato due tipologie diverse di reti jump: la prima, da noi chiamata tipologia A (Sezione 3.3), i cui ingressi erano completamente connessi sia allo strato nascosto che allo strato d'uscita; la seconda, che abbiamo definito tipologia B (Sezione 3.7), per la quale è stata invece effettuata un'analisi quantitativa dei pesi sinaptici che ci ha permesso di selezionare solo le connessioni più importanti della rete eliminando quelle con peso trascurabile. Tre modelli di complessità crescente sono stati infine qui proposti:

- la rete NNJA2, avente cinque neuroni nello strato nascosto e alimentata in ingresso con il segnale relativo alla serie glicemica, alla rate of appearance futura del glucosio e alla serie delle derivate prime di quest'ultima;
- la rete NNJA5, anch'essa con cinque neuroni nello strato nascosto e alimentata dalle serie di glicemia e della rate of appearance del glucosio futura, nonché dalle serie delle derivate prime di queste due;
- la rete NNJA13, avente invece otto neuroni nascosti e alimentata dal segnale relativo all'errore commesso dal predittore lineare calcolato all'istante attuale rispetto alla predizione fatta 30 minuti prima, la serie delle differenze prime di questo segnale e le serie della glicemia attuale e della rate of appearance della glicemia futura, nonché le serie delle differenze prime di questi due ultimi segnali.

Per ottimizzare l'architettura delle diverse reti abbiamo utilizzato un algoritmo di *k-fold cross validation*, illustrato nel Capitolo 3. Tutte le reti jump sono state sottoposte a training sia con dati simulati, sia con dati reali, utilizzando serie temporali con periodo di campionamento pari a 5 minuti, ottenute nel caso dei dati reali attraverso sensore per il continuous glucose monitoring Dexcom® 7-Plus. I risultati ottenuti con le reti jump A e B sono stati confrontati con quelli di altri due modelli stato dell'arte: il

predittore polinomiale di ordine uno proposto da Sparacino et al. in [21] e la rete neurale feedforward presentata da Zecchin et al. in [24].

8.2 Bilancio dei risultati

Le tre reti jump candidate sono state allenate per la stima della glicemia futura utilizzando un'orizzonte di predizione di 30 minuti. I risultati sono stati molto incoraggianti, soprattutto su dati simulati (Capitolo 5). Per quanto concerne i dati reali (Capitolo 6), il numero ridotto di soggetti a disposizione ha reso più difficile l'allenamento delle reti, che per loro natura necessiterebbero di un numero di informazioni maggiore per poter sviluppare buone capacità predittive e di generalizzazione.

Per completezza, le reti candidate sono state allenate anche su data set DIAAdvisor, con serie temporali ad 1 minuto, al fine di confrontare i risultati presentati in letteratura da modelli stato dell'arte con quelli delle reti jump qui proposte (Capitolo 7). Queste ultime sono risultate in grado di produrre performance in linea con i risultati dei predittori stato dell'arte.

I risultati ottenuti sul data set reale hanno evidenziato le grandi potenzialità delle reti jump che generano delle predizioni in generale migliori di quelle ottenibili con modello polinomiale del primo ordine e in grado di competere con quelle della rete NN-LPA, con un'accuratezza e un guadagno temporale molto elevati, mantenendo una buona regolarità del segnale.

I risultati ottenuti sul data set reale Dexcom® 7-Plus, sebbene di dimensioni ridotte rispetto al data set simulato, hanno nuovamente messo in luce le capacità delle reti jump di generare dei profili che, rispetto a quelli ottenuti con modello poly(1), risultano generalmente più accurati e con un guadagno temporale più alto, mantenendo comunque una regolarità buona. Anche con il data set reale Abbott Navigator™ le reti si sono dimostrate in grado di concorre bene con i modelli stato dell'arte considerati.

8.3 Possibili sviluppi futuri

Lo sviluppo di questa tesi ci ha permesso di sottolineare molti dei vantaggi ottenibili dall'uso delle reti jump per la predizione della glicemia futura. Nel corso dell'analisi, contemporaneamente, sono stati anche evidenziati alcuni limiti legati per esempio ai datasets a disposizione e all'approccio per la valutazione delle prestazioni. A tal proposito, presentiamo ora alcuni dei problemi ancora aperti che sarebbe interessante indagare in futuro.

a) Nuovi approcci all'allenamento: l'algoritmo per l'allenamento delle reti neurali utilizzato in questo elaborato prevede la minimizzazione del RMSE. Se da una parte questo fa sì che la funzione predetta risulti il più possibile aderente al target, contemporaneamente v'è il rischio che le reti seguano anche dei segnali non legati al target stesso, come per esempio il rumore sovrapposto. Inoltre, il RMSE non ingloba alcuna informazione sulla regolarità del profilo predetto, caratteristica molto importante nell'ambito del segnale glicemico.

