

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI
INGEGNERIA



Dipartimento di
Ingegneria
dell'Informazione

Corso di Laurea Magistrale in Bioingegneria

Tesi di Laurea

Reti Neurali per la Predizione della Glicemia Futura mediante Sensori per il Continuous Glucose Monitoring

RELATORE: Prof. Giovanni Sparacino

CORRELATORE: Ing. Andrea Facchinetti

LAUREANDA: Chiara Zecchin

ANNO ACCADEMICO: 2009-2010

Padova, 5 Ottobre 2010

Indice

| | |
|--|-----------|
| Introduzione | 1 |
| 1 Il diabete e il monitoraggio continuo della glicemia | 5 |
| 1.1 Fisiopatologia del diabete mellito | 6 |
| 1.2 Diagnosi e classificazione del diabete | 7 |
| 1.2.1 Criteri diagnostici del diabete | 7 |
| 1.2.2 Classificazione del diabete | 8 |
| 1.3 Effetti del diabete | 9 |
| 1.3.1 Complicanze a breve termine | 9 |
| 1.3.2 Complicanze a lungo termine | 10 |
| 1.4 Il monitoraggio della glicemia | 10 |
| 1.5 I sensori CGM | 12 |
| 1.5.1 Dispositivi minimamente invasivi | 12 |
| 1.5.2 Dispositivi non invasivi | 13 |
| 1.5.3 Abbott FreeStyle Navigator™ | 14 |
| 1.5.4 Punti di forza e problemi dei dispositivi CGM | 16 |
| 1.5.5 Aspetti critici dei sensori CGM | 17 |
| 1.6 Applicazione di sensori CGM per la gestione on-line della terapia: il progetto DIAAdvisor | 19 |
| 1.6.1 Aspetti generali del progetto | 19 |
| 1.6.2 Descrizione dei dati | 20 |
| 1.6.3 Tecniche per la predizione | 21 |
| 1.7 Obiettivi e presentazione della tesi | 22 |
| 2 Introduzione alle reti neurali | 23 |
| 2.1 Reti neurali biologiche | 23 |
| 2.1.1 Apprendimento | 25 |
| 2.2 Reti neurali artificiali (ANN) | 25 |
| 2.2.1 Il neurone artificiale | 27 |
| 2.2.2 Funzioni di attivazione | 30 |
| 2.3 Architettura della rete | 33 |
| 2.3.1 Reti neurali feedforward (FFNN) | 33 |
| 2.3.2 Reti neurali ricorsive (RNN) | 35 |

| | | |
|----------|---|-----------|
| 2.3.3 | Reti neurali radial-basis function (RBFNN) | 37 |
| 2.3.4 | Reti competitive e self organizing maps | 38 |
| 2.4 | Allenamento supervisionato della rete (supervised training) . | 39 |
| 2.4.1 | Apprendimento “errore-correzione” (error-correction learning) | 39 |
| 2.4.2 | Apprendimento hebbiano | 47 |
| 2.4.3 | Apprendimento competitivo | 47 |
| 2.4.4 | Algoritmi stocastici | 48 |
| 2.5 | Predizione di serie temporali con reti neurali | 51 |
| 3 | Predizione della glicemia futura da dati CGM: stato dell’arte | 53 |
| 3.1 | Filtro di Kalman | 54 |
| 3.2 | Modello autoregressivo del primo ordine (AR(1)) | 56 |
| 3.3 | Modello polinomiale di primo grado | 59 |
| 3.4 | Reti neurali con ingressi SMBG | 61 |
| 3.4.1 | Reti neurali ricorsive ottimizzate e testate su dati relativi ad un solo soggetto | 61 |
| 3.4.2 | Reti neurali per modellare il metabolismo di un diabetico: confronto delle performance di varie architetture | 62 |
| 3.4.3 | Reti neurali per il controllo della glicemia di pazienti diabetici di tipo 1 | 63 |
| 3.4.4 | Sviluppo e confronto di reti neurali per modellare il metabolismo di un paziente diabetico di tipo 2 | 64 |
| 3.4.5 | Reti neurali distinte per modellare la variabilità circadiana del segnale glicemico | 65 |
| 3.5 | Reti neurali con ingressi CGM | 68 |
| 3.5.1 | Rete neurale con in ingresso misure di glucosio e informazioni su pasti, insulina ed esercizio fisico ricavate dalla serie glicemica stessa | 68 |
| 3.5.2 | Rete ricorsiva implementata e testata su pazienti diabetici virtuali | 69 |
| 3.5.3 | Rete neurale per la predizione del valore glicemico successivo | 70 |
| 3.5.4 | Reti neurali e kernels per predire il profilo glicemico futuro | 71 |
| 3.5.5 | Rete neurale feedforward per la predizione on-line della glicemia, sfruttando esclusivamente misure CGM . . | 72 |
| 3.6 | Conclusioni | 74 |
| 4 | Approcci alla predizione della glicemia basati su reti neurali: possibili nuove architetture e criteri di valutazione | 77 |
| 4.1 | Architetture candidate | 77 |
| 4.1.1 | Architettura della rete neurale | 77 |

| | | |
|----------|--|------------|
| 4.1.2 | Target della predizione | 81 |
| 4.2 | Scelta della struttura migliore | 82 |
| 4.3 | Metriche di valutazione delle performance della rete | 83 |
| 4.3.1 | Regolarità del profilo predetto: J | 83 |
| 4.3.2 | Aderenza del profilo predetto al target: RMSE | 83 |
| 4.3.3 | Guadagno Temporale | 84 |
| 4.3.4 | Obiettivo della predizione | 86 |
| 4.4 | Conclusioni | 87 |
| 5 | Due nuovi approcci basati su reti neurali per predire la glicemia futura | 89 |
| 5.1 | Approccio 1: rete neurale feedforward per la predizione del livello glicemico | 89 |
| 5.1.1 | Semplificazione dell'architettura della rete | 90 |
| 5.1.2 | Nuovi ingressi considerati nella rete | 94 |
| 5.1.3 | Ri-ottimizzazione del numero di neuroni nascosti | 104 |
| 5.2 | Approccio 2: rete neurale feedforward per correggere l'errore del modello lineare | 108 |
| 5.2.1 | Potenzialità e problemi della rete per la correzione dell'errore del modello lineare | 110 |
| 5.3 | Conclusioni | 114 |
| 6 | Implementazione e risultati su dati simulati | 115 |
| 6.1 | Creazione del dataset | 115 |
| 6.1.1 | Modello per la simulazione dei profili glicemici | 115 |
| 6.1.2 | Difetti e limiti dei dati simulati | 116 |
| 6.2 | Risultati: predizione del profilo glicemico | 117 |
| 6.3 | Risultati: attraversamento di soglie critiche | 121 |
| 6.4 | Risultati: performance dei predittori al variare del PH | 124 |
| 6.5 | Conclusioni | 131 |
| 7 | Implementazione e risultati su dati reali | 133 |
| 7.1 | Creazione del dataset | 133 |
| 7.2 | Risultati: predizione del profilo glicemico | 134 |
| 7.3 | Risultati: attraversamento di soglie critiche | 140 |
| 7.4 | Risultati: performance dei predittori al variare del PH | 142 |
| 7.5 | Conclusioni | 147 |
| 8 | Conclusioni e possibili sviluppi futuri | 151 |
| 8.1 | Bilancio sui risultati ottenuti | 151 |
| 8.1.1 | Risultati su dati simulati | 152 |
| 8.1.2 | Risultati su dati reali | 153 |
| 8.2 | Possibili sviluppi futuri | 154 |
| 8.2.1 | Nuovi algoritmi per l'allenamento della rete neurale | 154 |

INDICE

| | | |
|-------|---|------------|
| 8.2.2 | Reti ricorsive | 155 |
| 8.2.3 | Classificazione del livello glicemico | 156 |
| 8.3 | Conclusione | 156 |
| | Bibliografia | 164 |

Elenco delle figure

| | | |
|------|--|----|
| 1.1 | Dispositivo pungi-dito per SMBG. | 11 |
| 1.2 | Confronto tra CGM e SMBG | 11 |
| 1.3 | FreeStyle Navigator TM , sensore [51]. | 15 |
| 1.4 | FreeStyle Navigator TM , trasmettitore [51]. | 15 |
| 1.5 | FreeStyle Navigator TM , ricevitore [51]. | 15 |
| 1.6 | Confronto tra misurazioni di glucosio plasmatiche e misurazioni di glucosio interstiziale ottenute con sensore CGM. | 18 |
| 1.7 | Serie temporali misurate con sensore CGM FreeStyle Navigator TM | 19 |
| | | |
| 2.1 | Immagine di un neurone. | 24 |
| 2.2 | Potenziale d'azione tipico di un neurone. | 24 |
| 2.3 | Esempi di reti neurali feedforward e di reti neurali ricorsive [20]. | 26 |
| 2.4 | Modello di neurone non lineare [19]. | 27 |
| 2.5 | Trasformazione affine prodotta dalla presenza del bias [19]. | 29 |
| 2.6 | Riformulazione del modello del neurone. | 29 |
| 2.7 | Funzioni di attivazione | 32 |
| 2.8 | Esempi di reti neurali feedforward [19]. In Figura 2.8(a) è rappresentata una rete neurale con un solo strato di neuroni. In Figura 2.8(b) è rappresentata una ffnn con uno strato nascosto costituito da 4 neuroni e con 2 neuroni d'uscita [19]. | 33 |
| 2.9 | Rete neurale ricorsiva con neuroni nascosti [19]. | 36 |
| 2.10 | Un esempio di funzione radiale. | 38 |
| 2.11 | Grafo indicante il flusso del segnale, dall'ingresso all'uscita del neurone j [19]. | 42 |
| 2.12 | Grafo di una rete neurale allenata attraverso apprendimento competitivo | 48 |
| | | |
| 3.1 | Predizione ottenuta col modello proposto da Palerm <i>et al.</i> in [29]. | 55 |
| 3.2 | Predizione ottenuta col modello proposto da Reifman <i>et al.</i> in [36]. | 57 |
| 3.3 | Profili predetti dal modello proposto da Gani <i>et al.</i> in [17]. | 58 |
| 3.4 | Performance dei modelli ARX proposti da Finan <i>et al.</i> in [16]. | 59 |

Elenco delle figure

| | | |
|------|--|-----|
| 3.5 | Predizione ottenuta con modello AR(1) e con modello lineare, entrambi con parametri tempo-varianti, proposti da Sparacino <i>et al.</i> in [39]. | 60 |
| 3.6 | Performance della rete neurale per la predizione della glicemia futura proposta da El-Jabali in [10]. | 64 |
| 3.7 | Performance della rete neurale proposta da Zitar <i>et al.</i> in [50]. | 64 |
| 3.8 | Performance delle quattro reti neurali proposta da Kok in [24]. | 66 |
| 3.9 | Plot delle performance di una delle reti neurali proposte da Zainuddin <i>et al.</i> in [48]. | 67 |
| 3.10 | Performance della rete neurale di Eskaf <i>et al.</i> | 69 |
| 3.11 | Performance della rete neurale di Pappada <i>et al.</i> [31]. | 71 |
| 3.12 | Performance del predittore proposto da Pereverzyev <i>et al.</i> | 72 |
| 3.13 | Performance della rete neurale di Pérez-Gandía <i>et al.</i> su profilo monitorato con sensore FreeStyle Navigator TM | 74 |
| 4.1 | Asimmetria della distribuzione dei valori glicemici. | 80 |
| 4.2 | Divisione del dataset per l'applicazione di k-fold cross validation. | 82 |
| 4.3 | Difetti del RMSE | 85 |
| 4.4 | Confronto tra una predizione aderente al target e una predizione pronta in corrispondenza dell'attraversamento di soglie critiche | 86 |
| 5.1 | Andamento del RMSE e di J all'aumentare del numero di neuroni nascosti della rete neurale. | 90 |
| 5.2 | Andamento del guadagno temporale all'aumentare del numero di neuroni nascosti della rete neurale. | 92 |
| 5.3 | Confronto tra profili predetti da reti neurali con un diverso numero di neuroni nello strato nascosto. | 93 |
| 5.4 | Confronto tra la distribuzione dei valori glicemici e la distribuzione delle loro differenze prime, ottenute a partire, rispettivamente, dal segnale glicemico e dalla sua derivata. | 95 |
| 5.5 | Confronto tra l'uscita della rete con 15 differenze prime in ingresso e l'uscita della rete con la media delle differenze prime in ingresso. | 96 |
| 5.6 | Rate of appearance dei carboidrati ingeriti nell'arco di una giornata. | 98 |
| 5.7 | Plot di BGL(t), Ra meal(t) e Ra insulin(t). | 100 |
| 5.8 | Plot di BGL(t) e Ra insulin(t-120). | 101 |
| 5.9 | Plot dell'Ra e dell'AUC a confronto. | 103 |
| 5.10 | Andamento del RMSE e di J all'aumentare del numero di neuroni nascosti. | 104 |
| 5.11 | Andamento del guadagno temporale all'aumentare del numero di neuroni nascosti. | 106 |

| | |
|--|-----|
| 5.12 Profili predetti da reti neurali con un diverso numero di neuroni nascosti. | 107 |
| 5.13 Schematizzazione del sistema realizzato nell'approccio 2 per predire la glicemia futura. | 109 |
| 5.14 Confronto tra le performance della rete che predice direttamente il livello glicemico e le performance della rete che predice l'errore del modello lineare. | 111 |
| 5.15 Confronto tra livello glicemico da predire e target effettivo delle reti neurali. | 113 |
| | |
| 6.1 Profilo glicemico di un paziente virtuale, prima e dopo l'aggiunta di rumore. | 117 |
| 6.2 Predizione di serie glicemiche simulate | 119 |
| 6.3 Boxplot e tabella riassuntiva con i valori dell'indice J, per i 4 algoritmi di predizione considerati. | 120 |
| 6.4 Boxplot e tabella riassuntiva con i valori del RMSE, per i 4 algoritmi di predizione considerati. | 120 |
| 6.5 Boxplot e tabella riassuntiva con i valori del TG medio, per i 4 algoritmi di predizione considerati. | 121 |
| 6.6 Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le ipoglicemie, con i 4 algoritmi di predizione considerati. | 122 |
| 6.7 Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le iperglicemie, con i 4 algoritmi di predizione considerati. | 122 |
| 6.8 Confusion Matrix relativa all'attraversamento delle soglie ipoglicemiche e iperglicemiche. | 123 |
| 6.9 Andamento di J al variare del PH. | 125 |
| 6.10 Andamento del RMSE al variare del PH. | 125 |
| 6.11 Guadagno temporale medio al variare del PH. | 126 |
| 6.12 Guadagno temporale relativo alla predizione delle ipoglicemie, al variare del PH. | 128 |
| 6.13 Confusion Matrix al variare del PH | 129 |
| 6.14 Guadagno temporale relativo alla predizione delle iperglicemie, al variare del PH. | 130 |
| 6.15 Profilo glicemico e rate of appearance dei carboidrati in un paziente virtuale. | 132 |
| | |
| 7.1 Plot del profilo glicemico del paziente 31, e delle relative predizioni. | 134 |
| 7.2 Plot del profilo glicemico del paziente 5, e delle relative predizioni. | 135 |
| 7.3 Boxplot e tabella riassuntiva con i valori dell'indice J, per i 4 algoritmi di predizione considerati. | 135 |

Elenco delle figure

| | | |
|------|---|-----|
| 7.4 | Boxplot e tabella riassuntiva con i valori del RMSE, per i 4 algoritmi di predizione considerati. | 136 |
| 7.5 | Boxplot e tabella riassuntiva con i valori del TG medio, per i 4 algoritmi di predizione considerati. | 136 |
| 7.6 | Predizione ottenuta su un profilo poco rumoroso. | 138 |
| 7.7 | Predizione ottenuta su un profilo molto rumoroso. | 139 |
| 7.8 | Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le ipoglicemie, con i 4 algoritmi di predizione considerati. | 140 |
| 7.9 | Confusion Matrix relativa all'attraversamento delle soglie ipoglicemiche e iperglicemiche. | 141 |
| 7.10 | Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le iperglicemie, con i 4 algoritmi di predizione considerati. | 142 |
| 7.11 | Andamento di J al variare del PH. | 143 |
| 7.12 | Andamento del RMSE al variare del PH. | 143 |
| 7.13 | Guadagno temporale medio al variare del PH. | 144 |
| 7.14 | Guadagno temporale in corrispondenza dell'attraversamento della soglia ipoglicemica, al variare del PH. | 145 |
| 7.15 | Guadagno temporale in corrispondenza dell'attraversamento della soglia iperglicemica, al variare del PH. | 146 |
| 7.16 | Percentuale di vere e false crisi ipo e iperglicemiche predette, al variare del PH. | 148 |

Elenco delle tabelle

| | | |
|-----|---|-----|
| 3.1 | Confronto tra le performance delle reti neurali proposte da Kok in [24], Baghdadi <i>et al.</i> in [4] e da Zainuddin <i>et al.</i> in [48] in termini di RMSE dei vari predittori. (Tabella tratta da [48]). | 67 |
| 7.1 | Media e standard deviation del RMSE minimo dei vari predittori. | 139 |
| 8.1 | Valori mediani degli indici utilizzati per quantificare la bontà della predizione (dati simulati). | 153 |
| 8.2 | Valori mediani degli indici utilizzati per quantificare la bontà della predizione (dati reali). | 154 |

Introduzione

Il glucosio è il substrato più importante per gli esseri umani e il suo livello ematico è finemente controllato dall'insulina, attraverso un complesso sistema a feedback negativo. Nei pazienti diabetici il pancreas è incapace di secernere insulina (diabete di tipo 1), o esiste un malfunzionamento nel meccanismo di secrezione e nell'azione dell'insulina (diabete di tipo 2). In entrambi i casi il risultato è un aumento cronico della glicemia, dovuto all'alterazione del metabolismo dei principali nutrienti.

Il diabete mellito è considerato una patologia sociale, esso affligge 220 milioni di individui in tutto il mondo, e si prevede che tale numero raggiungerà i 366 milioni entro il 2030. La diffusione epidemica di questa patologia è dovuta non solo alla crescita della popolazione e all'allungamento dell'aspettativa di vita, ma anche a diete poco idonee, a sedentarietà e obesità. Da questi dati si può facilmente intuire quanto sia elevato l'impatto economico del diabete, sia per le famiglie dei pazienti che ne sono affetti, sia per i sistemi sanitari nazionali, che utilizzano tra il 2.5% e il 15% delle loro risorse economiche per curare questa patologia.

Attualmente la terapia convenzionale del diabete è basata sulla somministrazione di insulina e farmaci ipoglicemizzanti, sulla dieta e sull'esercizio fisico, ottimizzati in base a 3-4 automonitoraggi giornalieri della glicemia (Self Monitoring of Blood Glucose, SMBG), e finalizzati al mantenimento del livello di glucosio ematico all'interno del range euglicemico ([80-120]mg/dl). Purtroppo il SMBG è invasivo, è infatti necessario prelevare una goccia di sangue per effettuare ogni misurazione. Di conseguenza questa tecnica non consente di monitorare adeguatamente il livello di glucosio ematico, in quanto non è in grado di identificare tutte le escursioni glicemiche al di fuori del range di normalità che si possono verificare durante la vita quotidiana.

Per questo motivo, a partire dal XXI secolo, sono stati sviluppati e introdotti nel mercato nuovi dispositivi minimamente invasivi, che consentono di effettuare il monitoraggio quasi in continuo della glicemia, fornendo una misurazione ogni 1-5 minuti. Questi sensori per il Continuous Glucose Monitoring (CGM) sono un efficace strumento per il miglioramento della terapia del diabete. Il monitoraggio continuo della glicemia consente di individuare un numero maggiore di episodi pericolosi per la salute del paziente, rispetto al SMBG, e la grande mole di dati raccolti permette di effettuare un'analisi

retrospettiva dell'andamento glicemico stesso, in base alla quale si può aggiustare opportunamente la terapia futura. Molti dispositivi CGM, inoltre, generano allarmi quando la glicemia eccede le soglie del range di normalità, consentendo al paziente di agire prontamente per evitare l'aggravarsi della crisi ipo o iperglicemica.

L'aspetto più attraente e potenzialmente utile dei sensori CGM è la possibilità di sfruttare la grande quantità di dati fornita per prevenire, invece che semplicemente curare, le ipo e le iperglicemie, generando un allarme 20 o 30 minuti prima che l'episodio si verifichi. Alcuni dispositivi incorporano al loro interno semplici algoritmi di proiezione, che estrapolano il trend glicemico dagli ultimi campioni misurati e producono un'allerta quando si prevede che la glicemia attraverserà una soglia critica nell'immediato futuro. Inoltre sono stati recentemente proposti in letteratura metodi più sofisticati, per predire la concentrazione di glucosio nel sangue con un anticipo di almeno 30 minuti, utilizzando i dati passati registrati dal sensore CGM e modelli adeguati a descrivere le serie glicemiche. I potenziali benefici apportati da questi algoritmi sono purtroppo limitati dall'alto numero di falsi allarmi che essi generano.

Il presente elaborato di tesi si inserisce all'interno di questo filone e ha lo scopo di testare nuovi metodi per la predizione del profilo glicemico. In particolare saranno sviluppati degli algoritmi basati su reti neurali feedforward che, sfruttando informazioni sui valori glicemici passati e sulla quantità di cibo ingerito, saranno in grado di predire la glicemia futura con un anticipo di 30 minuti. I due algoritmi proposti saranno testati sia su un dataset simulato, sia su un dataset reale. I principali punti di forza delle reti neurali sono la loro capacità di modellare autonomamente complessi sistemi altamente non lineari e il fatto che esse possono sfruttare le informazioni riguardanti i pasti e l'insulina somministrata per suggerire il probabile valore glicemico futuro. Come dimostreranno gli incoraggianti risultati ottenuti le reti neurali hanno grandi potenzialità e sono un modello molto promettente da utilizzare per la predizione del livello di glucosio ematico.

Nel *capitolo 1* verrà descritta la patologia del diabete e sarà presentata una breve panoramica sui dispositivi per il CGM attualmente in commercio. Il sensore FreeStyle NavigatorTM della Abbott Diabetes Care sarà descritto dettagliatamente, poichè da esso provengono tutti i dati utilizzati nel presente elaborato. Successivamente saranno fornite alcune informazioni sul progetto DIAdvisor, all'interno del quale sono stati raccolti i dati utilizzati e, infine, saranno presentati e spiegati gli obiettivi della tesi.

Nel *capitolo 2* sarà presentata la teoria di base sulle reti neurali, verranno descritte le principali architetture e i vari metodi a disposizione per il training. Successivamente saranno forniti dei validi motivi per utilizzarle come predittori.

Nel *capitolo 3* verranno descritti gli algoritmi recentemente proposti in letteratura per predire la glicemia futura utilizzando dati provenienti da

sensori CGM. Saranno inoltre presentate alcune interessanti reti neurali che utilizzano dati SMBG.

Nel *capitolo 4* si riassumeranno i vari passi che hanno portato allo sviluppo dei due algoritmi, basati su reti neurali, proposti. Saranno presentate le diverse architetture testate e verranno descritti la procedura utilizzata per ottimizzarne la struttura e gli indici valutati per decidere quale fosse la rete migliore.

Nel *capitolo 5* verranno dettagliatamente descritte le due reti neurali scelte per effettuare la predizione. Ne saranno presentati i pregi e i difetti e saranno accennate alcune possibili modifiche che potrebbero migliorarne le performance.

Nel *capitolo 6* e nel *capitolo 7* saranno riassunti i risultati ottenuti, rispettivamente su un dataset simulato e su un dataset reale. Il profilo predetto dalle reti neurali sarà valutato tenendo conto di vari indici, in modo da avere un quadro complessivo il più ampio possibile delle potenzialità e dei limiti degli algoritmi sviluppati. Tutti i risultati saranno confrontati con quelli forniti da un modello lineare tempo variante e da un'altra rete neurale, entrambi recentemente proposti in letteratura.

Nel *capitolo 8* verranno tracciate le conclusioni del lavoro svolto e saranno commentati brevemente, da un punto di vista globale, i risultati ottenuti. Infine saranno presentati alcuni dei molti possibili sviluppi futuri.

Capitolo 1

Il diabete e il monitoraggio continuo della glicemia

Il diabete mellito è una patologia caratterizzata dall'incapacità del pancreas di produrre l'insulina (diabete di tipo 1), o dal malfunzionamento della secrezione e dell'azione insulinica (diabete di tipo 2), con un conseguente aumento della concentrazione di glucosio nel sangue (iperglicemia) e disturbi del metabolismo di carboidrati, grassi e proteine. La terapia convenzionale, che consiste in iniezioni sub-cutanee di insulina, una corretta dieta ed esercizio fisico, mira al mantenimento della glicemia il più possibile entro il range euglicemico ([70-180 mg/dl]), con lo scopo di prevenire e limitare le complicanze micro- e macrovascolari legate al diabete. La somministrazione di insulina, indispensabile nel diabete di tipo 1, e a cui talvolta devono ricorrere anche i soggetti affetti da diabete di tipo 2, è particolarmente critica, in quanto livelli troppo bassi di questo ormone portano a iperglicemia cronica, mentre livelli troppo alti possono causare ipoglicemia (concentrazione di glucosio ematico inferiore ai 70 mg/dl), particolarmente rischiosa a breve termine, poichè può degenerare nel coma e nella morte.

Il numero attuale di diabetici nel mondo supera i 220 milioni di individui e si prevede raggiungerà i 366 milioni nel 2030 [62]. Questo dato è probabilmente una sottostima del fenomeno, poichè almeno metà dei pazienti affetti da diabete mellito sono inconsapevoli della loro patologia. Per quanto riguarda l'Italia la prevalenza del diabete è aumentata dal 3.9% nel 2001 al 4.8% nel 2010 [54], (pari a circa 2,900,000 persone).

Il diabete è quindi considerato un problema socio-economico, non solo per la sua diffusione, ma soprattutto per la sua relazione con complicanze altamente invalidanti. I pazienti diabetici presentano infatti, rispetto alla popolazione generale, una mortalità doppia, per l'accresciuto rischio di malattie cardiovascolari, che comporta un aumento del numero di infarti del miocardio, di ictus cerebrali, di ulcere e amputazioni degli arti inferiori. Oltre ad interessare il sistema cardiovascolare il diabete altera anche i vasi

di piccolo calibro, danneggia inoltre la retina, i reni ed i nervi. Il diabete è tuttora la causa principale di cecità nei soggetti in età lavorativa, ed è tra le principali cause di insufficienza renale grave. I danni a carico del sistema nervoso, noti come neuropatia diabetica, risultano invalidanti per il soggetto, in quanto causano sintomatologie dolorose negli arti inferiori, ipotensione ortostatica e disturbi digestivi. Tutte queste complicanze impongono, per la loro cura e diagnosi, un impegno finanziario tale da porre il diabete tra le prime voci di spesa del sistema sanitario (in Italia il 6.7% dell'intera spesa sanitaria, pubblica e privata, pari a circa 5.5 miliardi di euro nel 2004 [54]).

Per tutti questi motivi è sempre maggiore l'attenzione che il diabete suscita a livello sociale, medico, farmacologico e sperimentale e la sua cura ha come obiettivo non solo il miglioramento dello stato di salute attuale del paziente, ma soprattutto la prevenzione delle complicanze d'organo [8].

1.1 Fisiopatologia del diabete mellito

Il glucosio è la principale fonte di energia per l'organismo umano e la sua concentrazione nella circolazione sanguigna viene mantenuta entro limiti molto ristretti, grazie ad un sottile equilibrio dinamico tra produzione e utilizzazione di questo substrato.

Normalmente esso proviene dall'assorbimento dei carboidrati lungo il tubo gastrointestinale, mentre, nello stato di digiuno non prolungato, il rifornimento di glucosio per l'organismo è assicurato dal fegato e, in misura minore, dal rene.

I vari organi e tessuti possono essere classificati, in base alle caratteristiche funzionali del loro metabolismo del glucosio, come:

- *Glucosio dipendenti*: il cervello e i globuli rossi. Il glucosio rappresenta l'unica fonte di energia e viene utilizzato in maniera costante, indipendentemente dalla concentrazione di insulina. Quando la concentrazione di glucosio nel sangue scende al di sotto dei 65 mg/dl, i neuroni non riescono a captare una quantità di glucosio sufficiente alle loro esigenze metaboliche (pari, in un individuo adulto, a circa 100 mg/min), e si instaura la neuroglicopenia.
- *Insulino dipendenti*: tessuto muscolare, tessuto adiposo e fegato. Questi organi utilizzano il glucosio in modo fasico, assorbendone in proporzione alla quantità di insulina circolante all'interno dell'organismo. L'insulina è un ormone secreto dalle β -cellule del pancreas endocrino e rilasciato in notevole quantità ogni volta che ci nutriamo. Essa ha un effetto anabolico, in quanto predispone l'organismo a immagazzinare i nutrienti e a sintetizzare nuova materia vivente, ed è ipoglicemizzante, poichè riduce la produzione endogena di glucosio nel fegato e stimola l'utilizzazione di glucosio negli organi insulino dipendenti.

- *Gluco-sensori*: pancreas endocrino, fegato e ipotalamo ventro-mediale. Questi organi percepiscono la concentrazione di glucosio circolante ed organizzano un'appropriata risposta secretoria. Le β -cellule pancreatiche secernono insulina in risposta all'aumento della glicemia, il fegato produce glucosio in modo tale da mantenere costante la concentrazione di glucosio stesso, mentre l'ipotalamo ventro-mediale percepisce biochimicamente l'ipoglicemia e organizza la risposta ormonale controregolatoria, secernendo ormoni antagonizzanti l'insulina, con effetto iperglicemizzante.
- *Glucovori*: intestino e rene. Mediano l'ingresso (intestino) o il rientro (rene) di glucosio all'interno dell'organismo.

Nei soggetti sani la concentrazione ematica di glucosio è finemente monitorata e regolata dall'organismo, mentre nelle persone affette da diabete è presente un difetto assoluto o relativo nella produzione di insulina da parte del pancreas, o si riscontra una risposta non adeguata delle cellule all'insulina (insulinoresistenza).

1.2 Diagnosi e classificazione del diabete

Nel 1997, dopo oltre un anno di collaborazione tra esperti dell'ADA (American Diabetes Association) e dell'OMS (Organizzazione Mondiale della Sanità), è stato rilasciato un comunicato contenente le nuove raccomandazioni per la diagnosi e la classificazione del diabete mellito.

1.2.1 Criteri diagnostici del diabete

Le tre modalità diagnostiche per il diabete sono:

1. Riscontro casuale (in qualsiasi momento della giornata) di glicemia maggiore o uguale a 200 mg/dl e sintomi di diabete (poliuria, polidipsia e perdita di peso non altrimenti giustificata).
2. Glicemia a digiuno (dopo almeno 8 ore dall'ultimo apporto calorico) maggiore o uguale a 126 mg/dl.
3. Glicemia plasmatica due ore dopo un carico orale di glucosio di 75 g maggiore o uguale a 200 mg/dl.

Un ulteriore marker diagnostico, utilizzato soprattutto per valutare retrospettivamente il controllo della glicemia e per aggiustare la terapia insulinica è la percentuale di emoglobina glicosilata HbA_{1c} , un indice legato all'iperglicemia nelle ultime 5 settimane [58].

1.2.2 Classificazione del diabete

In base all'eziopatogenesi dell'iperglicemia si distinguono:

Diabete di tipo 1

È caratterizzato dalla distruzione delle β -cellule e da una conseguente insulino-deficienza solitamente assoluta. Ciò può avvenire a causa di un processo autoimmune (diabete mellito immuno-mediato), o per altri motivi non totalmente noti (diabete mellito idiopatico).

Questa forma di diabete interessa principalmente soggetti in età infantile o giovani adulti, normopeso, e i sintomi si manifestano in un breve arco di tempo, quando oltre l'80% delle β -cellule è già stato distrutto. La carenza di insulina porta la glicemia a superare i 180 mg/dl, il glucosio inizia quindi ad essere perso dal rene, assieme a crescenti quantità di acqua. Si nota un aumento della diuresi (poliuria) e della sete (polidipsia). In assenza di insulina si ha anche un'incontrollata lipolisi e un'ossidazione massiva di acidi grassi nel fegato, con produzione di corpi chetonici che si accumulano nel sangue, determinando acidosi. Questo sconvolgimento metabolico ha come segno visibile il calo ponderale, dovuto all'intaccamento delle riserve energetiche proteiche, per l'incapacità dell'organismo di utilizzare il glucosio e gli acidi grassi a fini energetici e per il venir meno dell'effetto anabolico dell'insulina. Se il diabete non viene prontamente diagnosticato e curato, somministrando insulina esogena, il paziente può andare incontro a coma chetoacidotico e a morte [8].

Diabete di tipo 2

È caratterizzato da insulino-resistenza, insulino-deficienza relativa e eccessiva produzione epatica di glucosio.

Questa forma di diabete si sviluppa soprattutto in età adulta e nella popolazione anziana, e si associa spesso a sovrappeso o obesità. I soggetti affetti da diabete di tipo 2 presentano *resistenza insulinica*, che causa iperglicemia post-prandiale, dovuta alla scarsa utilizzazione di glucosio da parte dei tessuti insulino-sensibili, e iperglicemia a digiuno, causata dall'aumento della produzione epatica di glucosio. Nelle prime fasi della patologia la resistenza insulinica viene compensata con un aumento del rilascio dell'ormone stesso. A lungo andare, le isole pancreatiche si stressano e, non essendo più in grado di sostenere lo stato iperinsulinemico, diminuiscono la secrezione di insulina. Questo fenomeno, assieme all'aumento della produzione epatica di glucosio, conduce al diabete.

Diabete mellito gestazionale

Viene definito come qualsiasi grado di intolleranza al glucosio insorto o individuato per la prima volta durante la gravidanza. Può richiedere di essere trattato con insulina o semplicemente con la dieta e l'esercizio fisico, e la condizione può persistere o meno dopo la gravidanza. Ciò non esclude la possibilità che un'intolleranza al glucosio possa essere retrodatata o fatta iniziare in concomitanza alla gravidanza [58], [8].

1.3 Effetti del diabete

Gli effetti e le complicanze del diabete sono legate principalmente a danni provocati dalle crisi iper e ipoglicemiche, a cui i soggetti affetti da questa patologia vanno spesso incontro. Tra essi si possono distinguere sintomi legati direttamente all'aumento della concentrazione plasmatica di glucosio (polidipsia e poliuria), e sintomi dovuti a complicanze a lungo termine (microangiopatia, neuropatia, aterosclerosi e infezioni cutanee e urinarie).

1.3.1 Complicanze a breve termine

La principale complicanza a breve termine del diabete è il coma, dovuto sia a iper che ipoglicemia.

- *Coma ipoglicemico*: l'ipoglicemia può manifestarsi a causa della somministrazione di una dose di insulina eccessiva, rispetto alla produzione endogena e all'apporto esogeno di glucosio, ma anche per esercizio fisico troppo intenso. Quando la glicemia scende a valori compresi tra 50 e 70 mg/dl il sistema nervoso centrale, che trae tutta la sua energia dal metabolismo del glucosio, diventa più eccitabile, un'ulteriore discesa glicemica a valori compresi tra 20 e 50 mg/dl causa, in genere, convulsioni e possibile perdita di coscienza, mentre al di sotto della soglia di 20 mg/dl il paziente cade in uno stato di coma [15].
- *Coma iperglicemico*: in carenza di insulina, nonostante l'alta concentrazione di glucosio ematico, l'organismo non riesce ad utilizzarlo e si verifica un passaggio dal metabolismo dei carboidrati a quello dei grassi, di conseguenza aumenta la concentrazione plasmatica di chetoacidi, che vengono assunti e ossidati dalle cellule. Si instaura una forte acidosi metabolica, aggravata dalla disidratazione dovuta all'eccessiva formazione di urina. Il paziente va così incontro a coma e a morte, se non viene immediatamente trattato con alte dosi di insulina [18].

1.3.2 Complicanze a lungo termine

La patogenesi delle complicanze microvascolari e neuropatiche a lungo termine è una delle più importanti questioni irrisolte riguardanti il diabete mellito. L'*ipotesi "metabolica"* le correla all'iperglicemia e ad altre anomalie metaboliche, dovute al deficit di insulina, mentre l'*ipotesi "genetica"* sostiene siano puramente dovute a fattori genetici. I dati attualmente a disposizione sono a netto favore della teoria metabolica, ma manca ancora una prova definitiva [15].

L'iperglicemia cronica porta ad alterazioni strutturali dei tessuti. Per spiegare questi effetti sono stati proposti due meccanismi:

- Ad alte concentrazioni il glucosio può essere ridotto a sorbitolo. Questa sostanza non può essere metabolizzata da tessuti come le lenti oculari e i nervi, nei quali si accumula creando danni osmotici. Tale meccanismo potrebbe essere alla base della *neuropatia periferica*, in cui è alterato il funzionamento dei nervi periferici, e di varie disfunzioni del sistema nervoso vegetativo.
- Il glucosio può anche reagire non enzimaticamente con gruppi liberi di aminoacidi proteici, glicando le proteine. Danneggiamenti al collagene potrebbero spiegare le *artriti*, di cui molti diabetici soffrono, e l'ispessimento delle membrane basali, associato al danneggiamento dei piccoli vasi sanguigni. La *retinopatia diabetica* è dovuta al danneggiamento dei capillari della retina ed è la principale causa di cecità nei soggetti in età lavorativa. Alla base dell'accresciuto rischio di *aterosclerosi* e di *ischemia cardiaca*, che si osserva nei pazienti diabetici, è probabile ci sia la glicazione dell'apolipoproteina B, mentre il danneggiamento dell' α -cristallino nella lente oculare può essere messo in correlazione con l'alta prevalenza della *cataratta* [5].

Un'ulteriore complicanza dell'iperglicemia è la *nefropatia diabetica*, che causa insufficienza renale anche grave, fino a rendere indispensabile la dializzazione del paziente o il trapianto di rene.

L'ipoglicemia, sebbene causi principalmente danni acuti, ha anche effetti a lungo termine, in quanto crisi ipoglicemiche ripetute, seppur di breve durata, possono provocare un sottile deterioramento cerebrale, con riduzione dell'intelligenza e disritmie cerebrali [15].

1.4 Il monitoraggio della glicemia

La terapia diabetica ha come scopo il mantenimento della concentrazione di glucosio nel sangue all'interno del range euglicemico ([70-180 mg/dl]), con l'obiettivo di minimizzare il rischio di complicanze d'organo legate al diabete. Fino agli anni '50 la terapia si basava esclusivamente sull'equilibrio

tra dieta ed esercizio fisico e sulla somministrazione di insulina, senza la conoscenza del livello glicemico preciso. A partire dagli anni '80 lo sviluppo di nuove tecnologie e dispositivi per l'automonitoraggio della glicemia (Self Monitoring Blood Glucose, SMBG), ha notevolmente migliorato la terapia.



Figura 1.1: Dispositivo pungi-dito per SMBG.

Questi dispositivi sono costituiti da un pungi-dito, che consente il prelievo di una goccia di sangue capillare da un polpastrello, e da un misuratore, che fornisce il valore glicemico del campione di sangue prelevato, generalmente in mg/dl (Figura 1.1). Il monitoraggio della glicemia avviene, in genere, 3-4 volte al giorno e le somministrazioni di insulina o di farmaci ipoglicemizzanti sono regolate in base alle misurazioni fornite dal dispositivo. Questi strumenti hanno indubbiamente permesso notevoli miglioramenti nel trattamento del diabete, ma presentano ancora delle limitazioni, connesse sia alla loro invasività (è necessario prelevare campioni di sangue più volte al giorno), sia alla loro incapacità di fornire informazioni sul trend glicemico e su tutti gli episodi di ipo e iperglicemia che si hanno nell'arco della giornata. Un chiaro esempio

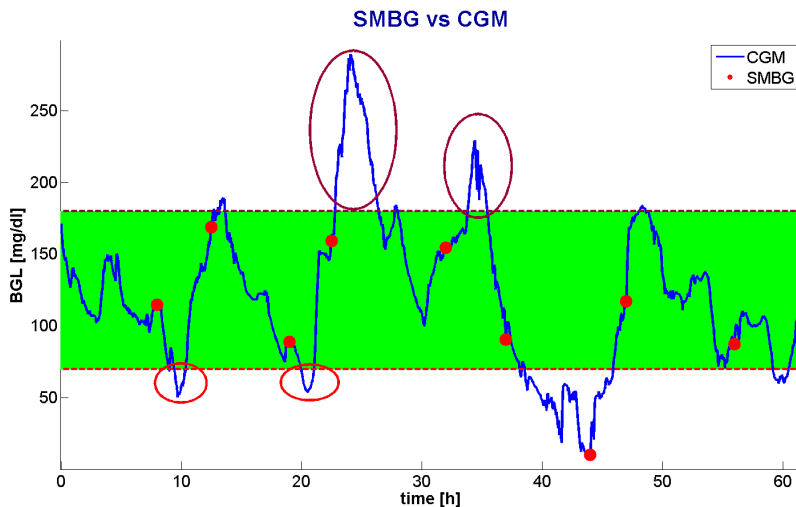


Figura 1.2: Confronto tra misure CGM (profilo blu) e misure SMBG (pallini rossi). L'area verde indica il range euglicemico.

dell'insufficienza delle poche misurazioni ottenute con riflettometro nell'identificazione di tutti gli episodi ipo e iperglicemici è fornito in Figura 1.2. I pallini rossi corrispondono alle 4 misure giornaliere ottenute con dispositivi per il SMBG, mentre la linea blu è il profilo glicemico reale. Risultano evidenti le due escursioni iperglicemiche, cerchiare in marrone, e le due crisi ipoglicemiche lievi, cerchiare in rosso, non individuate dalle misure

SMBG. Fino alla fine del XX secolo l'unico modo per ottenere una misura frequente della glicemia era il campionamento in vena, presso i laboratori clinici. Questa tecnica, tutt'ora utilizzata in ambito ospedaliero, è in grado di fornire una misura glicemica ogni 15 min, ma, data l'elevata invasività, non è destinata al monitoraggio vero e proprio, bensì ad un aggiustamento retrospettivo della terapia, grazie alle informazioni raccolte, e a scopi di ricerca. L'obiettivo di superare questo limite ha portato, a partire dal 2000, allo sviluppo di nuove tecnologie e strumenti per un monitoraggio quasi continuo e minimamente invasivo della glicemia (il cosiddetto Continuous Glucose Monitoring, CGM), mediante apparecchi portatili che il paziente può "indossare" nella vita quotidiana per diversi giorni.

1.5 I sensori CGM

I sensori per il monitoraggio continuo della glicemia sono strumenti in grado di fornire misure molto frequenti della concentrazione glicemica (ogni 1-10 min) per diversi giorni consecutivi (fino a 7 giorni). Sono evidenti i potenziali benefici che si possono avere nella terapia del diabete utilizzando questi dispositivi. Le frequenti misurazioni permettono di identificare eventi rischiosi in tempo reale, e danno indicazioni sul trend glicemico, che possono essere sfruttate per predire il livello di glucosio nell'immediato futuro. I molti dati raccolti, inoltre, possono essere analizzati anche retrospettivamente per aggiustare la terapia insulinica.

Dal punto di vista tecnologico i dispositivi CGM possono essere classificati come minimamente invasivi e non invasivi.

1.5.1 Dispositivi minimamente invasivi

Il sensore misura la concentrazione di glucosio non direttamente nel sangue, ma nel fluido interstiziale, attraverso il quale le sostanze nutritive provenienti dai vasi sanguigni diffondono prima di essere assorbite dalle cellule. Questi dispositivi sono composti da:

- un sensore impiantato nel sottocute (generalmente a livello addominale o nel braccio), e un trasmettitore indossabile, che fornisce misurazioni in tempo reale;
- un dispositivo tascabile costituito da un ricevitore, una memoria per salvare la grande mole di dati e un display per visualizzare le misure provenienti dal sensore.

Il sensore misura generalmente una corrente (in mA o mV), prodotta da reazioni chimiche e proporzionale alla concentrazione di glucosio nel sito

d'impianto del sensore. La conversione tra questo segnale grezzo e la concentrazione ematica di glucosio avviene attraverso un processo di calibrazione, che richiede alcune misure SMBG.