Potrebbe dunque essere utile, in primo luogo, studiare delle metodologie alternative per l'allenamento delle reti neurali: l'uso di nuove funzioni obiettivo, che magari inglobino opportunamente le informazioni sull'aderenza e la regolarità del profilo predetto, come anche l'ottimizzazione di parametri diversi a seconda del problema in analisi (vedi guadagno temporale), potrebbe risultare particolarmente vantaggioso .

b) Studio di nuovi indici di qualità: in seconda battuta, gli indici qualitativi utilizzati in questa tesi (RMSE, TG, $ESOD_{norm}$ e J) sono risultati molto buoni, ma a volte non sufficienti per dare un giudizio complessivo della validità delle predizioni ottenute prescindendo da un'analisi visiva. Potrebbe dunque essere interessante individuare dei nuovi parametri per la valutazione della qualità o affinare ulteriormente quelli qui adottati. Ad esempio, si potrebbe porre l'attenzione sull'individuare un metodo differente per il calcolo del guadagno temporale, pesando in maniera diversa alle modalità con cui intendiamo il ritardo in ipo/eu/iperglicemia.

c) Individuazione di nuovi segnali utili alla predizione delle ipoglicemie: per concludere, abbiamo visto nel corso di questo elaborato come le reti neurali, nello specifico le reti jump, riescano a predire molto bene le fasi di crescita della glicemia in seguito all'ingerimento di carboidrati, soprattutto grazie all'uso della funzione rate of appearance ad essi associata. Le reti, tuttavia, non hanno a disposizione alcuna informazione specifica sul come la glicemia invece decresca. In questo caso la storia passata del segnale glicemico è l'unico ingresso fruibile.

Potrebbe essere quindi molto interessante individuare un qualche segnale correlato a queste fasi di discesa, che abbinato alle misure glicemiche precedenti e alla velocità di comparsa del glucosio nel plasma fornisca alla rete informazioni utili per predire in modo più accurato anche gli eventi ipoglicemici, che ricordiamo essere quelli potenzialmente più dannosi e più difficili da prevedere.

8.4 Considerazioni finali

In questa tesi abbiamo messo in luce i vantaggi ottenibili nella predizione della glicemia futura attraverso l'adozione di nuove reti neurali: le reti jump. Grazie ad un'architettura relativamente semplice, queste reti sono state in grado di integrare facilmente informazioni provenienti da domini diversi, come ad esempio la storia passata del segnale e le informazioni sui carboidrati ingeriti, e di gestire attraverso un unico modello le dinamiche lineari/non lineari del profilo glicemico ottenuto per mezzo di sensori CGM. Le performance delle tre reti candidate sono risultate in linea con i predittori stato dell'arte considerati in questo elaborato: poly(1) (Sparacino et al. [21]) e NN-LPA (Zecchin et al. [24]).

Come accennato negli sviluppi futuri, vi sono ancora dei problemi aperti su come sfruttare al meglio le potenzialità di queste nuove architetture che, come si può dire per le reti neurali in generale, risultano molto affascinanti, soprattutto per la loro versatilità e capacità di risolvere problemi di svariata natura, che spaziano in ambiti come l'economia, la meteorologia, fino anche alla bioingegneria. I risultati di questa tesi, inoltre, incoraggiano l'approfondimento di queste nuove architetture per la predizione della glicemia futura nei pazienti diabetici, con l'idea di poterle negli anni a venire implementare *on-chip* con lo scopo di favorire e supportare i pazienti nella gestione della patologia diabetica.

Bibliografia

- [1] www.who.int
- [2] <http://www.progettodiabete.org>
- [3] www.diabete.net
- [4] <http://www.medcompare.com/details/33037/CGMS-System-Gold-Continuous-Glucose-Monitoring.html>
- [5] <http://www.menarini.com>
- [6] <http://www.dexcom.com/>
- [7] <http://www.abbottdiabetescare.com>
- [8] <http://www.minimed.com/products/guardian/>
- [9] <http://www.solianis.com>
- [10] Cobelli C., Bonadonna R., Bioingegneria dei sistemi metabolici, Patron editore, 1998
- [11] Haykin S., Neural Networks - A Comprehensive Foundation - Second Edition, Prentice Hall International, 1999
- [12] Ajoy K. P., Dobrivoje P., Computational Intelligence in Time Series Forecasting, Springer, 2005
- [13] Paul D. McNelis, "Neural Networks in Finance, gaining predictive edge in the market", Elsevier Academic Press, 2005
- [14] E. Hulthen and M. Wahde. Improved time series prediction using evolutionary algorithms for the generation of feedback connections in neural networks. In Computational Finance 2004.
- [15] Bishop C. M., "Neural Networks for Pattern Recognition", Oxford University Press, 2005.
- [16] Zecchin C., Reti Neurali per la Predizione della Glicemia Futura mediante Sensori per il Continuous Glucose Monitoring, 2010
- [17] Tierney M.J., Tamada J.A., Potts R.O., Eastman R.C., Pitzer K., Ackerman N.R., Fermi S.J., The GlucoWatch Biographer: a frequent, automatic and noninvasive glucose monitor, Ann Med 2000 Dec;32(9):632-41.
- [18] Weiss R., Yegorchikov Y., Shusterman A., Raz I., Noninvasive continuous glucose monitoring using photoacoustic technology-results from the first 62 subjects, Diabetes Technol Ther, 2007 Feb;9(1):68-74.