Tra i più importanti ed utilizzati dispositivi minimamente invasivi citiamo:

- CGMS[®] System Gold[™] (Medtronic MiniMed, northridge, CA), che ha ricevuto l'approvazione della FDA a Marzo 2001 [55];
- Guardian[®] Real-Time (Medtronic MiniMed, northridge, CA), che ha ricevuto l'approvazione della FDA a Giugno 2005 [57];
- GlucoDay[®] (Menarini Diagnostics, Firenze, Italia), che ha ricevuto il marchio CE in Europa (non è stato richiesto il marchio FDA) [56];
- STS-7[®] (Dexcom, San Diego, CA), che ha ricevuto l'approvazione della FDA a Giugno 2007 [52];
- FreeStyle Navigator[™] (Abbott Laboratories, Alameda, CA), che ha ricevuto l'approvazione della FDA a Marzo 2008 [51].

Il dispositivo FreeStyle Navigator[™], da cui proviene la totalità dei dati reali utilizzati nella presente tesi sarà descritto più dettagliatamente nella Sezione 1.5.3.

1.5.2 Dispositivi non invasivi

I dispositivi CGM non invasivi sono tutt'ora in fase di sviluppo e prima della loro messa in commercio dovranno trascorrere ancora alcuni anni, necessari per perfezionarli e per migliorare la qualità della misura fornita. Questi dispositivi, a differenza di quelli minimamente invasivi, che misurano il glucosio nel fluido interstiziale, sfruttano metodi di misura transcutanei, come l'irradiazione del tessuto e l'analisi delle radiazioni assorbite e di scatter, e metodi ottici, quali le radiazioni infrarosse, la spettroscopia di Raman, la polarimetria, la luce scatter e la spettroscopia fotoacustica.

Di seguito è riportata una breve rassegna dei più interessanti dispositivi non invasivi e del relativo principio di funzionamento.

- GlucoWatch G2 Biographer (Cygnus, Inc., Redwood City, CA), approvato dalla FDA nel 2001 [41]. Questo dispositivo misura la corrente di glucosio attraverso una membrana: sfruttando una carica elettrica causa l'espulsione di sodio e cloro all'esterno della pelle del paziente. Durante questo processo anche il glucosio viene espulso passivamente, con l'acqua di idratazione dei sali. La soluzione così estratta viene ossidata e se ne misura il contenuto di glucosio. Questo prodotto è stato ritirato dal mercato nel 2006, per i numerosi casi di irritazione cutanei causati dal suo utilizzo.

- Aprise (Glucon Inc., Boulder, CO): il dispositivo si basa su un metodo fotoacustico [45]. In corrispondenza ad un vaso sanguigno viene applicata alla pelle una luce laser che provoca un piccolo e rapido incremento della temperatura del vaso stesso e produce un leggero suono. Il dispositivo percepisce questo suono e, in base alle sue caratteristiche acustiche, determina il livello di glucosio nel sangue. Questo sensore misura il glucosio direttamente nel sangue, e non nell'interstizio, come fanno invece gli altri. L'Aprise non è ancora stato approvato dalla FDA.
- NMB (OrSende Ltd., Nes Ziona, Israele): questo dispositivo si basa sulla spettroscopia di occlusione [2]. La misurazione è effettuata attraverso una sonda posizionata sul dito, 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 ultrasistolica tale da occludere il vaso. Questa tecnologia sfrutta direttamente gli effetti del glucosio sulle proprietà di scattering dell'organo: il glucosio abbassa il coefficiente di scattering del tessuto, di conseguenza un minor numero di fotoni viene assorbito ed aumenta l'intensità della luce. Il dispositivo è ancora un prototipo e non è stata richiesta né l'approvazione della CE, né l'approvazione della FDA.
- Solianis Multisensor (Solianis Monitoring AG, Zürich, CH): il dispositivo sfrutta la spettroscopia d'impedenza (IS) per rilevare le variazioni della concentrazione di glucosio [59]. 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. Il sensore è ancora in fase di sviluppo e non è stata richiesta l'approvazione della CE.

1.5.3 Abbott FreeStyle Navigator™

Il FreeStyle Navigator™ prodotto da Abbott Diabetes Care è un sensore elettrochimico sottocutaneo, che può essere indossato per un massimo di 5 giorni e che è in grado di fornire una misurazione di glucosio al minuto. Il dispositivo è composto da:

- un *sensore*, Figura 1.3, posizionato sulla parte superiore del braccio o sull'addome e fissato con uno speciale adesivo. Un sottile filamento lungo 5 mm penetra nel sottocute e misura il livello di glucosio nel fluido interstiziale.
- un *trasmettitore*, Figura 1.4, che è collegato al sensore e che invia le misurazioni di glucosio al ricevitore wireless, fino ad una distanza di 3 m.

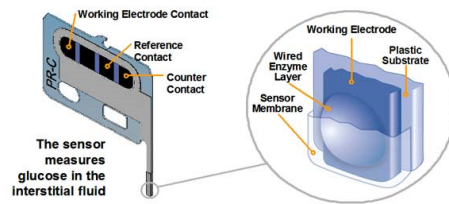


Figura 1.3: FreeStyle NavigatorTM, sensore [51].



Figura 1.4: FreeStyle NavigatorTM, trasmettitore [51].

- un ricevitore, Figura 1.5, paragonabile a un piccolo computer. Memorizza tutti i valori di glucosio, per un periodo che può durare fino a 60 giorni, e fornisce informazioni accurate sull'andamento della glicemia in tempo reale. Può essere programmato per predire iper- e ipoglicemie, in base a soglie settate dal paziente, e genera allarmi quando i valori di glucosio si avvicinano alle soglie stesse. Al ricevitore è integrato anche un riflettometro, in modo che il paziente non necessiti di un ulteriore dispositivo per misurare la glicemia ematica e rendere così più agevole la procedura di calibrazione.

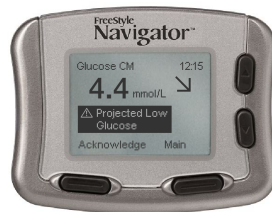
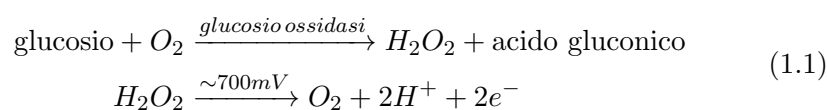


Figura 1.5: FreeStyle NavigatorTM, ricevitore [51].

La misurazione del glucosio sfrutta la tecnologia Wired-EnzymeTM, simile al processo di misura basato sulla glucosio-ossidasi, ma meno dipendente dall'ossigeno e più precisa. La misurazione basata sulla glucosio-ossidasi sfrutta la generazione di perossido idrogeno, a partire da glucosio e ossigeno, grazie all'enzima glucosio-ossidasi:



Un amperometro misura la corrente generata dall'ossidazione del perossido idrogeno, nel sito dell'elettrodo. Il sensore, sfruttando questa reazione chimica, misura, però, non tanto la concentrazione del glucosio, ma principalmente quella dell'ossigeno. Affinchè la reazione avvenga è necessario sia presente, per ogni molecola di glucosio una molecola di ossigeno, ma il glucosio è presente all'interno dell'organismo in quantità molto superiori a quelle dell'ossigeno, che diventa, di conseguenza, un reagente limitante. Il FreeStyle NavigatorTM risolve questo problema attraverso la tecnologia Wired-EnzymeTM che, sfruttando un mediatore all'osmio (il Wired-Enzyme), riduce la dipendenza della reazione dall'ossigeno e ne evita la saturazione. In questo modo si ottengono una maggiore stabilità del sensore, una ridotta suscettibilità alle variazioni in vivo delle concentrazioni di ossigeno e la minimizzazione della risposta alle comuni interferenze elettroattive [13].

Il segnale elettrico prodotto dalla reazione chimica è trasformato nel corrispondente livello glicemico attraverso una procedura di calibrazione, basata su misure SMBG registrate dopo 10, 12, 24 e 72 ore dall'inserimento del sensore [14].

Per testare la bontà delle performance del FreeStyle NavigatorTM le misure prelevate dal sensore sono state confrontate sia con misurazioni ematiche [26], sia con misure ottenute con altri sensori CGM [7], [25]. Questi studi hanno dimostrato un'alta accuratezza delle letture del FreeStyle NavigatorTM, (98.4% di letture corrette, usando i criteri della FDA [7]), specialmente nelle zone di ipoglicemia.

1.5.4 Punti di forza e problemi dei dispositivi CGM

Come già accennato in precedenza, uno dei grandi vantaggi dei sensori CGM è la loro capacità di fornire misurazioni complete e quasi continue della glicemia nell'arco della giornata. Le informazioni in tempo reale sul livello glicemico e sul trend che esso sta seguendo permettono al paziente modifiche tempestive della dieta e della terapia insulinica, mirate al mantenimento dell'euglicemia. Inoltre molti dispositivi sono in grado di generare allarmi quando la glicemia oltrepassa le soglie critiche e alcuni, tra i quali anche il FreeStyle NavigatorTM, sfruttano i molti dati a disposizione per cercare di predire possibili ipo e iperglicemie future e generare un'allerta con un anticipo di 20-30 min.

La generazione di allarmi, unita alla portabilità di questi dispositivi, rende i pazienti diabetici molto più autosufficienti nella gestione della loro terapia e degli eventi critici. Bisogna però sottolineare come la grande mole di informazioni che il paziente si trova a gestire renda necessario, da parte del paziente stesso, una maggior conoscenza degli aspetti fisiologici, legati alla sua malattia, e degli aspetti tecnologici, legati al sensore. È inoltre opportuno sottolineare che nessun sensore CGM è stato approvato per la gestione autonoma della terapia, è sempre necessario, infatti, effettuare anche un

test SMBG prima di prendere qualsiasi decisione suggerita dai dati CGM, principalmente a causa della scarsa accuratezza delle letture del sensore. Infine il costo dei dispositivi non può essere ignorato. Per esempio, nel caso del FreeStyle Navigator™ il kit contenente il trasmettitore, le batterie, le lancette e una confezione di strisce per il riflettometro incorporato al trasmettitore costa approssimativamente \$960-\$1040, mentre il prezzo del sensore, che può essere indossato per un massimo di 5 giorni consecutivi, è di circa \$60.

1.5.5 Aspetti critici dei sensori CGM

Uno degli aspetti più critici legati all'utilizzo dei sensori CGM è indubbiamente la loro scarsa accuratezza. Ci sono inoltre problemi legati alla calibrazione dei dispositivi e al filtraggio dei dati da essi raccolti, nonché alla possibile generazione di allarmi in corrispondenza all'attraversamento, reale o predetto, di soglie critiche.

Aspetti fisiologici

Per ridurre l'invasività dei dispositivi CGM, il segnale misurato è proporzionale alla concentrazione di glucosio nell'interstizio, e non nel plasma. Poiché il metabolismo del glucosio è regolato in base alla sua concentrazione plasmatica, è necessario descrivere la dinamica tra plasma e interstizio. Il modello più popolare, tra i vari proposti in letteratura, è il modello bicompartimentale di Rebrin *et al.* [35], la cui funzione di trasferimento è paragonabile a quella di un filtro passa-basso, ed è caratterizzata dal gradiente di diffusione da plasma a interstizio e dalla costante di tempo di diffusione. L'identificazione del modello su alcuni dataset ha dimostrato che, mentre il gradiente di diffusione è pressochè costante, la costante di tempo risulta tempo-variante e il ritardo introdotto dalla diffusione non è trascurabile. La Figura 1.6 illustra chiaramente la distorsione del segnale di glucosio nell'interstizio (linea blu), rispetto al glucosio plasmatico (asterischi rossi). Tale distorsione consiste sia in uno scostamento orizzontale tra le due misurazioni, dovuto al ritardo fisiologico con cui il glucosio diffonde dal plasma, attraverso la parete capillare, fino all'interstizio, sia in uno scostamento verticale del segnale misurato rispetto a quello reale, dovuto principalmente al deterioramento del sensore.

Calibrazione

La maggior parte dei sensori CGM misura una corrente, che deve essere convertita nel corrispondente livello di glucosio usando, come riferimento, misure SMBG. Il segnale misurato, però, non è direttamente proporzionale al glucosio ematico, a causa del ritardo introdotto dalla dinamica plasma-interstizio precedentemente discussa. In particolare, durante i cambiamenti

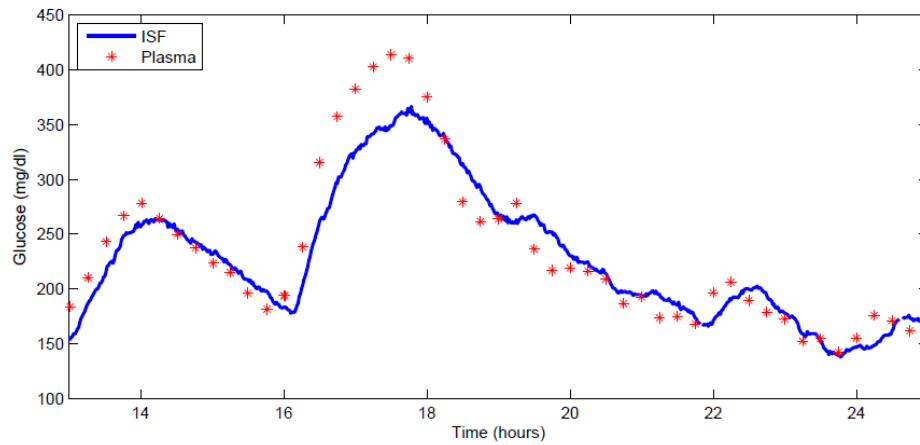


Figura 1.6: Confronto tra misurazioni di glucosio plasmatiche e misurazioni di glucosio interstiziale ottenute con sensore CGM.

rapidi della glicemia, la corrispondenza puntuale tra glucosio plasmatico e interstiziale può essere significativamente diversa e una calibrazione effettuata in questi momenti risulterebbe inutile, poichè poco veritiera. Inoltre la calibrazione deve essere effettuata più volte, dopo l’inserimento del sensore, (per il FreeStyle Navigator™ dopo 10, 12, 24 e 72 ore), a causa del deterioramento di alcune parti del sensore stesso a contatto con i liquidi biologici. Il problema della calibrazione è stato solo parzialmente investigato in letteratura ed è tutt’ora oggetto di studio.

Filtraggio

I dati registrati dai sensori CGM sono inevitabilmente corrotti da rumore di misura, quindi, per poterli utilizzare al fine di generare allarmi e predire livelli di glucosio futuro è necessario applicarvi un filtraggio causale. La Figura 1.7 riporta due profili registrati entrambi con sensore FreeStyle Navigator™, ma relativi a due soggetti differenti. Nella serie rappresentata nel pannello in basso il rumore è visibilmente maggiore che nella serie presente nel pannello in alto (variabilità interindividuale). Inoltre, nella seconda serie temporale, si può notare come il rumore si accentui nell’intervallo [700, 900] min (variabilità intraindividuale).

Ciò rende necessario lo sviluppo di algoritmi di filtraggio ottimi per migliorare il rapporto segnale rumore dei dati registrati dal sensore CGM.

Predizione

Un’applicazione molto interessante dei sensori CGM è la prevenzione delle ipo e iperglicemie, cercando di predirle con un opportuno anticipo. Il segnale glicemico è predicibile [6], ma, per minimizzare l’errore di predizione

1.6 Applicazione di sensori CGM per la gestione on-line della terapia: il progetto DIAAdvisor

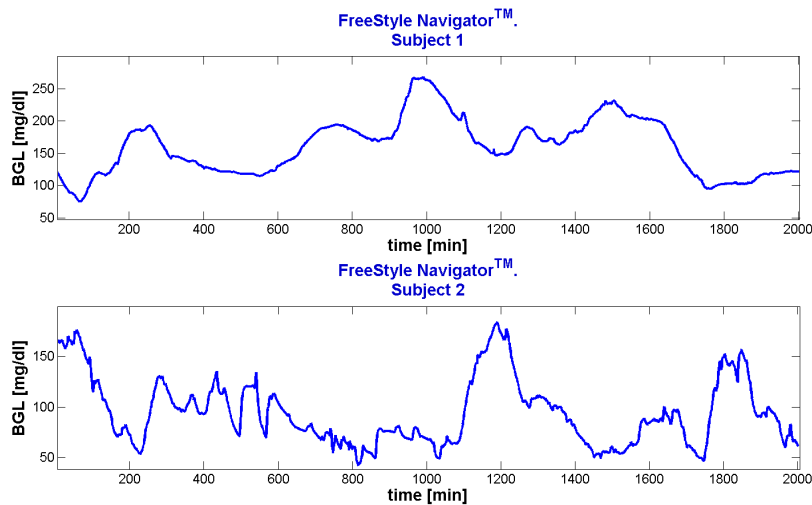


Figura 1.7: Serie temporali misurate con sensore CGM FreeStyle Navigator™. È evidente la variabilità inter- e intra- personale del rumore di misura.

e generare pochi falsi allarmi i dati devono essere accurati. Di conseguenza predizione e filtraggio risultano strettamente collegate.

Poichè l'obiettivo di questa tesi è proprio la predizione del segnale glicemico, nel capitolo 3 verranno indagati alcuni metodi utilizzati in letteratura per predire il livello di glucosio ematico. Nei capitoli successivi verranno illustrati due nuovi predittori, basati sull'utilizzo di reti neurali, e ne verranno valutate le performance su dati simulati e su dati reali.

1.6 Applicazione di sensori CGM per la gestione on-line della terapia: il progetto DIAAdvisor

1.6.1 Aspetti generali del progetto

DIAAdvisor™ [53] è un progetto finanziato dalla UE nell'ambito del 7° progetto quadro, è partito nel 2008 e ha una durata di 4 anni. Lo scopo di tale progetto è lo sviluppo di un dispositivo che utilizzi dati e informazioni passati, su valori glicemici, pasti, dosi di insulina e parametri fisiologici, per aiutare i diabetici di tipo 1 e di tipo 2 ad ottimizzare la loro terapia. In particolare sfrutta anche un sistema di predizione a breve termine della glicemia e di guida al trattamento della malattia. Il dispositivo dovrebbe accorciare l'esposizione temporale alle ipo e iperglicemie, diminuendo le complicanze a lungo termine del diabete. La predizione è basata sulle misurazioni di glucosio, le dosi di insulina somministrata e parametri specifici del paziente e, oltre a fornire indicazioni sul livello glicemico nell'imminente futuro, lo strumento è anche in grado di dare suggerimenti su come modificare e

adattare la terapia insulinica, per ottenere un controllo più stabile della patologia [53].

La prima parte del progetto, conclusasi a fine 2009, consisteva nella raccolta di una ricca serie di informazioni e dati bioclinici, potenzialmente o direttamente coinvolti nel controllo glicemico del paziente diabetico. Successivamente, sulla base di questi dati, sono stati costruiti modelli del sistema glucosio-insulina, per identificare le relazioni tra glicemia e insulina e mettere a punto algoritmi di controllo. La grande quantità di dati raccolti contribuisce ad aumentare le conoscenze sul controllo glicemico nei pazienti diabetici e sulle correlazioni tra concentrazione di glucosio e i vari parametri misurati.

1.6.2 Descrizione dei dati

Tutti i dati utilizzati nel presente elaborato sono dati raccolti nella prima fase del progetto DIAadvisor™.

Nello studio sono stati arruolati 30 pazienti¹, 24 con diabete di tipo 1 e 6 con diabete di tipo 2, equamente divisi, all'interno di ogni gruppo, tra maschi e femmine. I pazienti avevano un'età compresa tra 18 e 70 anni, il diabete era stato loro diagnosticato da almeno un anno e risultavano in cura con i medici coinvolti nello studio da almeno 3 mesi. Era inoltre richiesto che lo schema di trattamento insulinico dei pazienti fosse del tipo basal-bolus, ovvero composto da iniezioni multiple di insulina o da somministrazione continua con microinfusore.

Il monitoraggio in continua della glicemia è stato eseguito con l'Abbott FreeStyle Navigator™, le glicemie plasmatiche sono state misurate con il sensore HemoCue Glucose 201+ Analyzer ed è stato applicato anche il sensore non invasivo della Solianis Monitoring AG. Ai pazienti è stato inoltre chiesto di indossare la maglietta Clinical LifeShirt® della VivoMetrics, disegnata appositamente per trials clinici, in grado di registrare ed analizzare in continuo i segni vitali (ovvero la funzione respiratoria, l'ECG, le variazioni della postura, della posizione e dell'attività fisica, la pressione arteriosa, la saturazione dell'ossigeno, l'EEG, i movimenti degli arti inferiori, la temperatura corporea e della pelle, la produzione di CO₂ e la tosse). Per quanto riguarda le informazioni relative ai pasti, ai partecipanti è stato chiesto di fotografare con cellulare standard GSM i cibi e le bevande assunte, in tal modo, per ogni foto, viene aggiunto il dato relativo all'orario del pasto e viene calcolata la quantità di carboidrati ingeriti. Le dosi di insulina somministrate e i relativi orari sono state invece registrate manualmente dai pazienti. In aggiunta sono stati effettuati anche dei prelievi ematici per eseguire le analisi biologiche, come la determinazione della glicemia e la misurazione della concentrazione d'insulina.

¹Poichè per uno dei pazienti non era disponibile in realtà nessuna serie glicemica, sono stati effettivamente utilizzati i dati di 29 pazienti.

1.6.3 Tecniche per la predizione

Per poter suggerire la terapia insulinica più appropriata e cercare di prevenire le crisi ipo e iperglicemiche il dispositivo studiato nell'ambito del progetto DIAAdvisor deve incorporare un predittore che, in base alle misurazioni passate di glucosio e alle altre informazioni raccolte, sia in grado di indicare un valore attendibile della glicemia futura.

Attualmente vengono utilizzati con successo modelli lineari e modelli AR(1) [39], entrambi con parametri tempovarianti, determinati attraverso la tecnica ricorsiva dei minimi quadrati pesati. Queste strategie predittive, benchè semplici, sono in grado di individuare con un anticipo di 20-25 minuti gli attraversamenti delle soglie ipo e iperglicemiche. Questi modelli predicono però anche un elevato numero di falsi attraversamenti, dovuto principalmente al fatto che essi non possono incorporare al loro interno informazioni sui pasti ingeriti dal paziente e sulle dosi di insulina iniettata. Tali eventi perturbano in maniera significativa la glicemia, e i loro effetti non sono predicibili utilizzando esclusivamente misure di glucosio passate. Dato che il dataset contiene informazioni complete e dettagliate sui carboidrati ingeriti e sulla terapia insulinica di ogni paziente, sarebbe opportuno sviluppare un predittore che sappia sfruttare questi importanti dati.

In questa direzione si sta muovendo uno dei partner del progetto, che sta sviluppando un predittore che sfrutta la teoria relativa all'intelligenza artificiale e ai problemi inversi malposti per predire il profilo glicemico futuro (e non semplicemente il suo valore puntuale), dall'istante attuale fino all'orizzonte di predizione prestabilito [32]. Il modello proposto da Pereverzyev *et al.* è composto da due reti neurali, la prima, in base ai dati in ingresso, sceglie uno spazio funzionale tra un insieme di kernels disponibili, la seconda rete neurale ottimizza i parametri del kernel scelto, in modo da ottenere una funzione attraverso cui predire il profilo glicemico futuro. Quest'architettura, inizialmente sviluppata utilizzando esclusivamente misure glicemiche passate, registrate con sensori CGM, è molto flessibile, in quanto possono essere agevolmente introdotte informazioni sui pasti o sulle dosi di insulina, semplicemente affiancando nuovi kernels ai precedenti.

Il presente elaborato di tesi si inserisce proprio all'interno del settore della predizione della glicemia futura. L'obiettivo è sviluppare nuove metodologie, basate su semplici reti neurali, per predire il livello di glucosio ematico, con un orizzonte di predizione di almeno 30 minuti. Lo scopo è incorporare nei nuovi predittori, oltre alle informazioni relative ai valori glicemici passati, registrate mediante sensori CGM, anche le abbondanti informazioni sui pasti e sulle dosi di insulina. Questi nuovi ingressi dovrebbero migliorare l'accuratezza della predizione, soprattutto in corrispondenza alle inversioni del trend glicemico, solitamente dovute proprio a un pasto e all'iniezione di un bolo di insulina.

1.7 Obiettivi e presentazione della tesi

Lo scopo del presente elaborato è lo sviluppo di nuovi algoritmi per la predizione della glicemia futura, utilizzando misurazioni fornite da sensori per il continuous glucose monitoring. La predizione verrà effettuata sfruttando reti neurali feedforward, che permettono di integrare, tra i loro ingressi, le informazioni relative ai pasti ingeriti dal soggetto, ed eventuali informazioni relative ad altri parametri che notoriamente influenzano il livello di glucosio ematico. Le nuove reti neurali sono state ottimizzate per predire la glicemia futura con un anticipo di 30 minuti, in modo da consentire al soggetto di intervenire opportunamente per evitare crisi ipo e iperglicemiche. Si è inoltre cercato di minimizzare la quantità di informazioni aggiuntive che il paziente deve registrare, per ottenere la predizione glicemica. Le reti create sono modelli di popolazione, allenate sfruttando i profili glicemici di pazienti diabetici di tipo 1, monitorati col sensore AbbottTM FreeStyle NavigatorTM, che fornisce una misurazione glicemica al minuto.

Nel *capitolo 2* verrà presentata la teoria di base sulle reti neurali. Successivamente, nel *capitolo 3*, sarà fatta una breve panoramica dello stato dell'arte relativo alla predizione del livello glicemico. A causa della recente introduzione sul mercato dei sensori CGM, sono pochi in letteratura i lavori in cui vengono utilizzate reti neurali con in ingresso dati di glucosio campionati frequentemente per predire la glicemia futura. Per questo motivo nella presente tesi sono state considerate reti neurali con architetture molto semplici, per indagarne le potenzialità e i difetti, con lo scopo di risolvere il problema della predizione nel modo più semplice ed efficiente possibile. Nei *capitoli 4 e 5* verranno descritti i passi che hanno portato alla scelta delle configurazioni utilizzate per le nuove reti neurali, e nei *capitoli 6 e 7* verranno presentati e commentati i risultati ottenuti, rispettivamente, su un dataset simulato e su un dataset reale. Infine, nel *capitolo 8* si introdurranno i possibili sviluppi futuri.

Capitolo 2

Introduzione alle reti neurali

Una rete neurale (o NN, dall'inglese Neural Network) è un sistema di calcolo parallelo e distribuito con la naturale propensione ad immagazzinare conoscenze basate sull'esperienza, rendendole disponibili per l'uso. Le conoscenze:

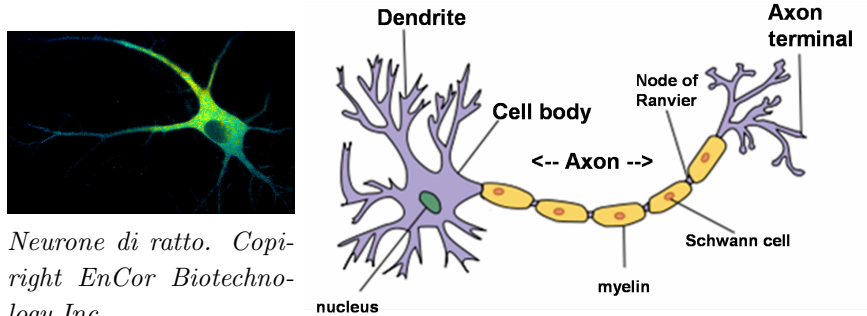
- sono acquisite attraverso un processo di apprendimento (*learning*);
- sono immagazzinate come forza delle interconnessioni tra neuroni (*pesi sinaptici*) [19].

Le reti neurali artificiali si ispirano alle proprietà e alla struttura delle reti neurali biologiche che costituiscono il cervello di uomini e animali e che sono uno degli oggetti più complessi che esista, con una capacità di calcolo ed elaborazione sorprendenti. Nonostante la velocità di calcolo dei moderni processori ($\sim 10^{-9}$ s) sia nettamente superiore alla velocità di elaborazione di un neurone biologico ($\sim 10^{-3}$ s), il numero complessivo di operazioni al secondo compiute dal cervello supera abbondantemente il numero di calcoli svolti da un computer. Questa superiorità è dovuta alla natura distribuita del cervello, dove moltissime unità elementari, dette neuroni, elaborano semplici informazioni in parallelo, anziché in modo sequenziale, come avviene in un computer, e comunicano tra di loro attraverso sinapsi [44].

2.1 Reti neurali biologiche

Una rete neurale biologica (come, per esempio, il sistema nervoso centrale) è un sistema di calcolo massivo parallelo, strutturato in livelli altamente connessi, tra i quali sono scambiati messaggi relativamente semplice.

L'unità di base è il neurone, schematizzato in Figura 2.1. Esso è dotato di un corpo cellulare contenente il nucleo, detto *soma*, da cui si propagano numerosi e brevi processi extracellulari, detti *dendriti*, e un lungo *assone*, che ad una certa distanza dal soma si suddivide in numerosi filamenti che



Neurone di ratto. Copyright EnCor Biotechnology Inc.

Schematizzazione di un neurone biologico [61]

Figura 2.1: Immagine di un neurone.

terminano con i bottoni sinaptici. Ogni neurone acquisisce, attraverso i suoi dendriti, i segnali provenienti dai neuroni ad esso afferenti, elabora la risposta a tali stimoli nel soma e la trasmette, attraverso l'assone, ai neuroni ad esso collegati. Il segnale che si propaga attraverso il neurone è di natura elettrica e, quando raggiunge le sinapsi, viene trasferito chimicamente ai dendriti degli altri neuroni. Quando il segnale raggiunge i bottoni sinaptici

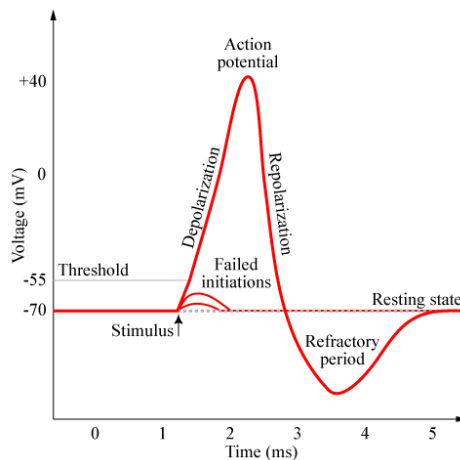


Figura 2.2: Potenziale d'azione tipico di un neurone [60].

dell'assone vengono rilasciate delle sostanze che diffondono verso i dendriti degli altri neuroni, realizzando la connessione e attivando un potenziale elettrico (Figura 2.2). Oltre alle sinapsi chimiche esistono anche alcune sinapsi elettriche, principalmente situate in zone del cervello in cui il segnale deve essere trasmesso estremamente rapidamente. La sinapsi può essere eccitatoria o inibitoria, a seconda che il segnale da essa trasmesso vada o meno a favorire l'attivazione del neurone ricevente. Dopo aver generato un impulso elettrico il neurone necessita di un periodo di riposo e rilassamento, detto tempo refrattario, prima di poter generare un altro impulso. Durante

questo intervallo, pari a circa 1ms, il neurone è insensibile ai segnali che giungono ai suoi dendriti. È opportuno sottolineare la natura binaria del comportamento dei neuroni: essi o generano un potenziale d'azione, (che, in un determinato tipo di cellule, assume sempre la stessa forma e ampiezza), o non lo generano.

La velocità con cui il cervello elabora le informazioni, nonostante la lentezza dei neuroni (se confrontati con i computers), è dovuta all'altissimo numero di neuroni e sinapsi (circa 10^{12} neuroni e $10^{14} - 10^{15}$ sinapsi nel cervello umano), che lavorano contemporaneamente in maniera coordinata. Se si aggiungono o tolgono sinapsi si modifica l'architettura della rete e, di conseguenza, il suo modo di elaborare le informazioni. Per questo si può affermare che una delle funzioni delle sinapsi è proprio immagazzinare le conoscenze [44].

2.1.1 Apprendimento

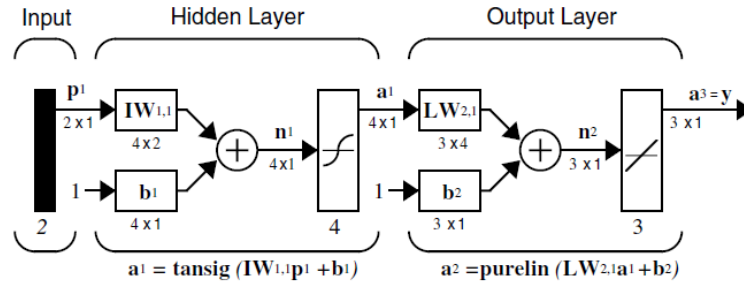
Una delle caratteristiche più sorprendenti degli animali è la loro capacità di memorizzare e apprendere da avvenimenti passati e di modificare il loro comportamento in base all'esperienza. L'informazione è immagazzinata nelle sinapsi, quindi l'apprendimento e la memorizzazione possono avvenire in virtù di cambiamenti dei pesi sinaptici. Questi due processi, che ritroviamo in tutti gli animali, permettono loro di pesare gli eventi passati, per valutare il probabile risultato di un'azione o di una situazione, e di modificare il loro comportamento futuro, quando un'azione o una situazione ha portato ad un risultato negativo e indesiderato [44].

Dopo aver introdotto alcune nozioni basilari sulle reti neurali biologiche passiamo alla trattazione della reti neurali artificiali che ad esse si ispirano.

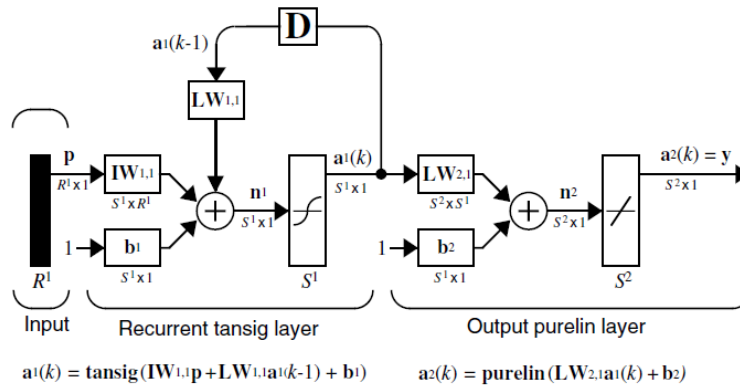
2.2 Reti neurali artificiali (ANN)

Una rete neurale artificiale è un paradigma di calcolo ispirato alla struttura e al funzionamento del cervello. È costituita da un gruppo di neuroni interconnessi che elaborano semplici informazioni. Solitamente una rete neurale modifica la sua struttura in base alle informazioni esterne ed interne che le sono presentate durante il processo di apprendimento. Esistono molti tipi differenti di reti neurali e una scelta comune è classificarle in base alla loro struttura:

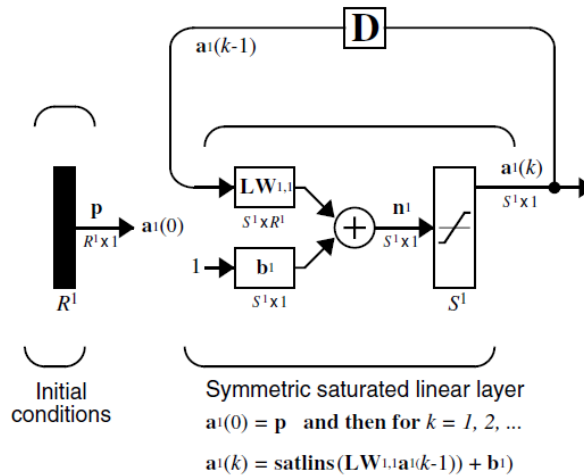
- una *feedforward neural network* (Figura 2.3(a)) contiene solo connessioni "in avanti", ovvero l'informazione fluisce ordinatamente e sequenzialmente dall'ingresso verso l'uscita;
- una *recurrent neural network* (Figura 2.3(b) e Figura 2.3(c)), invece, contiene connessioni che puntano sia in avanti che indietro, renden-



(a) Rete neurale feedforward, costituita da un unico strato di neuroni nascosti, con funzione d'attivazione sigmoidale e da uno strato d'uscita, con funzione d'attivazione lineare [20].



(b) Elman neural network. La ricorsione avviene dall'uscita dello strato nascosto all'ingresso della rete.



(c) Hopfield neural network. La ricorsione avviene dall'uscita all'ingresso della rete.

Figura 2.3: Esempi di reti neurali feedforward e di reti neurali ricorsive [20].

do così necessaria l'introduzione di ritardi. Ad ogni connessione è associato un peso con un certo valore numerico.

Ci sono vari modi per allenare una rete neurale, ovvero per determinarne i pesi in modo che risolva il compito desiderato. Tipicamente si distingue tra:

- *apprendimento supervisionato*, durante il quale per ogni vettore di ingresso è noto il vettore d'uscita desiderato. Di conseguenza in questo metodo d'apprendimento l'obiettivo è minimizzare l'errore, ovvero la differenza tra l'uscita desiderata e l'uscita reale della rete;
- *apprendimento debolmente guidato*, in cui non si riesce a generare un segnale d'errore per ogni vettore d'ingresso presentato alla rete. Anche in questo caso è comunque dato un certo feedback alla rete, ma non dopo ogni istante temporale, bensì più saltuariamente e solitamente alla fine di una lunga valutazione;
- *apprendimento non supervisionato*, durante il quale la rete impara da sola a riconoscere proprietà statistiche importanti degli ingressi che le vengono presentati e risponde in modo appropriato a insiemi di vettori simili in ingresso. Questa tecnica è solitamente utilizzata per la classificazione, quando non sono note le classi a cui i diversi vettori appartengono [44].

2.2.1 Il neurone artificiale

L'unità di base di una ANN, analogamente a una rete neurale biologica, è il neurone artificiale. Figura 2.4 mostra il modello di neurone più utilizzato.

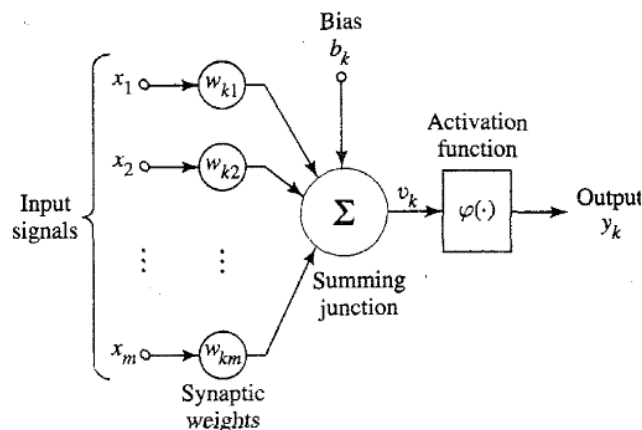


Figura 2.4: Modello di neurone non lineare [19].

In esso identifichiamo tre elementi di base:

1. Un insieme di *sinapsi* o *connessioni*, ognuna caratterizzata da un *peso*, indicante la sua forza. Un ingresso x_j collegato al neurone k è moltiplicato per il peso sinaptico w_{kj} ¹. Il peso w_{kj} è positivo se associato a una sinapsi eccitatoria, negativo se associato a una inibitoria.
2. Un sommatore che combina i segnali in ingresso al neurone pesati dalle rispettive sinapsi.
3. Una funzione di attivazione ($\varphi(\cdot)$), che limita l'ampiezza dell'uscita del neurone in un determinato intervallo di valori finiti (tipicamente in $[0, 1]$ o in $[-1, 1]$).

Il modello di neurone mostrato in Figura 2.4 include anche un bias b_k che ha l'effetto di accrescere l'ingresso della funzione di attivazione.

In termini matematici un neurone può essere descritto dalle seguenti equazioni:

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

e

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

dove x_1, x_2, \dots, x_m sono gli *ingressi* della rete; $w_{k1}, w_{k2}, \dots, w_{km}$ sono i *pesi sinaptici* relativi al neurone k ; v_k è l'*uscita del sommatore lineare*; b_k è il *bias*; $\varphi(\cdot)$ è la *funzione d'attivazione*; e y_k è l'*uscita* della rete. L'effetto del bias è quello di applicare una trasformazione affine all'uscita v_k del sommatore lineare del modello di neurone di Figura 2.4. In particolare, in base al segno del bias, la relazione tra il potenziale di attivazione effettivo del neurone k e l'uscita del sommatore lineare è modificata come illustrato in Figura 2.5. Poichè il bias b_k è un parametro esterno del neurone k , così come lo sono i pesi, possiamo riformulare l'equazione (2.1) come

$$v_k = \sum_{j=0}^m w_{kj} x_j \quad (2.3)$$

e la (2.2) diventa

$$y_k = \varphi(v_k) \quad (2.4)$$

Questo risultato si ottiene semplicemente aggiungendo una nuova sinapsi, il cui ingresso è fissato e vale

$$x_0 = +1 \quad (2.5)$$

¹È opportuno sottolineare che il peso sinaptico w_{kj} indica un collegamento che va al neurone k (come indicato dal primo pedice) partendo dalla sinapsi j (come indica il secondo pedice). In letteratura è possibile trovare anche l'opposto di questa notazione.

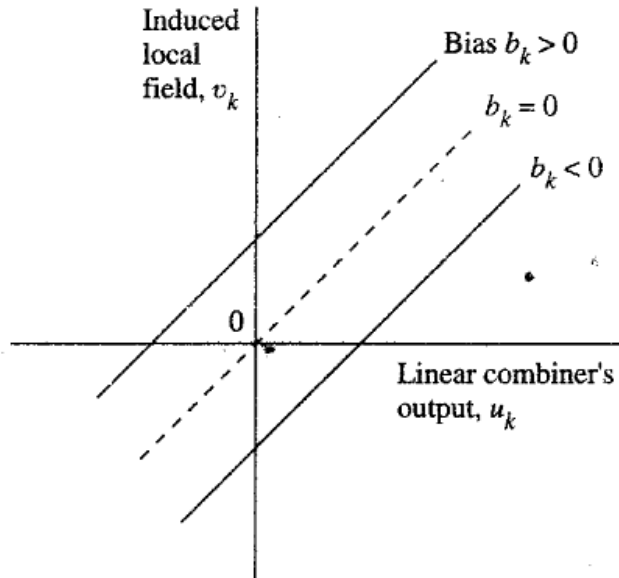


Figura 2.5: Trasformazione affine prodotta dalla presenza del bias [19].

e il cui peso è

$$w_k = b_k \quad (2.6)$$

Il modello del neurone diventa quindi quello illustrato in Figura 2.6.

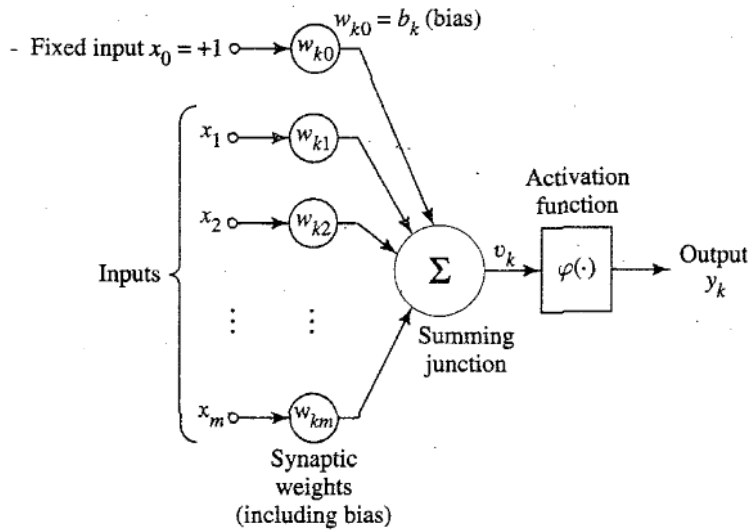


Figura 2.6: Riformulazione del modello del neurone. Il bias è stato incluso tra i pesi grazie all'aggiunta di un nuovo ingresso il cui valore è sempre +1 [19].