- [19] Amir O., Weinstein D., Zilberman S., Less M., Perl-Treves D., Primack H., Weinstein A., Gabis E., Fikhte B., Karasik A., Continuous Noninvasive Glucose Monitoring Technology Based on “Occlusion Spectroscopy”, *J Diabetes Sci Technol*, 2007 Jul;1(4):463-9.
- [20] Rossetti P., Bondia J., Vehì J., Fanelli C.G., Estimating Plasma Glucose from Interstitial Glucose: The Issue of Calibration Algorithms in Commercial Continuous Glucose Monitoring Devices, *Sensors* 2010, 10, 10938-10939.
- [21] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, and C. Cobelli, “Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series,” *IEEE Trans Biomed Eng*, vol. 54, no. 5, pp. 931–937, 2007.
- [22] C. P´erez-Gand´ia, A. Facchinetti, G. Sparacino, C. Cobelli, E. G´omez, M. Rigla, A. de Leiva, and M. Hernando, “Artificial neural network algorithm for on-line glucose prediction from continuous glucose monitoring,” *Diabetes Tech Ther*, vol. 12, no. 1, pp. 81–88, 2010.
- [23] S. Pappada, B. Cameron, P. Rosman, R. Bourey, T. Papadimos, W. Oloruntu, and M. Borst, “Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes,” *Diabetes Tech Ther*, vol. 13, no. 2, pp. 135–141, 2011.
- [24] C. Zecchin, A. Facchinetti, G. Sparacino, G. De Nicolao, and C. Cobelli, “A New Neural Network Approach for Short-Term Glucose Prediction Using Continuous Glucose Monitoring Time-Series and Meal Information”, 10.1109/IEMBS.2011.6091368.
- [25] C. Dalla Man, R.A. Rizza, and C. Cobelli. Meal simulation model of the glucose insulin system. *IEEE Transactions on Biomedical Engineering*, 54(10):1740-1749, 2007.
- [26] Facchinetti A, Sparacino G, Trifoglio E, Cobelli C. A new index to optimally design and compare continuous glucose monitoring glucose prediction algorithms. *Diabetes Technol Ther*. 2011 Feb;13(2):111-9.
- [27] Dalla Man C., Toffolo G, Basu R, Rizza R, Cobelli C. A model of glucose production during a meal. *Proceedings of IEEE-EMBS conference* pp 5647-5650, 2006
- [28] Facchinetti A, Sparacino G, Trifoglio E, Cobelli C. A new index to optimally design and compare continuous glucose monitoring glucose prediction algorithms. *Diabetes Technol Ther*. 2011 Feb;13(2):111-9. PubMed PMID: 21284477.
- [29] I. M. E. Wentholt . J. B. L. Hoekstra . A. Zwart . J. H. DeVries. Pendra goes Dutch: lessons for the CE mark in Europe. *Diabetologia* (2005) 48: 1055–1058
- [30] Giovanni Sparacino, Andrea Facchinetti and Claudio Cobelli. “Smart” Continuous Glucose Monitoring Sensors: On-Line Signal Processing Issues. *Sensors* 2010, 10, 6751-6772;
- [31] Reifman, J.; Rajaraman, S.; Gribok, A.; Ward, W.K. Predictive monitoring for improved management of glucose levels. *J. Diabetes Sci. Technol*. 2007, 1, 478-486.
- [32] Eren-Oruklu, M.; Cinar, A.; Quinn, L.; Smith, D. Estimation of future glucose concentrations with subject-specific recursive linear models. *Diabetes Technol. Ther*. 2009, 11, 243-253
- [33] Gani, A.; Gribok, A.V., Rajaraman, S.; Ward, W.K.; Reifman, J. Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. *IEEE Trans. Biomed. Eng*. 2009, 56, 246-254

- [34] Lu, Y.; Gribok, A.; Ward, K.; Reifman, J. The importance of different frequency bands in predicting subcutaneous glucose concentration in type 1 diabetic patients. *IEEE Trans. Biomed. Eng.* 2010 (in press).
- [35] Palerm, C.C.; Bequette, W. Hypoglycemia detection and prediction using continuous glucose monitoring—a study on hypoglycemic clamp data. *J. Diabetes Sci. Technol.* 2007, 1, 624-629.