2.2.2 Funzioni di attivazione

La funzione di attivazione, indicata con $\varphi(\cdot)$, calcola l'uscita del neurone in base al livello di attività al suo ingresso. I principali tipi di funzione di attivazione sono:

1. *Funzione a soglia.* Questa funzione, rappresentata nel pannello 1 di Figura 2.7, è definita dall'espressione

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad (2.7)$$

Quindi, se v_k è l'attività interna del neurone, definita da

$$v_k = \sum_{j=0}^m w_{kj} x_j \quad (2.8)$$

l'uscita del neurone stesso, utilizzando questa funzione di attivazione, sarà

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

Questo neurone segue il comportamento tutto o nulla, in quanto la sua uscita può assumere solo due valori, ovvero 1 se la sua attività interna supera una certa soglia, 0 altrimenti. Questa funzione è particolarmente usata nel caso di reti utilizzate per la classificazione.

2. *Funzione segno.* In alcune situazioni può essere opportuno che l'uscita della funzione di attivazione sia nel range $[-1, 1]$ e assuma una forma antisimmetrica rispetto all'origine. In tal caso la *funzione soglia* (2.7) è riformulata nel seguente modo:

$$\varphi(v) = \begin{cases} -1 & \text{se } v < 0 \\ 0 & \text{se } v = 0 \\ 1 & \text{se } v > 0 \end{cases} \quad (2.10)$$

3. *Funzione lineare a tratti.* Questa funzione è definita da

$$\varphi(v) = \begin{cases} 0 & \text{se } v \leq -\frac{1}{2} \\ v & \text{se } -\frac{1}{2} < v < \frac{1}{2} \\ 1 & \text{se } v \geq \frac{1}{2} \end{cases} \quad (2.11)$$

dove il fattore di amplificazione, nella regione lineare, è unitario. Questa funzione tende alla funzione soglia se il fattore di amplificazione nella zona lineare è reso infinitamente grande.

4. *Funzione lineare.* È semplicemente definita come

$$\varphi(v) = v \quad (2.12)$$

ed è molto utilizzata come funzione d'attivazione dei neuroni dello strato d'uscita, soprattutto se l'uscita target da riprodurre varia in un certo intervallo di valori, non necessariamente finito.

5. *Funzione sigmoideale.* È una funzione strettamente crescente, continua, derivabile e limitata tra 0 e 1. Un esempio di funzione sigmoideale è la funzione logistica

$$\varphi(v) = \frac{1}{1 + e^{-cv}} \quad (2.13)$$

dove c è un parametro che ne influenza la pendenza (come illustrato nel pannello 5 di Figura 2.7).

6. *Funzione tangente iperbolica.* Assieme alla funzione sigmoideale è la più utilizzata tra le funzioni di attivazione. Ha proprietà simili alla funzione sigmoideale, ma, a differenza di questa, assume anche valori negativi ed è antisimmetrica rispetto all'origine. Essa è definita da

$$\varphi(v) = \frac{1 - e^{-2cv}}{1 + e^{-2cv}} \quad (2.14)$$

Questa funzione torna particolarmente utile quando l'uscita del neurone deve essere nulla se la somma pesata dei suoi ingressi è nulla, e se si desidera esso abbia un comportamento antisimmetrico rispetto alla soglia d'attivazione interna raggiunta. Un'altra caratteristica importantissima che questa funzione condivide con la funzione sigmoideale è il suo comportamento non lineare che permette di ottenere uscite che dipendono non linearmente dagli ingressi del neurone

La teoria sulle reti neurali qui riportata è stata tratta principalmente da [19].

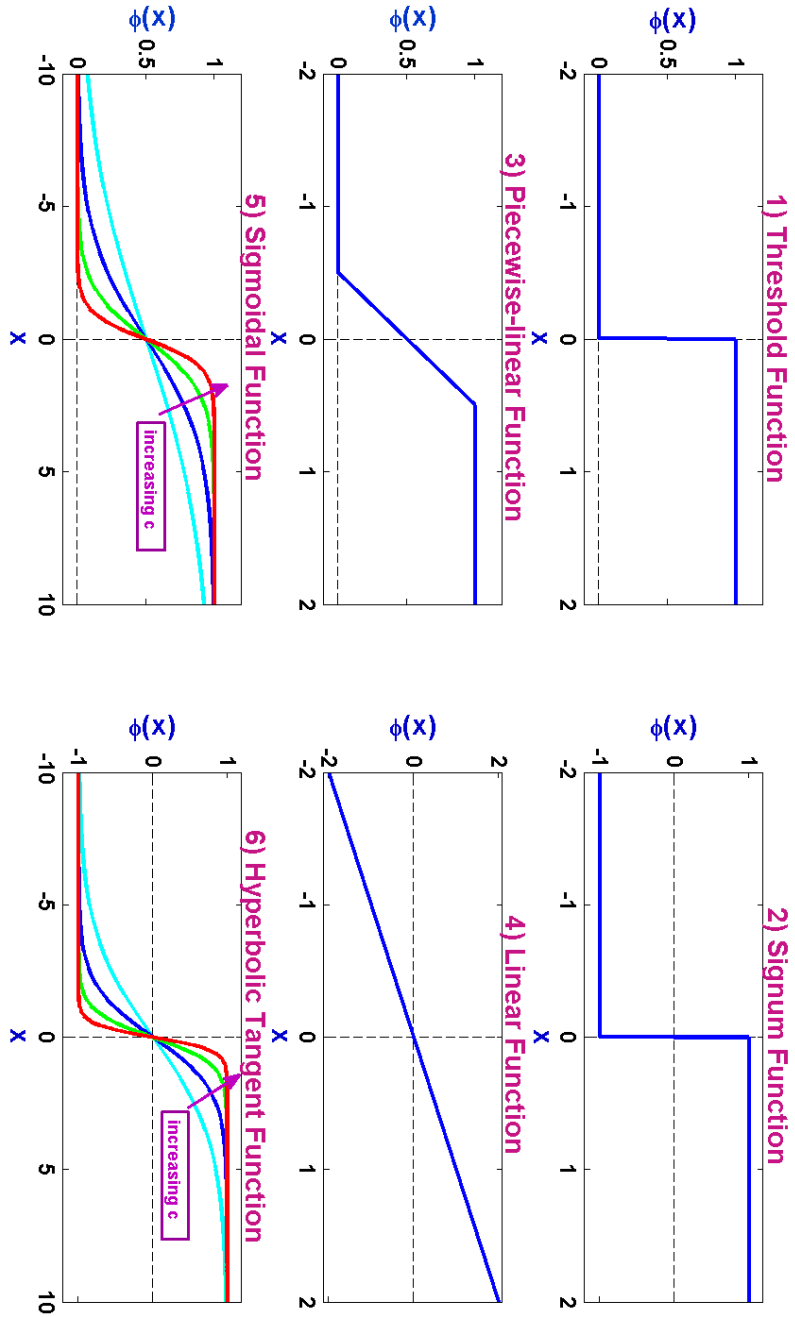


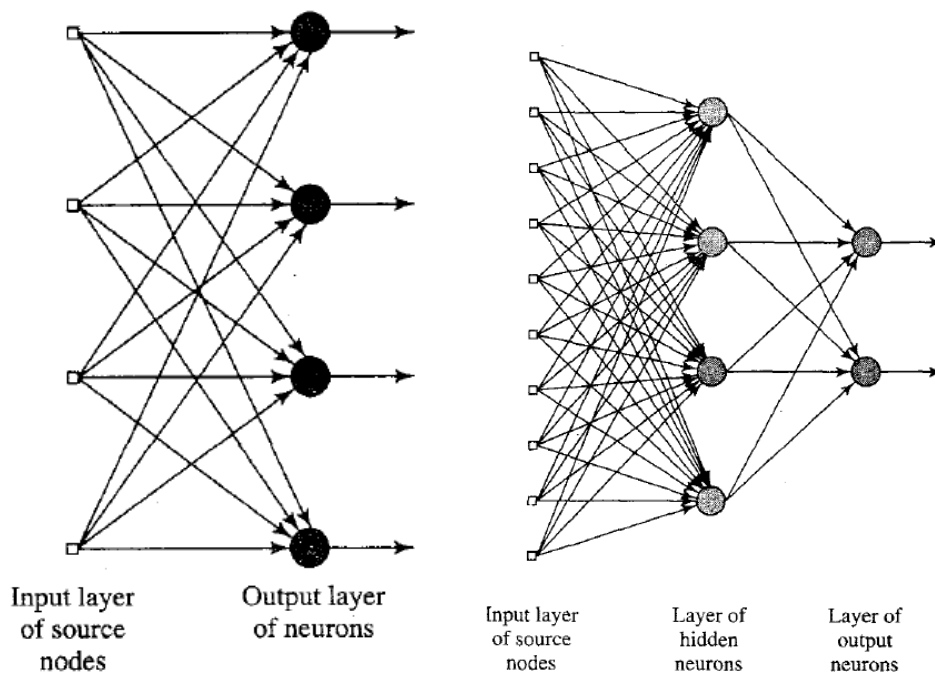
Figura 2.7: Funzioni di attivazione

2.3 Architettura della rete

Il modo in cui i neuroni di una rete sono collegati e strutturati ne definisce l'architettura ed è strettamente connessa all'algoritmo di training utilizzato per allenare la rete.

2.3.1 Reti neurali feedforward (FFNN)

Una rete neurale feedforward (Feedforward Neural Network, FFNN) è una rete in cui i neuroni sono organizzati in strati. Nella sua forma più semplice, illustrata in Figura 2.8(a), c'è solo uno strato d'ingresso che dirige l'informazione allo strato dei neuroni d'uscita, dove essa viene processata. Questa rete è definita a *singolo strato*, (*single-layer network*), dato che lo strato d'ingresso non è contato, poichè in esso non avviene nessuna operazione sui dati. In questa rete gli ingressi, opportunamente pesati, vengono elaborati direttamente dai neuroni dello strato d'uscita.



(a) Rete neurale feedforward a singolo strato. (b) Rete neurale feedforward multistrato.

Figura 2.8: Esempi di reti neurali feedforward [19]. In Figura 2.8(a) è rappresentata una rete neurale con un solo strato di neuroni. In Figura 2.8(b) è rappresentata una ffn con uno strato nascosto costituito da 4 neuroni e con 2 neuroni d'uscita [19].

In Figura 2.8(b) è rappresentata una rete neurale *multistrato*, (*multilayer network*), caratterizzata da uno o più strati di neuroni nascosti, la cui

funzione è intervenire tra ingresso esterno e uscita della rete, elaborando le informazioni fornite alla rete ed estraendo statistiche di ordine superiore rispetto ad una rete monostrato. Solitamente la funzione d'attivazione è di tipo sigmoideale o tangente iperbolica per i neuroni degli strati nascosti, mentre per i neuroni dell'uscita può essere sia di forma sigmoideale che lineare, in base al compito che la rete neurale dovrà risolvere.

I nodi nel primo strato ricevono ognuno un elemento dal vettore d'ingresso e lo trasferiscono ai neuroni del secondo strato (ovvero il primo strato nascosto), i quali elaborano le informazioni ricevute e le passano ai neuroni dello strato immediatamente successivo, proseguendo così fino ai neuroni dell'ultimo strato della rete, la cui uscita corrisponde alla risposta della rete al vettore d'ingresso. Tipicamente i neuroni di ogni strato della rete hanno in ingresso solamente i segnali d'uscita dello strato precedente e dirigono la loro uscita esclusivamente ai neuroni dello strato successivo.

Una rete neurale con p nodi d'ingresso, h_1 neuroni nel primo strato nascosto, h_2 neuroni nel secondo strato nascosto e q neuroni nello strato d'uscita è brevemente indicata come una rete $p - h_1 - h_2 - q$ [19].

Una rete è detta *totalmente connessa* se ogni neurone, in ogni strato, è connesso con ogni neurone dello strato successivo. Se invece mancano alcuni collegamenti la rete è definita *parzialmente connessa*. Solitamente una rete neurale parzialmente connessa ha una struttura che riflette le informazioni note a priori sulle caratteristiche della relazione tra ingresso e uscita che la rete deve modellare [44].

Generalmente il numero di strati e il numero di neuroni in ogni strato vengono settati in base a una procedura di tipo *trial and error*, ovvero testando varie configurazioni possibili e scegliendo quella che produce il risultato migliore. Ciò porta indubbiamente ad una soluzione sub-ottima, poichè non si può testare tutte le architetture immaginabili. D'altra parte, non esistendo regole, se non empiriche, per la costruzione di una rete neurale feedforward, questa è l'unica via percorribile.

Le reti neurali feedforward sono le più utilizzate, sia grazie alla loro semplicità che alla loro versatilità. In particolare sono spesso usate per il riconoscimento vocale ed ottico, per la classificazione, per l'identificazione e il controllo di sistemi, per la predizione e per l'approssimazione di funzioni [19].

Per quanto riguarda l'approssimazione di funzioni, è opportuno citare il

*** Teorema dell'approssimazione universale:**

Sia $\varphi(\cdot)$ una funzione continua, non costante, limitata e monotona crescente.

Sia I_p l'iper-cubo unitario p -dimensionale $[0, 1]^p$.

Sia $C(I_p)$ lo spazio delle funzioni continue in I_p .

Allora, dati una funzione $f \in C(I_p)$ e $\epsilon > 0$, esiste un intero M ed un insieme di costanti reali $\alpha_i, \theta_i,$ e w_{ij} , dove $i = 1, \dots, M$ e $j = 1, \dots, p$ così

che possiamo definire

$$F(x_1, \dots, x_p) = \sum_{i=1}^M \alpha_i \varphi \left(\sum_{j=1}^p w_{ij} x_j - \theta_i \right) \quad (2.15)$$

come una realizzazione approssimata della funzione $f(\cdot)$, ovvero

$$|F(x_1, \dots, x_p) - f(x_1, \dots, x_p)| < \epsilon$$

per ogni $\{x_1, \dots, x_p\} \in I_p$.

*

Questo teorema è direttamente applicabile alle reti feedforward con un unico strato di neuroni nascosti. Sia la funzione logistica che la tangente iperbolica, normalmente utilizzate come funzioni di attivazione dei neuroni degli strati nascosti, soddisfano le condizioni imposte per la funzione $\varphi(\cdot)$. Notiamo, inoltre, che la 2.15 corrisponde all'uscita di una FFNN con p nodi d'ingresso ed un singolo strato nascosto di M neuroni; gli ingressi sono indicati come x_1, \dots, x_p , il neurone nascosto i -esimo ha pesi w_{i1}, \dots, w_{ip} e bias $-\theta_i$, l'uscita della rete è una combinazione lineare delle uscite dei neuroni nascosti, pesate con i coefficienti $\alpha_1, \dots, \alpha_M$.

È opportuno sottolineare che il *teorema di approssimazione universale* è importante soprattutto dal punto di vista teorico, poichè dichiara la possibilità di approssimare una funzione con una rete neurale feedforward con un singolo strato, ma non dà indicazioni sull'architettura della rete da usare; per questo ha un valore pratico molto limitato. Inoltre esso ipotizza che la funzione da approssimare sia nota, e che sia possibile utilizzare un numero arbitrariamente grande di neuroni nello strato nascosto. Entrambe queste assunzioni sono solitamente violate nella realtà.

Il problema delle FFNN con un unico strato nascosto è che i neuroni tendono ad interagire tra loro globalmente e ciò può rendere difficile migliorare l'approssimazione in un punto, senza peggiorarla altrove. Per questo molte reti neurali feedforward sono costituite da due strati nascosti [19].

2.3.2 Reti neurali ricorsive (RNN)

Una *rete neurale ricorsiva* (*recurrent neural network, RNN*), a differenza di una rete feedforward, ha almeno una connessione che va ad uno strato precedente. Un esempio è la rete di Figura 2.9, costituita da uno strato di ingressi, uno strato di neuroni nascosti e due neuroni d'uscita. La ricorsione avviene sia dai neuroni nascosti, sia dai neuroni d'uscita. La presenza di feedback ha un grandissimo impatto sulle performance e sulle capacità di apprendimento della rete. Inoltre, la presenza di ricorsioni richiede l'introduzione di elementi di ritardo, (coerentemente a quanto rappresentato in Figura 2.9 il ritardo unitario è indicato con z^{-1}) [19].

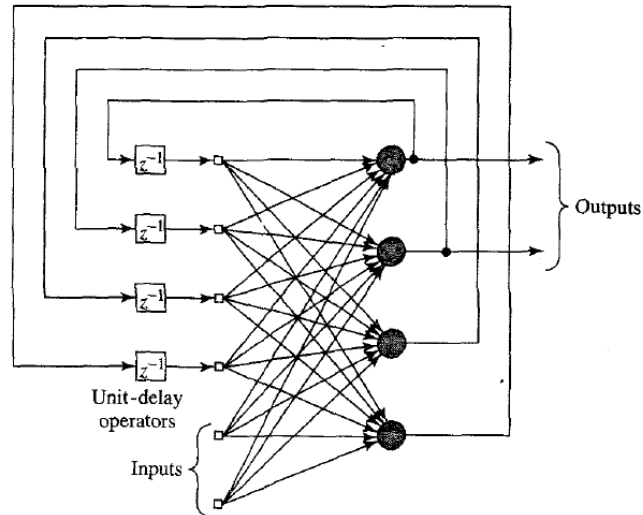


Figura 2.9: Rete neurale ricorsiva con neuroni nascosti [19].

Proprio la presenza di ritardi differenzia profondamente il comportamento di una rete ricorsiva da quello di una rete feedforward. Anche se una FFNN può modellare una grande varietà di relazioni ingresso-uscita, essa ha il difetto di mancare totalmente di memoria dinamica. Ovvero, l'uscita di una FFNN sarà sempre la stessa se in ingresso le viene presentato lo stesso vettore. Chiaramente questo comportamento non è biologico e, come già sottolineato nella Sezione 2.1.1, la memoria è una caratteristica essenziale degli animali e permette loro di rispondere in modo diverso quando si ripresenta un evento a cui avevano precedentemente risposto in maniera inappropriata, arrivando ad un epilogo poco piacevole. Una rete ricorsiva, invece, ha una memoria teoricamente infinita, proprio grazie alla presenza di feedback e ritardi, e la sua uscita non dipende solamente dal vettore in ingresso, ma anche dallo stato interno della rete, permettendole di rispondere in maniera diversa ad uno stesso ingresso [21].

Due reti ricorsive storiche sono la *Elman NN* e la *Hopfield NN*.

- Una *Elman network* (Figura 2.3(b)) è una rete con due strati e feedback dall'uscita del primo strato al suo ingresso. Questa ricorsione permette alla rete di individuare e anche generare relazioni temporvarianti.
- Una *Hopfield network* (Figura 2.3(c)) è una rete in cui l'uscita è rimandata in ingresso al primo strato di neuroni. Questa rete ha la capacità di imparare uno specifico insieme di punti di equilibrio così che, quando le viene presentato un certo ingresso, essa evolve fino ad arrestarsi su uno di questi. Uno dei problemi connessi a questa

architettura è la possibile presenza di punti di equilibrio spuri, oltre a quelli desiderati [19].

2.3.3 Reti neurali radial-basis function (RBFNN)

Questi modelli sono nati interpretando la costruzione di una rete neurale come un problema di approssimazione di una curva in uno spazio a molte dimensioni. Coerentemente con questo punto di vista, l'apprendimento da parte della rete corrisponde a trovare una superficie che fitti bene i dati utilizzati per allenare la rete, con la bontà del risultato misurata in qualche modo statistico. Analogamente l'abilità della rete di generalizzare coincide con la sua capacità di interpolare bene, con questa superficie multidimensionale, anche i dati del test set. Nell'ambito delle reti neurali i singoli neuroni forniscono un insieme di funzioni che costituiscono una base per espandere i vettori d'ingresso nello spazio dello strato nascosto. Queste funzioni sono le *funzioni radiali* (o *radial-basis functions*).

Una *rete neurale radial-basis function* è composta da tre strati: un primo strato d'ingresso, un secondo strato con molti neuroni e lo strato d'uscita. La trasformazione tra l'ingresso e il primo strato nascosto è non lineare, mentre quella tra lo strato nascosto e l'uscita è lineare. Una giustificazione a questa architettura viene dal teorema di Cover che afferma che un problema di classificazione, mappato non linearmente in uno spazio ad alta dimensionalità, ha più probabilità di essere linearmente separabile che in uno spazio a poche dimensioni.

Una RBFNN differisce dalle reti finora viste in due aspetti fondamentali:

1. l'ingresso di ogni neurone del primo strato nascosto è 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. (Mentre in una rete "classica" l'ingresso di un neurone è la somma del bias e degli elementi del vettore d'ingresso della rete, moltiplicati per i pesi delle sinapsi che li collegano al neurone in esame).
2. la funzione di trasferimento dei neuroni dello strato nascosto è una funzione radiale. Una tale funzione, di cui troviamo un esempio in Figura 2.10, assume valori crescenti man mano che diminuisce la distanza tra i pesi e gli elementi del vettore d'ingresso, fino a valere 1 quando tale distanza è nulla.

Esistono due approcci nella costruzione di una RBFNN. Il primo rende nullo l'errore commesso dalla rete in fase di training, ma porta ad una rete con un numero di neuroni enorme. Il secondo approccio, invece, contiene il numero di neuroni della rete limitandosi a richiedere che l'errore sia inferiore ad un certo valore prestabilito. Nel primo caso il numero di neuroni creato eguaglia il numero di coppie ingresso-uscita presentate alla rete durante

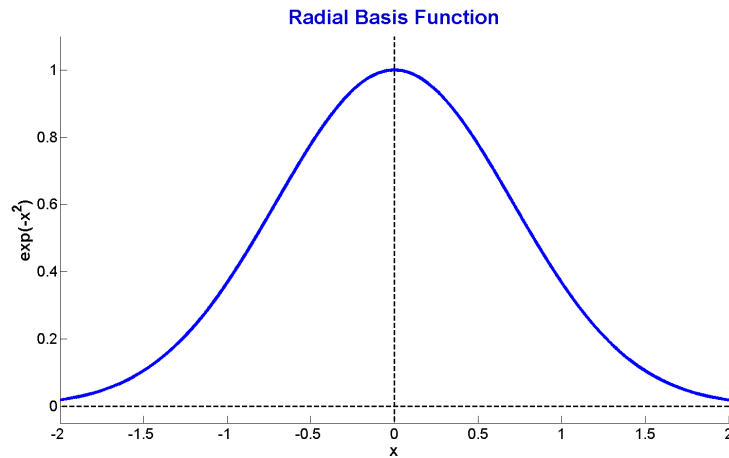


Figura 2.10: Un esempio di funzione radiale.

l'allenamento, e i pesi di ogni neurone corrispondono con gli elementi di uno dei vettori d'ingresso. In tal caso ogni neurone riconosce il corrispondente ingresso e produce l'uscita target. Chiaramente questa soluzione non è accettabile quando il numero di vettori con cui la rete viene allenata deve essere grande, per ottenere un comportamento adeguato su tutto lo spazio degli ingressi. La seconda tecnica, invece, crea un neurone alla volta e si arresta non appena l'errore è sufficientemente piccolo. I pesi di ogni nuovo neurone aggiunto eguagliano il valore degli elementi di uno degli ingressi, in particolare, tra tutti quelli non ancora utilizzati, si sceglie quello che abbassa maggiormente l'errore di predizione.

Le RBFNN vengono utilizzate per l'analisi d'immagini, per il riconoscimento vocale, per l'analisi di serie temporali, per la classificazione e, spesso, anche per la predizione [19].

2.3.4 Reti competitive e self organizing maps

Le *reti competitive* sono reti che imparano da sole a individuare certe regolarità e correlazioni tra i loro ingressi e adattano la loro risposta di conseguenza. I neuroni di queste reti, particolarmente adatte per la classificazione, riescono a riconoscere gruppi di ingressi simili e appartenenti con alta probabilità ad una stessa classe. Le *self organizing maps* imparano a riconoscere questi gruppi di ingressi simili in modo che neuroni vicini tra loro rispondano a ingressi simili. Le due reti finora descritte imparano automaticamente come classificare gli ingressi. A volte, però, si può desiderare che la rete classifichi gli ingressi entro classi prestabilite dall'utente: questo obiettivo è raggiunto utilizzando le *learning vector quantization nets* [20].

2.4 Allenamento supervisionato della rete (supervised training)

Tra le molte proprietà delle reti neurali una delle più interessanti è la loro capacità di imparare dall'ambiente esterno e di migliorare le loro performance attraverso un processo di *apprendimento* o *learning*. Esso avviene modificando iterativamente i pesi sinaptici e il bias di ogni neurone. Le modifiche ai pesi possono essere effettuate dopo la presentazione di ogni coppia ingresso-uscita del training set (*on-line training*), oppure dopo la presentazione dell'intero training set (*batch training*) alla rete neurale.

In particolare, si definisce *epoca* una presentazione completa di tutti gli elementi del training set alla rete.

La teoria sulle tecniche di apprendimento utilizzabili durante il training delle reti neurali, riportata nei paragrafi seguenti, è stata tratta da [19], salvo ove diversamente specificato.

2.4.1 Apprendimento “errore-correzione” (error-correction learning)

Sia $d_k(n)$ l'uscita desiderata (*target*) per il neurone k al tempo n . Sia $y_k(n)$ l'uscita reale del neurone k al tempo n , in risposta al vettore d'ingresso $\mathbf{x}(n)$.

Tipicamente l'uscita della rete è diversa dalla risposta target. Possiamo quindi calcolare l'errore

$$e_k(n) = d_k(n) - y_k(n) \quad (2.16)$$

L'obiettivo di questa tecnica di apprendimento è la minimizzazione di una funzione costo basata su $e_k(n)$. Solitamente si utilizza come funzione costo l'*errore quadratico medio*, definito da

$$J = E \left[\frac{1}{2} \sum_{k \in C} e_k^2(n) \right] \quad (2.17)$$

dove E indica l'aspettazione o media statistica e C è l'insieme di tutti i neuroni dello strato d'uscita della rete. La minimizzazione di J , rispetto ai parametri della rete, è solitamente eseguita utilizzando il *metodo del gradiente*. Poichè il calcolo dell'aspettazione richiede la conoscenza delle caratteristiche statistiche del processo, che solitamente non sono note, essa viene in pratica sostituita da

$$\mathcal{E} = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad (2.18)$$

All'istante n , la modifica dei pesi è

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) \quad (2.19)$$

dove η è una costante positiva che determina la *percentuale di apprendimento* (*rate of learning*). In pratica questa modifica è proporzionale al prodotto tra l'errore e il segnale d'ingresso in questione. Infine, il peso assume il valore

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (2.20)$$

Il valore del parametro η è critico, in quanto un valore piccolo produce cambiamenti piccoli dei pesi, permettendo di analizzare molto finemente la superficie dell'errore, ma dilatando notevolmente il tempo necessario per arrivare a convergenza; di contro η grande velocizza il training della rete, ma c'è il rischio che compaiano comportamenti oscillatori e che l'algoritmo non analizzi adeguatamente le varie regioni dello spazio dei pesi.

In base al tipo di funzioni di attivazione utilizzate nella rete distinguiamo due casi in cui l'algoritmo descritto si comporta in modo diverso:

1. La rete neurale è costituita esclusivamente da neuroni lineari. In tal caso la superficie dell'errore è una funzione quadratica dei pesi, concava e con un unico minimo, che il metodo del gradiente è in grado di identificare con certezza.
2. I neuroni della rete sono non lineari. La superficie dell'errore ha uno o più minimi globali e molti minimi locali. In tal caso il metodo del gradiente potrebbe restare imprigionato in un minimo locale e non è garantita l'identificazione della soluzione ottima.

L'algoritmo di back-propagation (BP)

Questo algoritmo è basato sulle regole dell'apprendimento *error-correction* appena descritto. È l'algoritmo più utilizzato per allenare le reti neurali feedforward e ne esistono molte varianti che ne ottimizzano la velocità e l'abilità di individuare un minimo globale della funzione errore, senza restare bloccato in un minimo locale. Ne è stata elaborata anche una versione "dinamica" (*back-propagation through time*), spesso utilizzato per allenare le reti ricorsive. Risulta, però, meno efficiente di quello utilizzato per le FFNN, sia in termini di tempo richiesto, sia in termini di performance finali della rete.

L'algoritmo è costituito da due parti fondamentali: un *passaggio in avanti* (*forward pass*) e un *passaggio all'indietro* (*backward pass*).

- Nella *prima fase* l'ingresso viene presentato alla rete e viene calcolata l'uscita corrispondente. I pesi rimangono inalterati. L'uscita della rete viene confrontata con l'uscita desiderata e si calcola l'errore commesso da ogni neurone dello strato d'uscita.
- Nella *seconda fase* i pesi sinaptici vengono modificati applicando la regola "errore-correzione". L'errore viene propagato all'indietro attraverso la rete (dall'uscita verso l'ingresso) e si cerca di capire l'errore

2.4 Allenamento supervisionato della rete (supervised training)

commesso da ogni neurone nascosto, di cui non sono note né l'uscita target, né quella reale, in modo da modificare adeguatamente i pesi ad esso relativi.

L'errore del neurone d'uscita j all'iterazione n (ovvero quando viene presentata alla rete l' n -esima coppia ingresso-uscita) vale

$$e_j(n) = d_j(n) - y_j(n) \quad (2.21)$$

Il valore istantaneo della somma del quadrato degli errori di tutti i neuroni d'uscita all'istante n vale

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (2.22)$$

dove C indica l'insieme di neuroni d'uscita. Se indichiamo con N il numero totale di coppie ingresso-uscita appartenenti al *training set* l'errore quadratico medio vale

$$\begin{aligned} \mathcal{E}_{av} &= \frac{1}{N} \sum_{n=1}^N \mathcal{E}(n) = \frac{1}{2N} \sum_{n=1}^N \sum_{j \in C} e_j^2(n) \\ &= \frac{1}{2N} \sum_{n=1}^N \sum_{j \in C} (d_j(n) - y_j(n))^2 \end{aligned} \quad (2.23)$$

L'errore istantaneo e l'errore medio sono entrambi funzione dei parametri liberi della rete. Per un dato training set \mathcal{E}_{av} rappresenta la funzione costo da minimizzare.

L'allenamento della rete può essere eseguito con due modalità:

- Si parla di *pattern* o (*on-line*) *learning* se i pesi della rete sono modificati dopo la presentazione di ogni ingresso del training set. (Si minimizza, ad ogni passo, $\mathcal{E}(n)$).
- Si parla di *batch learning* se i pesi vengono modificati dopo la presentazione dell'intero training set, sulla base dell'errore totale commesso dalla rete. (Si va a minimizzare solitamente \mathcal{E}_{av}).

Sia $\mathbf{y}(n) = [y_1(n), \dots, y_p(n)]$ il segnale in ingresso al neurone j al passo n . Lo stato di attivazione del neurone è

$$v_j(n) = \sum_{i=0}^p w_{ji}(n) y_i(n) \quad (2.24)$$

dove il peso w_{j0} (corrispondente all'ingresso fisso $y_0 = -1$) assume il valore del bias b_j . L'uscita del neurone j vale

$$y_j(n) = \varphi_j(v_j(n)) \quad (2.25)$$

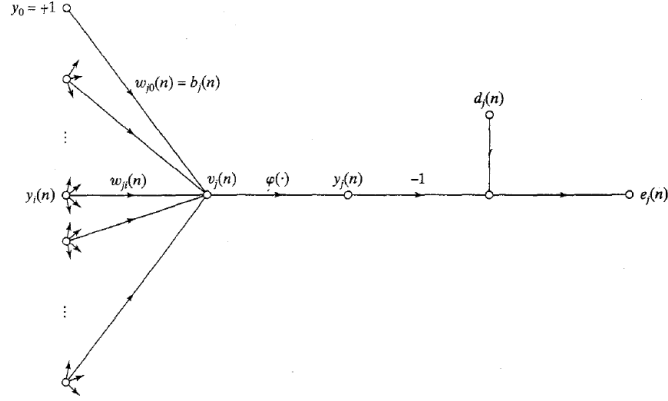


Figura 2.11: Grafo indicante il flusso del segnale, dall'ingresso all'uscita del neurone j [19].

L'algoritmo di back-propagation corregge il peso w_{ji} sommandogli un termine proporzionale al gradiente dell'errore, istantaneo ($\partial \mathcal{E}(n) / \partial w_{ji}(n)$) o medio ($\partial \mathcal{E}_{av} / \partial w_{ji}$), in base al metodo di training adottato. Presenteremo qui di seguito le formule utilizzate dalla modalità *pattern training*.

Il gradiente dell'errore istantaneo può essere espresso come

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (2.26)$$

Sviluppiamo il termine di destra eseguendo le varie derivate parziali. Dalla 2.22 ricaviamo

$$\frac{\partial \mathcal{E}(n)}{\partial e_j(n)} = e_j(n) \quad (2.27)$$

dalla 2.21 otteniamo

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (2.28)$$

dalla 2.25 abbiamo

$$\frac{\partial y_i(n)}{\partial v_j(n)} = \varphi'_j(v_j(n)) \quad (2.29)$$

dove $\varphi'_j(\cdot)$ indica la derivata di $\varphi_j(\cdot)$ rispetto al suo argomento. Infine, dalla 2.24 otteniamo

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n). \quad (2.30)$$

Sostituendo le equazioni dalla 2.27 alla 2.30 nella 2.26 ricaviamo

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = -e_j(n) \varphi'_j(v_j(n)) y_i(n) \quad (2.31)$$

La correzione $\Delta w_{ji}(n)$ da applicare a $w_{ji}(n)$, in base alla regola del gradiente, è

$$\begin{aligned}\Delta w_{ji}(n) &= -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \\ &= \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \\ &= \eta \delta_j(n) y_i(n)\end{aligned}\tag{2.32}$$

dove è stato introdotto il *gradiente locale* $\delta_j(n)$ definito da

$$\begin{aligned}\delta_j(n) &= -\frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ &= e_j(n) \varphi'_j(v_j(n))\end{aligned}\tag{2.33}$$

Dalle equazioni 2.32 e 2.33 notiamo che l'errore $e_j(n)$ è un fattore chiave nel calcolo delle modifiche $\Delta w_{ji}(n)$ da apportare a $w_{ji}(n)$. In base alla posizione del neurone j all'interno della rete possiamo distinguere due casi: se j è un neurone d'uscita il calcolo dell'errore da esso commesso è particolarmente semplice, se j è un neurone nascosto non è nota l'uscita target ad esso relativa. Anche se i neuroni nascosti non sono direttamente accessibili, essi sono comunque in parte responsabili degli errori commessi dai neuroni d'uscita. Questo problema, noto con il nome di *credit-assignment problem*, è risolto dall'algoritmo di back-propagation propagando all'indietro l'errore, attraverso la rete. Il segnale d'errore di un neurone nascosto è calcolato ricorsivamente, tenendo conto dei segnali d'errore di tutti i neuroni a cui quel neurone dirige la sua uscita.

Partendo dall'equazione 2.33 possiamo ridefinire il *gradiente locale* di un neurone nascosto come

$$\begin{aligned}\delta_j(n) &= -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ &= -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \varphi'_j(v_j(n))\end{aligned}\tag{2.34}$$

dove, per passare alla seconda riga, abbiamo sfruttato il risultato ottenuto in 2.29.

Come già visto in 2.22, se k è un nodo d'uscita, l'errore istantaneo ad esso associato vale

$$\mathcal{E}(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n)\tag{2.35}$$

Vale la pena sottolineare che questa equazione differisce dalla versione precedentemente proposta solo per la sostituzione dell'indice j con l'indice k , necessaria per evitare confusione, dato che in questo contesto l'indice j indica

un neurone nascosto.

Differenziando l'equazione 2.35 otteniamo

$$\begin{aligned}\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} &= \sum_{k \in C} e_k \frac{\partial e_k(n)}{\partial y_j(n)} \\ &= \sum_{k \in C} e_k \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)}\end{aligned}\quad (2.36)$$

Dall'equazione 2.16 sappiamo che

$$\begin{aligned}e_k(n) &= d_k(n) - y_k(n) \\ &= d_k(n) - \varphi_k(v_k(n))\end{aligned}\quad (2.37)$$

quindi

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'_k(v_k(n))\quad (2.38)$$

Sappiamo inoltre, dall'equazione 2.24 che, se il neurone j ha q ingressi il suo stato d'attivazione interna vale

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n)\quad (2.39)$$

quindi

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n)\quad (2.40)$$

Quindi, sostituendo i risultati ottenuti in 2.38 e in 2.40 nell'equazione 2.36 otteniamo

$$\begin{aligned}\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} &= - \sum_{k \in C} e_k(n) \varphi'_k(v_k(n)) w_{kj}(n) \\ &= - \sum_{k \in C} \delta_k(n) w_{kj}(n)\end{aligned}\quad (2.41)$$

Infine, sfruttando il risultato di 2.41 in 2.34 otteniamo

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_{k \in C} \delta_k(n) w_{kj}(n)\quad (2.42)$$

che esprime il gradiente locale di un neurone nascosto.

Per quanto riguarda i fattori coinvolti nel calcolo del gradiente locale:

- $\varphi'_j(v_j(n))$ dipende esclusivamente dalla funzione di attivazione del neurone j ;

- i $\delta_k(n)$, con $k \in C$, richiedono la conoscenza dell'errore $e_k(n)$ di tutti i neuroni che fanno parte dello strato immediatamente successivo a quello del neurone j in esame e a cui il neurone j indirizza la sua uscita;
- i termini $w_{kj}(n)$ sono i pesi sinaptici associati alle connessioni tra il neurone j e i neuroni dello strato successivo a cui esso è collegato.

Quindi, riassumendo i passi fondamentali dell'algoritmo di back-propagation:

1. Il nuovo ingresso viene presentato alla rete e si calcola l'errore come differenza tra l'uscita target e l'uscita reale prodotta dalla rete;
2. Si calcola la correzione da applicare ai pesi sinaptici, in base alla formula

$$\begin{pmatrix} \text{Termine} \\ \text{correttivo} \\ \Delta w_{ji}(n) \end{pmatrix} = \begin{pmatrix} \text{learning} \\ \text{rate} \\ \eta \end{pmatrix} \cdot \begin{pmatrix} \text{gradiente} \\ \text{locale} \\ \delta_j(n) \end{pmatrix} \cdot \begin{pmatrix} \text{segnale in ingresso} \\ \text{al neurone } j \\ y_i(n) \end{pmatrix}$$

dove $\delta_j(n)$ assume espressioni diverse se il neurone in esame appartiene allo strato d'uscita (in tal caso è calcolato come in 2.33) o ad uno strato nascosto (in tal caso si utilizza 2.42, ovvero $\delta_j(n)$ è il prodotto della derivata della funzione di attivazione del neurone, e della somma pesata dei gradienti locali dei neuroni dello strato successivo allo strato del neurone considerato e a cui il neurone considerato dirige la sua uscita).

3. Si reitera fino a raggiungere un minimo dell'errore \mathcal{E}_{av} .

Se si applica la modalità *batch learning* l'unica differenza, dal punto di vista delle formule da applicare, è che si minimizza \mathcal{E}_{av} anzichè $\mathcal{E}(n)$, quindi la 2.26 diventa

$$\frac{\partial \mathcal{E}_{av}}{\partial w_{ji}} = \frac{1}{N} \sum_{n=1}^N \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (2.43)$$

e, a parte l'introduzione della sommatoria su tutti gli elementi del training set, si procede esattamente allo stesso modo, ottenendo per il termine correttivo da sommare al peso w_{ji} l'espressione

$$\Delta w_{ji} = \eta \frac{1}{N} \sum_{n=1}^N \delta_j(n) y_i(n) \quad (2.44)$$

Per quanto riguarda il procedimento da seguire in modalità *batch learning*, c'è da ricordare che il valore numerico dei pesi è mantenuto costante durante il calcolo della risposta della rete relativa a tutti gli elementi del training set;

solo successivamente si ricava l'errore medio e, minimizzandolo, si ottengono i termini correttivi da sommare ai pesi. In modalità *pattern training*, invece, l'errore viene calcolato dopo la presentazione di ogni ingresso e i pesi vengono conseguentemente modificati.

Pregi e difetti dell'algoritmo di back-propagation

L'algoritmo di back-propagation è il più utilizzato per il training supervisionato delle reti neurali feedforward. Esso è semplice da applicare e implementa il metodo del gradiente nello spazio dei pesi della rete. I difetti principali di questo metodo di allenamento sono la necessità di avere funzioni di attivazione differenziabili per poterlo applicare, la necessità di conoscere per ogni ingresso il valore target corrispondente e il fatto che la soluzione trovata non è necessariamente un minimo globale della funzione errore. È noto che la capacità del metodo del gradiente di individuare il vero minimo globale di una funzione è fortemente influenzata dal punto in cui inizia la ricerca. Per contro, l'inizializzazione dei pesi della rete, prima di applicare la back-propagation, avviene in maniera random.

Capacità della rete di generalizzare

Si dice che una rete neurale generalizza bene quando l'uscita generata dalla rete in corrispondenza ad un determinato ingresso è corretta (o l'errore è molto piccolo) anche per ingressi non contenuti nel training set. Quando una rete impara durante la fase di training troppo nei dettagli le relazioni ingresso-uscita c'è il rischio che essa memorizzi i dati con cui è stata allenata e perda la capacità di produrre uscite ragionevoli e corrette quando le vengono presentati ingressi simili a quelli già visti. Questo problema può essere risolto utilizzando la tecnica di *cross-validation* per arrestare il training della rete. Ovvero, ad ogni epoca, dopo aver calcolato l'errore e aver modificato i pesi come indicato dall'algoritmo di back-propagation, viene calcolato l'errore commesso dalla rete su un altro insieme di dati, mai visti in fase di training: il *validation set*. Poiché questi dati sono simili ai dati del training set, man mano che procede l'allenamento della rete dovrebbero migliorare sia le performance sul training che quelle sul validation set. All'aumentare del numero di epoche si nota, invece, che l'errore della rete sui dati del validation set, dopo aver raggiunto un punto di minimo, inizia ad aumentare, mentre l'errore sul training set continua a diminuire. Ciò significa che la rete sta perdendo capacità di generalizzare, poiché sta apprendendo anche il rumore dei dati, o caratteristiche insignificanti di essi. A questo punto l'allenamento della rete va fermato e si sceglie la configurazione che aveva dato i risultati migliori sul validation set.

2.4.2 Apprendimento hebbiano

Il postulato di Hebb sull'apprendimento dice che:

- se due neuroni su terminazioni opposte di una sinapsi si attivano simultaneamente la forza di quella sinapsi è accresciuta;
- se due neuroni su terminazioni opposte di una sinapsi si attivano in modo asincrono allora quella sinapsi viene indebolita o eliminata.

Il modello matematico che descrive queste affermazioni porta a modificare il peso w_{kj} all'istante n con il termine correttivo $\Delta w_{kj}(n)$ definito da

$$\Delta w_{kj}(n) = \eta y_k(n)x_j(n) \quad (2.45)$$

dove η è il *parametro di apprendimento*, $x_j(n)$ è l'attività presinaptica, ovvero l'ingresso del neurone k proveniente dal neurone j , mentre $y_k(n)$ è l'attività postsinaptica, ovvero l'uscita del neurone k .

Questa semplice formula potrebbe portare il peso w_{kj} a saturazione. Per evitare questa evenienza si impone un limite alla crescita dei pesi sinaptici, per esempio introducendo un coefficiente d'oblio. La 2.45 diventa

$$\Delta w_{kj}(n) = \eta y_k(n)x_j(n) - \alpha y_k(n)x_kj(n) \quad (2.46)$$

con α costante positiva e $w_{kj}(n)$ valore numerico del peso all'istante n .

2.4.3 Apprendimento competitivo

In questo tipo di apprendimento i neuroni dello strato d'uscita della rete neurale competono per essere l'unico neurone attivo. Quindi, in una rete allenata in questo modo, ad ogni istante un solo neurone produce l'uscita, rendendo questa tecnica particolarmente adatta per reti che funzionano come classificatori.

Nella sua forma più semplice l'apprendimento competitivo è applicato a una rete con un solo strato di neuroni d'uscita, connessi a tutti gli ingressi ed eventualmente connessi lateralmente tra loro, come mostrato in Figura 2.12. Le connessioni laterali hanno la funzione inibire gli altri neuroni d'uscita.

Affinchè il neurone j sia il neurone vincente in corrispondenza ad un certo vettore d'ingresso \mathbf{x} , è necessario che il suo livello di attivazione interna sia superiore a quello di tutti gli altri neuroni. In tal caso l'uscita del neurone j è settata a 1, mentre quella di tutti gli altri neuroni vale 0.

Sia w_{ji} il peso che collega l'ingresso i al neurone j . Tutti i pesi sinaptici sono positivi e vale

$$\sum_i w_{ji} = 1 \quad \forall j \quad (2.47)$$

ovvero la somma totale di tutti i pesi in ingresso ad un neurone vale 1. Un neurone apprende aumentando la forza dei pesi sinaptici che lo collegano

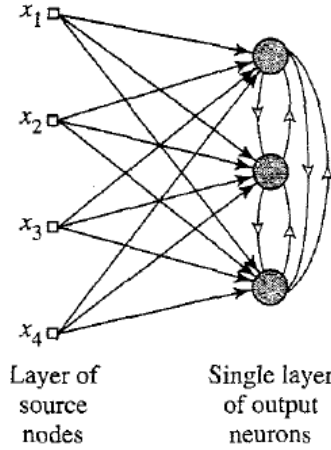


Figura 2.12: Grafo di una rete neurale allenata attraverso apprendimento competitivo. Le sinapsi che collegano gli ingressi alle uscite sono eccitatorie, mentre quelle laterali, tra i vari neuroni d'uscita, sono inibitorie [19].

agli ingressi con cui è attivo e diminuendo il valore dei pesi che lo uniscono agli ingressi con cui è inattivo. La modifica Δw_{ji} applicata al peso w_{ji} vale

$$\Delta w_{ji} = \begin{cases} \eta(x_i - w_{ji}) & \text{se il neurone } j \text{ è vincente} \\ \mathbf{0} & \text{se il neurone } j \text{ è perdente} \end{cases} \quad (2.48)$$

e questo ha l'effetto di muovere il vettore \mathbf{w}_j del neurone vincente verso il vettore d'ingresso \mathbf{x} corrispondente.

Al termine della fase di training ogni neurone d'uscita dovrebbe aver individuato una classe di ingressi, muovendo i suoi pesi sinaptici verso il baricentro di quella particolare classe.

2.4.4 Algoritmi stocastici

Dato che l'allenamento di una rete neurale si riduce alla minimizzazione di una funzione obiettivo che dipende dai pesi della rete, teoricamente si può utilizzare qualsiasi metodo di minimizzazione funzionale. Oltre ai metodi classici, finora descritti, sempre più frequentemente si ricorre ad algoritmi stocastici. Essi hanno il pregio di saper trovare con altissima probabilità il vero minimo globale della funzione obiettivo, e di non richiedere né che la funzione da minimizzare sia derivabile, né che sia continua. Inoltre, sfruttando questi metodi, si può contemporaneamente ottimizzare sia la struttura che i pesi della rete, ovvero si minimizza una funzione obiettivo in cui, oltre ai pesi, anche il numero di strati nascosti e il numero di neuroni in ogni strato sono variabili da settare.

Simulated annealing

Questo algoritmo ha anch'esso come obiettivo la minimizzazione dell'errore commesso dalla rete in fase di training, e per ottenere ciò utilizza un metodo stocastico per migliorare gradualmente le performance del sistema fino a giungere ad una configurazione ottima. L'algoritmo iterativamente seleziona in modo random una sinapsi e apporta un piccolo cambiamento al suo valore numerico. Se questa modifica migliora le performance della rete il cambiamento è mantenuto, in caso contrario il cambiamento è rifiutato. Per evitare che questa procedura si blocchi in un minimo locale, in verità, non tutti i cambiamenti che peggiorano le performance della rete vengono scartati; in una percentuale prestabilita di volte queste modifiche sono comunque accettate [1].

Algoritmi genetici (GA) ed evolutivi (EA)

Gli *algoritmi genetici* (*Genetic Algorithms*, GAs) e gli *algoritmi evolutivi* (*Evolutionary Algorithms*, EAs) sono tecniche di ottimizzazione ispirate ai meccanismi genetici dell'evoluzione della specie e basati sul concetto darwiniano di sopravvivenza del più forte e più dotato. Questi algoritmi iniziano la loro ricerca generando una popolazione di individui random. Ogni individuo è una possibile soluzione del problema. Successivamente i vari individui vengono valutati, (ovvero si calcola la funzione obiettivo nel punto dell'iperspazio che essi rappresentano) e viene generata una nuova popolazione applicando alla precedente operatori di riproduzione (basati sulla sopravvivenza del migliore), mutazione e crossover. La nuova popolazione viene valutata e l'algoritmo prosegue allo stesso modo finché non sono soddisfatti i criteri di terminazione. Questa tecnica si basa fortemente sull'operazione di crossover che, però, risulta poco adeguata se applicata a reti neurali [44]. Infatti, vista la natura distribuita delle reti neurali, quando durante il crossover una rete viene divisa in due parti, ognuna delle quali viene poi ricombinata con due parti di un'altra rete, si rischia di ottenere due nuove reti neurali non funzionali. Metà di una rete neurale non compie metà delle operazioni dell'intera rete neurale, di conseguenza, suddividendola in varie parti, se ne deteriorano irrimediabilmente le proprietà.

In letteratura troviamo svariati esempi in cui questi algoritmi sono stati applicati. In [47] una rete feedforward per predire la richiesta energetica è allenata con un GA, ottenendo performance migliori che con backpropagation. In [46] viene proposto un GA "ibrido", nel quale ad ogni generazione si applica l'algoritmo di backpropagation sull'individuo migliore, per velocizzare l'ottimizzazione, mentre in [21] una FFNN allenata con backpropagation viene trasformata in una rete ricorsiva, aggiungendo nuove sinapsi la cui posizione e il cui peso sono determinate con un algoritmo genetico. In [3] viene utilizzato un algoritmo evolutivo per ottimizzare contemporaneamente

pesi, numero di strati, numero di connessioni e posizione delle connessioni di una rete neurale ricorsiva.

Particle swarm optimization algorithm (PSO)

Il *PSO* è un algoritmo stocastico particolarmente adatto all'allenamento delle reti neurali, come dimostra il suo ampio utilizzo in letteratura. Esso viene utilizzato, per esempio, in [28], per risolvere un problema in cui l'algoritmo di backpropagation convergeva troppo lentamente e rimaneva spesso bloccato in minimi locali, mentre in [40] si ricorre a questo algoritmo per allenare la rete minimizzando una funzione non differenziabile. Il PSO proposto in [23], invece, ottimizza contemporaneamente sia i pesi che la struttura della rete.

I principali vantaggi, rispetto agli algoritmi genetici, sono il minor numero di parametri liberi da settare empiricamente, l'assenza dell'operatore di crossover e il fatto che viene introdotto il concetto di velocità nello spazio delle variabili: in questo modo la ricerca della soluzione ottima, pur muovendosi stocasticamente, ha una direzione ben definita.

Questi algoritmi generano inizialmente uno sciame di particelle distribuite casualmente nello spazio di ricerca. Ogni particella rappresenta una possibile soluzione del problema in questione ed è caratterizzata da una posizione e da una velocità. Per ogni elemento dello sciame viene calcolato il fitness (ovvero il valore della funzione obiettivo sul punto in cui la particella è posizionata nell'iperspazio delle variabili), quindi vengono modificate la velocità di ogni particella e, coerentemente, la sua posizione. La nuova velocità è determinata in modo da dirigere la particella verso la sua miglior posizione passata finora occupata, ma anche verso la miglior posizione trovata da un altro membro dello sciame. L'algoritmo si arresta quando lo sciame di particelle converge ad un punto di ottimo della funzione.

Covariance matrix adaptation-evolution strategy (CMA-ES)

Di particolare interesse è anche la *Covariance Matrix Adaptation-Evolution Strategy* (CMA-ES), una strategia evolutiva per l'ottimizzazione di funzioni non lineari e non convesse, ed eventualmente non derivabili.

Ad ogni passo l'algoritmo genera una popolazione di possibili candidate soluzioni, campionandole da una distribuzione gaussiana multivariata, quindi calcola il fitness di ogni individuo (ovvero il valore della funzione obiettivo nel punto che esso rappresenta). Prima di creare la generazione successiva la media della distribuzione da cui la popolazione stessa sarà campionata viene modificata in modo da massimizzare la probabilità di riottenere gli individui più dotati della generazione attuale. La matrice di covarianza della distribuzione è modificata in modo da massimizzare la probabilità di muoversi in direzioni dell'iperspazio che erano risultate precedentemente

molto buone. Si tiene inoltre memoria dei risultati ottenuti in due passi consecutivi dell'algoritmo: se passi consecutivi sono nella stessa direzione si aumenta la lunghezza del passo stesso, altrimenti la si tiene invariata. Le modifiche apportate alla matrice di covarianza mirano a concentrare i nuovi individui nelle direzioni dello spazio in cui sono stati individuati un maggior numero di candidate soluzioni buone, il controllo della lunghezza del passo ha lo scopo sia di prevenire la convergenza prematura, sia di velocizzare la convergenza stessa verso il punto di ottimo [22].

2.5 Predizione di serie temporali con reti neurali

Uno dei principali campi di applicazione delle reti neurali è la predizione di serie temporali. Ciò che le rende così attraenti è la loro capacità di apprendere da esempi la relazione che lega ingresso e uscita, nonché il fatto che liberano dalla necessità di definire un modello che descriva i dati, che, per essere costruito, necessita di assunzioni a priori sul processo da analizzare. Un'altra proprietà apprezzabile è la loro capacità di generalizzare e produrre risultati corretti anche se in ingresso hanno dati mai visti in precedenza e parzialmente corrotti da rumore. Infine le reti neurali sono, per loro natura, non lineari (basta utilizzare funzioni d'attivazione non lineari) e, quindi, particolarmente adatte a descrivere le componenti non lineari delle serie temporali da predire.

Oltre ai molti pregi, anche le reti neurali presentano dei limiti. Uno dei maggiori problemi è l'assenza di risultati teorici sulla struttura ottima della rete da utilizzare. Di conseguenza è necessario procedere a tentativi, testando varie configurazioni e scegliendo la migliore. Va poi sottolineato che, nel caso di processi quasi lineari, spesso i metodi di predizione classici producono risultati migliori o, per lo meno non peggiori, richiedendo meno dati e meno tempo per la fase di training [49].

In questa tesi sono state utilizzate reti neurali feedforward per predire la glicemia futura, con in ingresso misure glicemiche passate quasi continue, registrate attraverso sensori CGM, e informazioni sui pasti. Nel prossimo capitolo verrà affrontato dettagliatamente il problema specifico della predizione della serie glicemica e verranno illustrate alcune soluzioni proposte in letteratura.

Capitolo 3

Predizione della glicemia futura da dati CGM: stato dell'arte

Lo sviluppo dei dispositivi CGM a partire dall'inizio del XXI secolo ha reso possibile un notevole miglioramento della cura del diabete, poichè l'informazione continua sul livello glicemico può essere utilizzata per aggiustare adeguatamente la terapia, nonchè per generare allarmi quando si manifestano episodi rischiosi, ovvero quando la glicemia oltrepassa le soglie di ipo o iperglicemia. Come già introdotto in Sezione 1.5.5 (pag 18) sarebbe però più utile cercare di prevenire tali episodi, predicendoli con un adeguato anticipo.

Per predire una serie temporale servono la storia passata del segnale, eventualmente informazioni su altre grandezze che ne influenzano il valore futuro e, nella maggior parte dei casi, un modello del segnale. Inoltre la serie in esame deve essere predicibile, ovvero deve essere presente correlazione tra campioni successivi del segnale.

Il segnale glicemico è predicibile, poichè risulta sufficientemente stazionario lontano da pasti e medicazioni insuliniche, e poichè in esso è presente una certa correlazione [6]. Alcuni dispositivi, tra i quali anche il FreeStyle NavigatorTM, sono già dotati di un semplice algoritmo di proiezione che, tenendo conto del trend degli ultimi dati registrati, genera un allarme sulla base del valore di glucosio predetto. Il problema di questi algoritmi proiettivi è la loro scarsa precisione e, quindi, l'alto numero di falsi allarmi che generano. Sarebbe, quindi, preferibile utilizzare metodi di predizione più raffinati, in grado di fornire, con sufficiente anticipo, ma anche precisione, il valore glicemico futuro, così da poter intervenire in tempo e prevenire gli episodi di ipo e iperglicemia, nonchè limitare tutti i danni a breve e lungo termine ad essi connessi.

Nei prossimi paragrafi verranno brevemente illustrati alcuni lavori finora

compiuti nell'ambito della predizione del livello glicemico e verrà accennata la teoria su cui si basano.

3.1 Filtro di Kalman

Il filtro di Kalman è un filtro causale ottimo, a minima varianza d'errore, che imposta il problema di filtraggio come un problema di stima dello stato del sistema, ovvero di un vettore che ne racchiude tutta la storia passata.

Si fa riferimento ad una situazione in cui segnali incogniti e misure possono essere messi in relazione mediante un modello di stato nella forma

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) & \text{Equazione di aggiornamento dello stato} \\ \mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t) & \text{Equazione delle misure} \end{cases} \quad (3.1)$$

Dove:

\mathbf{x} è il vettore di stato di dimensione $N \times 1$;

\mathbf{F} è una matrice quadrata $N \times N$, detta *matrice di transizione di stato*;

\mathbf{y} è il *vettore delle misure*, di dimensione $M \times 1$

\mathbf{H} è una matrice rettangolare $M \times N$ detta *matrice di trasferimento stato-uscita*.

I vettori \mathbf{w} e \mathbf{v} sono aleatori, tra loro scorrelati, a media nulla e con matrici di covarianza $\text{cov}(\mathbf{w}) = \mathbf{Q}$ e $\text{cov}(\mathbf{v}) = \mathbf{R}$.

Impostando formalmente il problema: supponiamo di disporre di $\hat{\mathbf{x}}(t|t)$, ovvero la *stima dello stato* in t , originata dalla conoscenza dei vettori di misura $\mathbf{y}(t), \mathbf{y}(t-1), \dots, \mathbf{y}(1)$, ovvero di tutti i dati, fino al tempo t compreso. Supponiamo di conoscere anche l'*errore di stima*, che è a media nulla e matrice di covarianza $\mathbf{P}(t|t)$. L'obiettivo è stimare lo stato e l'errore di stima al tempo $t+1$.

1. Ad ogni istante t si fa un primo *passo predittivo* al tempo $t+1$, senza usare le misure al tempo $t+1$. Imponendo $\mathbf{w}(t) = \mathbf{0}$ si ottiene

$$\hat{\mathbf{x}}(t+1|t) = \mathbf{F}\hat{\mathbf{x}}(t|t) \quad (3.2)$$

$$\mathbf{P}(t+1|t) = \mathbf{F}\mathbf{P}(t|t)\mathbf{F}^T + \mathbf{Q} \quad (3.3)$$

2. Appena sono note le misure al tempo $t+1$ si esegue un *passo correttivo*, in cui si aggiornano la stima dello stato e della matrice di covarianza dell'errore, utilizzando la nuova misura e la stima precedente come prior.

$$\begin{aligned} \hat{\mathbf{x}}(t+1|t+1) &= \hat{\mathbf{x}}(t+1|t) + \\ &+ \mathbf{P}(t+1|t)\mathbf{H}^T[\mathbf{H}\mathbf{P}(t+1|t)\mathbf{H}^T + \mathbf{R}]^{-1} \cdot \\ &\cdot [\mathbf{y}(t+1) - \mathbf{H}\hat{\mathbf{x}}(t+1|t)] \end{aligned} \quad (3.4)$$

$$\begin{aligned} \mathbf{P}(t+1) &= \mathbf{P}(t+1|t) - \mathbf{P}(t+1|t)\mathbf{H}^T \cdot \\ &\cdot [\mathbf{H}\mathbf{P}(t+1|t)\mathbf{H}^T + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}(t+1|t) \end{aligned} \quad (3.5)$$

Reiterando la 3.2 e la 3.3 PH volte (con PH Prediction Horizon, ovvero orizzonte di predizione) si ottiene la predizione PH passi in avanti. In particolare si ha

$$\hat{\mathbf{x}}(t+PH|t) = \mathbf{F}\hat{\mathbf{x}}(t|t) \quad (3.6)$$

$$\mathbf{P}(t+PH|t) = \mathbf{F}^{PH}\mathbf{P}(t)(\mathbf{F}^T)^{PH}\mathbf{Q} \quad (3.7)$$

L'utilizzo del filtro di Kalman per la predizione della serie glicemica è stato suggerito da Palerm *et al.* in [30] e in [29]. Nel primo lavoro, che risale al 2005, viene utilizzato un dataset simulato (l'andamento del glucosio interstiziale viene modellato con delle sinusoidi), mentre il secondo lavoro, del 2007, utilizza dati provenienti da clamp ipoglicemico di 13 soggetti diabetici e può essere considerato un raffinamento di quanto svolto in precedenza. L'algoritmo elaborato stima le derivate prima e seconda del segnale glicemico e le utilizza per predire le ipoglicemie. Non vengono però quantificati i

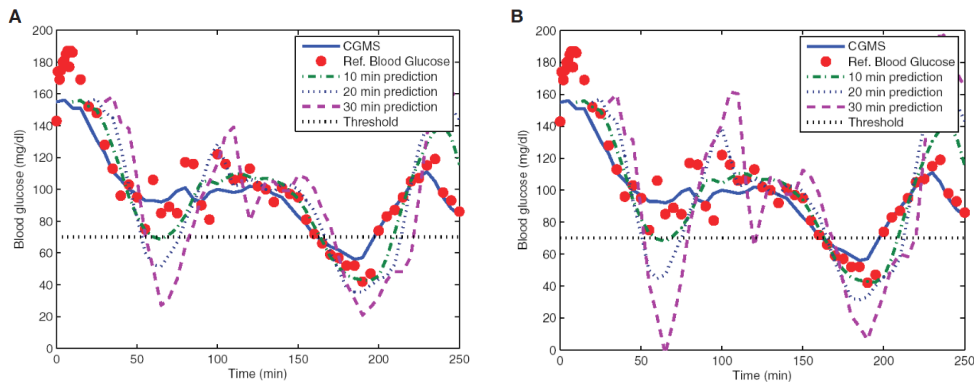


Figura 3.1: Predizione ottenuta col modello proposto da Palerm *et al.* in [29]. Il profilo di sinistra è ottenuto imponendo le derivate seconde del segnale pari a 0. Il profilo di destra è ottenuto tenendo conto anche delle derivate seconde del segnale [29].

risultati in termini di RMSE (Root Mean Square Error) e di ritardo della predizione, in quanto il fine ultimo del lavoro era l'indagine degli effetti della frequenza di campionamento, dei livelli di soglia ipoglicemica e dell'orizzonte di predizione sulla sensibilità (percentuale di ipoglicemie correttamente predette) e sulla specificità (numero di non-ipoglicemie correttamente predette) della predizione.

3.2 Modello autoregressivo del primo ordine (AR(1))

Un modello autoregressivo di ordine p è descritto dalla relazione ingresso-uscita

$$\mathbf{y}_n = - \sum_{k=1}^p \mathbf{a}_k \mathbf{y}_{n-k} + \mathbf{v}_n \quad (3.8)$$

ovvero il dato all'istante n è una combinazione lineare dei dati nei p istanti precedenti, opportunamente pesati dai coefficienti \mathbf{a}_k , e di un rumore bianco, \mathbf{u}_n . In forma matriciale la 3.8 diventa

$$\mathbf{y}_n = [\mathbf{y}_{n-1}, \mathbf{y}_{n-2}, \dots, \mathbf{y}_{n-p}] \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_p \end{bmatrix} + \mathbf{v}_n \quad (3.9)$$

e, in forma compatta

$$\mathbf{y}_n = \Psi_n^T \mathbf{a} + \mathbf{v}_n \quad (3.10)$$

Per un modello del primo ordine $p = 1$ e l'equazione che descrive l'uscita in relazione agli ingressi è semplicemente

$$\mathbf{y}_n = -\mathbf{a}_1 \mathbf{y}_{n-1} + \mathbf{v}_n \quad (3.11)$$

Per seguire i cambiamenti del segnale è opportuno utilizzare un modello tempo variante, quindi il modello viene ri-identificato ogni volta che arriva una nuova misura. Una delle tecniche di identificazione più utilizzate è quella dei minimi quadrati pesati ricorsivi (RWLS, Recursive Weighted Least Square), implementata dalle seguenti equazioni:

$$\mathbf{P}_{n+1} = \frac{1}{\mu} \left[\mathbf{P}_n - \frac{\mathbf{P}_n \Psi_{n+1} \Psi_{n+1}^T \mathbf{P}_n}{\mu + \Psi_{n+1}^T \mathbf{P}_n \Psi_{n+1}} \right] \quad (3.12)$$

$$\mathbf{k}_{n+1} = \frac{\mathbf{P}_n \Psi_{n+1}}{\mu + \Psi_{n+1}^T \mathbf{P}_n \Psi_{n+1}} \quad (3.13)$$

$$\mathbf{e}(n+1) = \mathbf{y}(n+1) - \Psi_{n+1}^T \hat{\mathbf{a}}_n \quad (3.14)$$

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n + \mathbf{e}(n+1) \mathbf{k}_{n+1} \quad (3.15)$$

con \mathbf{a} vettore dei coefficienti del modello AR, Ψ vettore dei p campioni precedenti a quello attuale, \mathbf{P} matrice di covarianza dei dati, \mathbf{k} vettore del guadagno utilizzato per correggere la stima fatta al passo precedente, \mathbf{e} è l'errore di stima nella predizione ad un passo e $\mu \in [0 \ 1]$ è il forgetting factor, o coefficiente di oblio. In particolare, un dato relativo a q istanti precedenti quello attuale avrà peso μ^q . Il valore di questo parametro ha un

ruolo fondamentale. Un μ troppo piccolo dà scarsa importanza al passato, portando ad un modello tempestivo, ma molto sensibile al rumore e, quindi, instabile. Al contrario, μ troppo grande porta a dare troppa importanza al passato e il sistema risulterà troppo lento a cogliere i cambiamenti del segnale.

Per il modello AR(1) tutte le quantità vettoriali delle equazioni 3.12, 3.13, 3.14 e 3.15 si riducono a scalari.

Dopo aver identificato il modello, per effettuare la predizione PH passi in avanti basta applicare PH volte l'equazione 3.8, alimentandola con rumore bianco. Per il modello AR(1) si ottiene

$$y_{n+PH} = a_n^{PH} y_n \quad (3.16)$$

Un modello autoregressivo del primo ordine, con parametri tempo varianti stimati ad ogni passo è utilizzato da Sparacino *et al.* in [39]. Il modello viene identificato utilizzando dati CGM prefiltrati con filtro passa basso per eliminare gli ampi spikes saltuariamente presenti nelle serie temporali. In questo lavoro vengono utilizzati vari forgetting factor e si analizza la bontà del modello in termini di RMSE e di ritardo della predizione. È da sottolineare che questo lavoro è uno dei pochi in cui vengono esplicitamente utilizzati i ritardi per valutare il predittore.

Un modello autoregressivo viene utilizzato anche da Reifman *et al.* in [36]. L'approccio di questo lavoro è profondamente diverso da [39], infatti in questo caso si utilizza un modello di ordine 10, tempo invariante e soggetto invariante. Il modello viene identificato una sola volta, su dati CGM, utilizzando i minimi quadrati regolarizzati. I risultati sono analizzati in termini di RMSE e utilizzando la Clark Error Grid Analysis. Non vengono fornite informazioni sui ritardi.

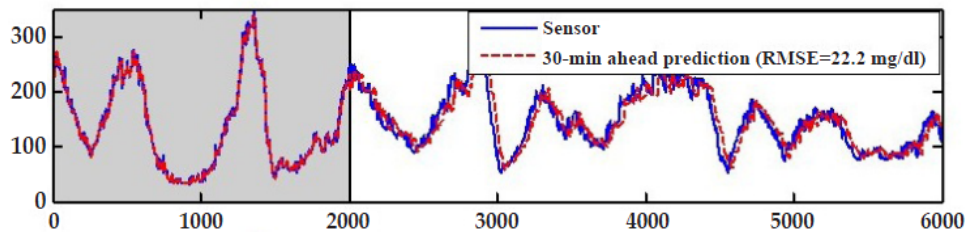


Figura 3.2: Predizione ottenuta col modello proposto da Reifman *et al.* in [36], identificato e testato su dati dello stesso soggetto [36].

In Gani *et al.*, in [17], viene utilizzato un modello AR(30) tempoinvariante per la predizione. I dati CGM usati per stimare i parametri sono prefiltrati, anche in questo caso, come in [36] vengono utilizzati i minimi quadrati regolarizzati. Erano stati analizzati anche altri due scenari, nel primo si utilizzavano dati grezzi, nel secondo dati prefiltrati, e in entrambi

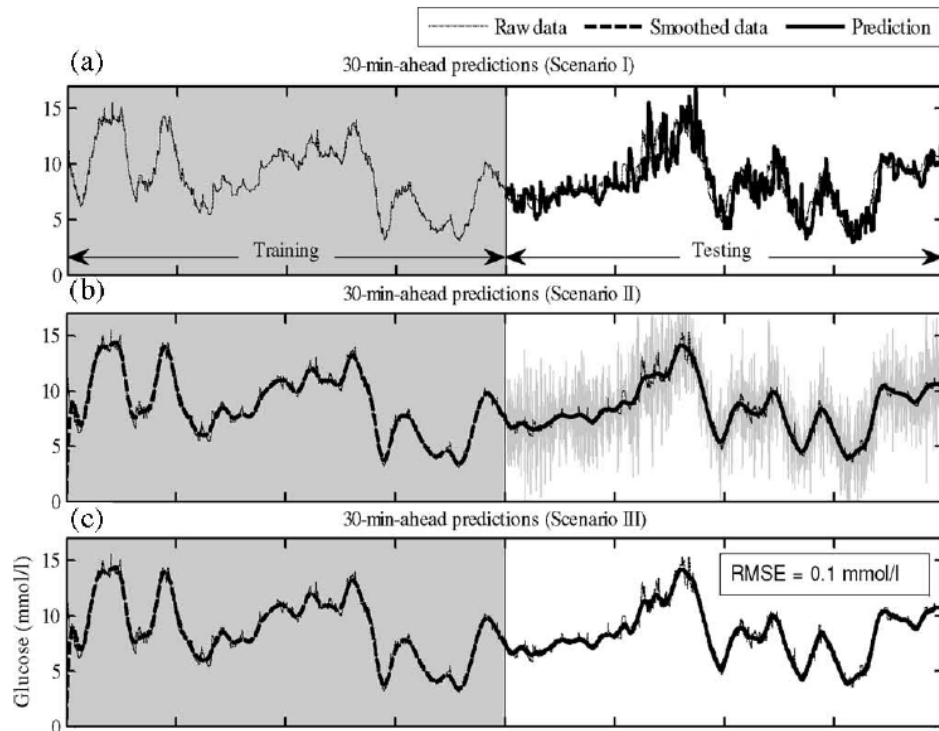


Figura 3.3: Profili predetti dal modello proposto da Gani *et al.* in [17]. L'orizzonte di predizione è pari a 30 minuti. In Figura (a) (Scenario I) vengono utilizzati dati grezzi. In Figura (b) (Scenario II) vengono utilizzati dati prefiltrati. In Figura (c) (Scenario III) vengono utilizzati dati prefiltrati e i parametri del modello sono regolarizzati [17].

i parametri del modello non venivano regolarizzati. I risultati, però, non erano fisiologicamente accettabili. Anche in questo caso nella discussione dei risultati non vengono considerati i ritardi.

Nel lavoro di Finan *et al.* [16] la predizione viene eseguita con un modello ARX, ovvero un modello autoregressivo con ingresso esogeno, che in questo caso è rappresentato dai valori dei carboidrati ingeriti e delle dosi di insulina somministrata. Il modello utilizzato è del terzo ordine e ne vengono proposte due versioni: una tempo invariante e l'altra tempo variante, identificata ricorsivamente ad ogni passo. Dai risultati il modello tempo variante non sembra superiore di quello tempo invariante. Le metriche utilizzate per esprimere la bontà dei risultati sono il FIT dei dati e il RMSE.

Il grande successo dei modelli autoregressivi per la predizione a breve termine del segnale glicemico è legata alla lentezza d'azione dei pasti ingeriti e delle dosi di insulina somministrate sulla glicemia. Proprio questo ritardo permette ai modelli autoregressivi di usare con successo la correlazione tra i campioni glicemici, precedentemente imparata, per inferire sul valore futuro.

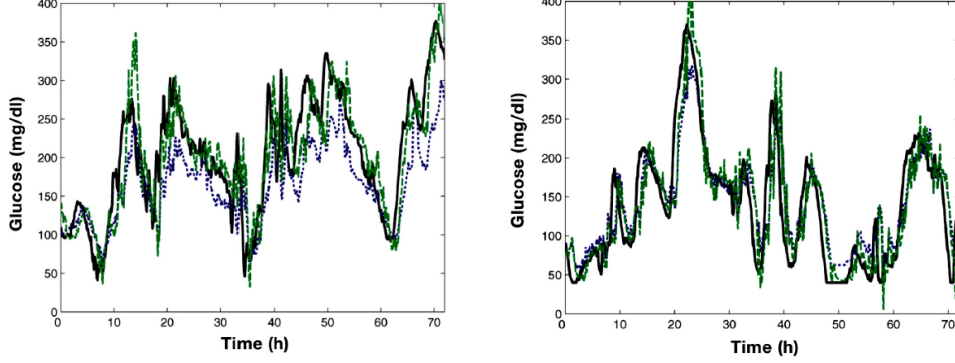


Figura 3.4: Performance dei modelli ARX proposti da Finan *et al.* in [16]. L'orizzonte di predizione è pari a 60 minuti. La linea nera è relativa al profilo glicemico misurato dal sensore, la linea blu punteggiata è relativa alla predizione del modello ARX con parametri tempo invariati, la linea verde tratteggiata è relativa alla predizione ottenuta col modello ARX con parametri tempo varianti, identificati ricorsivamente ad ogni passo. (Tratto da [16]).

3.3 Modello polinomiale di primo grado

Il modello polinomiale di grado m è descritto dall'equazione

$$\mathbf{y}(t) = a_1 t^{m-1} + a_2 t^{m-2} + \dots + a_{m-1} t + a_m \quad (3.17)$$

Per il modello lineare $m = 1$ e il modello equivale alla retta di regressione

$$\mathbf{y}(t) = a_1 t + a_2 \quad (3.18)$$

Anche in questo caso, come per i modelli autoregressivi, è opportuno ristimare i parametri ad ogni istante temporale, per permettere al predittore di adattarsi ai cambiamenti dei dati. Per identificare ricorsivamente il modello, partendo dalla stima effettuata all'istante precedente, si possono utilizzare le stesse equazioni del modello AR, ovvero la 3.12, la 3.13, la 3.14 e la 3.15, modificando solamente il vettore Ψ , che in questo caso sarà

$$\Psi_n = [t_n^m, t_n^{m-1}, \dots, 1]^T \quad (3.19)$$

Per il modello lineare $\Psi_n = [t_n, 1]^T$ e $\mathbf{a} = [a_1, a_2]$, con a_1 coefficiente angolare e a_2 termine noto della retta di regressione.

Dopo aver identificato il modello per effettuare la predizione PH passi in avanti è sufficiente calcolare il valore della retta in quell'istante temporale, ovvero

$$\mathbf{y}(t_{n+PH}) = a_1 t_{n+PH} + a_2 \quad (3.20)$$

CAPITOLO 3. Predizione della glicemia futura da dati CGM: stato dell'arte

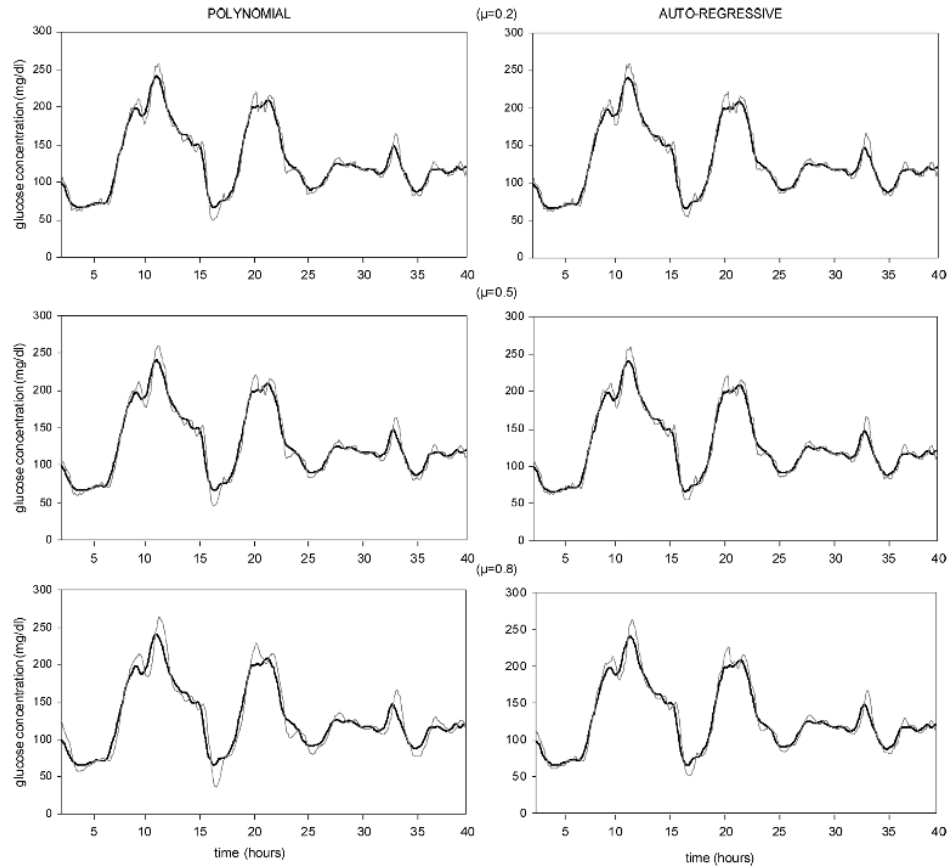


Figura 3.5: Predizione ottenuta con modello $AR(1)$ e con modello lineare, entrambi con parametri tempo varianti, proposti da Sparacino *et al.* in [39]. L'orizzonte di predizione è di 30 minuti. La linea più spessa rappresenta il profilo glicemico misurato dal sensore, la linea sottile è relativa al profilo predetto. Vengono riportati i risultati ottenuti con vari forgetting factors. (Figura tratta da [39]).

Nel lavoro di Sparacino *et al.* [39], oltre al modello $AR(1)$ viene utilizzato anche un modello lineare tempo variante, identificato ricorsivamente con la tecnica dei minimi quadrati pesati, per la predizione del livello glicemico. Le performance di questo modello, in termini di RMSE e ritardo della predizione risultano confrontabili con quelle del modello $AR(1)$ tempo variante precedentemente descritto.

3.4 Reti neurali con ingressi SMBG

La predizione del livello glicemico ha sempre destato interesse anche prima dell'avvento dei dispositivi CGM. Sono presenti in letteratura vari lavori relativi all'utilizzo di reti neurali per inferire la glicemia futura. Queste tecniche sono basate su misurazioni glicemiche di tipo SMBG, che vengono comunemente acquisite in un numero pari a 3-4 campioni al giorno. Di conseguenza l'approccio adottato è necessariamente diverso da quello che si potrebbe applicare a dati CGM. Per poter allenare la rete neurale l'orizzonte di predizione deve essere almeno pari al periodo di campionamento, che nel caso di misurazioni SMBG risulta essere di almeno 5 ore. Con dati CGM, invece, la frequenza con cui vengono effettuate le misurazioni è sufficientemente elevata da poter utilizzare un orizzonte di predizione di 30 o addirittura 15 minuti. Ciò nonostante verrà comunque presentata, in ordine cronologico, una breve panoramica di alcuni lavori recenti¹ presenti in letteratura.

3.4.1 Reti neurali ricorsive ottimizzate e testate su dati relativi ad un solo soggetto

Nel lavoro di Sandham *et al.* [37], che risale al 1998, viene utilizzata una rete neurale ricorsiva, ottimizzata su dati specifici del paziente, per predire il livello glicemico della successiva misura SMBG. Gli ingressi della rete sono:

- il vettore del *BGL* (ovvero il vettore contenente il livello di glucosio ematico, dall'inglese Blood Glucose Level), contenente l'ultimo valore glicemico misurato, l'ora della misurazione e l'ora a cui si vuole effettuare la predizione;
- il vettore dell'*insulina*, contenente il numero di unità di insulina ad azione rapida, intermedia, lenta, l'ora d'iniezione e il sito²;
- il vettore dei *pasti*, contenente i grammi totali di carboidrati ingeriti e l'ora del pasto;
- il vettore dell'esercizio fisico, contenente durata, intensità e orario a cui è stato effettuato;
- un vettore indicante la presenza di stress o malattie.

Da una procedura di tipo trial-and-error la struttura ottima della rete era risultata composta da un'unica strato nascosto di 95 neuroni. L'allenamento della rete è effettuato per un numero prestabilito di epoche (500). I risultati

¹Verranno presentati lavori relativi all'ultimo decennio.

²Ad ogni possibile sito d'iniezione viene associato un numero naturale che costituirà l'ingresso della rete.

sono valutati in base all'errore tra la predizione della rete e il valore di glucosio effettivamente registrato e sono definiti soddisfacenti. La rete, però, è stata testata solamente su 2 soggetti, di cui, tra l'altro, si possedevano pochi dati (122 misurazioni, di cui 97 sono state utilizzate per il training, e 25 per testare la rete).

3.4.2 Reti neurali per modellare il metabolismo di un diabetico: confronto delle performance di varie architetture

Un lavoro particolarmente articolato è quello di Tresp *et al.* [42], in cui vengono valutate varie architetture e se ne confrontano le performance tra di loro e con un modello compartimentale. I dati utilizzati sono relativi ad un unico soggetto, monitorato per 63 giorni (con una media di 7-8 misure SMBG al giorno). Oltre alla glicemia vengono registrati anche: l'ora e la dose di insulina somministrata (distinta in lenta e rapida), l'ora e la quantità di carboidrati ingeriti (differenziati in carboidrati ad assorbimento rapido, intermedio e lento), e l'ora e la durata dell'esercizio fisico (classificato come moderato o intenso). I dati vengono registrati con una risoluzione temporale di 15 minuti, ponendoli uguali a 0 quando assenti (il dataset risulta, quindi, composto prevalentemente da 0). Di seguito verranno elencate le varie architetture proposte per modellare il sistema glucosio-insulina e predire il livello glicemico futuro.

- Rete Neurale Ricorsiva. La rete ha 5 ingressi: insulina, carboidrati ingeriti, esercizio fisico e le stime dei livelli glicemici attuale e precedente. In particolare, per quanto riguarda insulina, carboidrati ed esercizio fisico, vengono utilizzati non gli ingressi grezzi, bensì la curva che descrive l'andamento del loro effetto nel tempo (ovvero la loro *rate of appearance*), campionata all'istante desiderato. L'uscita della rete è il valore glicemico futuro, ovvero con un PH di 15 minuti. Il problema principale in questo approccio è il fatto che le misure glicemiche sono registrate ad intervalli di circa 3 ore, quindi non è possibile confrontare ogni predizione della rete col reale livello di glucosio nel sangue. Inoltre, dato che la rete riceve un vettore di ingresso ogni 15 minuti, al posto del livello di glucosio attuale e precedente vengono utilizzate le ultime due predizioni (questo feedback determina la natura ricorsiva della rete).

Questa architettura è proposta in due varianti: nella prima i livelli glicemici precedenti in ingresso alla rete sono sempre le ultime due predizioni della rete stessa, mentre nella seconda, appena è disponibile una misura glicemica reale, la si usa al posto del valore predetto. Nel secondo caso si ottengono risultati decisamente migliori.

- Rete "di convoluzione". La rete in questo caso ha gli stessi ingressi della precedente, ma non è ricorsiva ed è strutturata con un approccio

diverso. Non si calcola la rate of appearance delle varie quantità, ma si integra ogni valore in una finestra temporale di ampiezza adeguata alla lunghezza dell'effetto della grandezza in esame. Il valore integrato è quindi utilizzato come valore effettivo in ingresso alla rete.

Di questa architettura sono proposte 3 varianti: nella prima le finestre d'integrazione sono rettangolari, disgiunte e con centri e ampiezze precedentemente determinati; nella seconda le finestre sono gaussiane e parzialmente sovrapposte, con centri e ampiezze già determinati; la terza variante differisce dalla seconda in quanto centri e ampiezze delle finestre vengono ottimizzati, assieme ai pesi, durante l'allenamento della rete.

Le stessa rete, nelle 3 versioni appena descritte, è proposta anche con 2 ulteriori ingressi: l'ultimo livello di glucosio misurato (non integrato in una finestra temporale), e il tempo trascorso dalla misura stessa. Questa architettura è risultata migliore della precedente, in termini di qualità della predizione.

I risultati sono forniti in termini di varianza del segnale glicemico che il predittore è in grado di modellare e gli autori affermano che tutti questi modelli non lineari producono risultati migliori dei loro analoghi modelli lineari.

3.4.3 Reti neurali per il controllo della glicemia di pazienti diabetici di tipo 1

Al 2005 risalgono due lavori, uno di El-Jabali [10] e l'altro di Zitar e El-Jabali [50]. In [10] viene proposto un sistema di controllo per diabetici di tipo 1. Tale sistema è composto da 3 reti neurali feedforward, una per la predizione del livello glicemico futuro (a 4 o più ore dall'ultima misurazione effettuata), le altre due per la stima ottima delle dosi di insulina ad azione rapida e ad azione media da somministrare al paziente per mantenere la glicemia futura nel range euglicemico. Gli ingressi della rete per la predizione sono: l'ultimo valore di glicemia misurato, le dosi di insulina rapida ed intermedia somministrate, la quantità di carboidrati ingeriti e il tempo a cui si vuole ottenere la predizione della rete, calcolato come tempo relativo rispetto all'ultima misura effettuata. Informazioni su esercizio fisico e insulina lenta sono state eliminate poichè effettuando un'*analisi delle componenti principali* (PCA) esse risultavano irrilevanti. La rete è stata allenata su dati di 70 pazienti, le performance sono valutate in termini di errore quadratico medio (Mean Square Error, MSE) e sono definite buone.

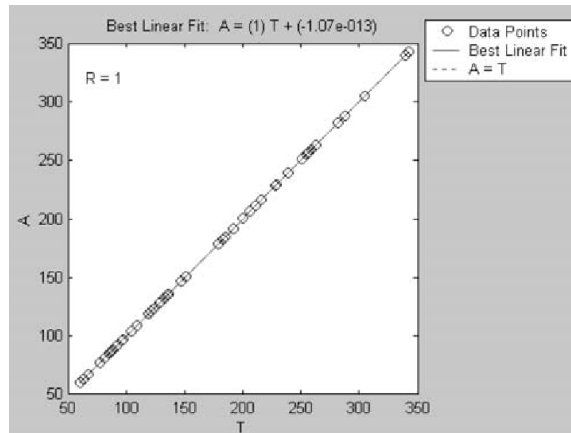


Figura 3.6: Performance della rete neurale per la predizione della glicemia futura proposta da El-Jabali in [10]. Retta di regressione tra uscita della rete neurale e valore glicemico successivo misurato [10].

3.4.4 Sviluppo e confronto di reti neurali per modellare il metabolismo di un paziente diabetico di tipo 2

In [50] vengono invece testate una rete feedforward e una rete polinomiale per predire il livello glicemico in pazienti diabetici di tipo 2. La rete polinomiale si è dimostrata totalmente inadatta alla predizione, mentre i risultati ottenuti con la rete feedforward erano incoraggianti, come dimostra la regressione tra uscita predetta e target, riportata in Figura 3.7. Anche in

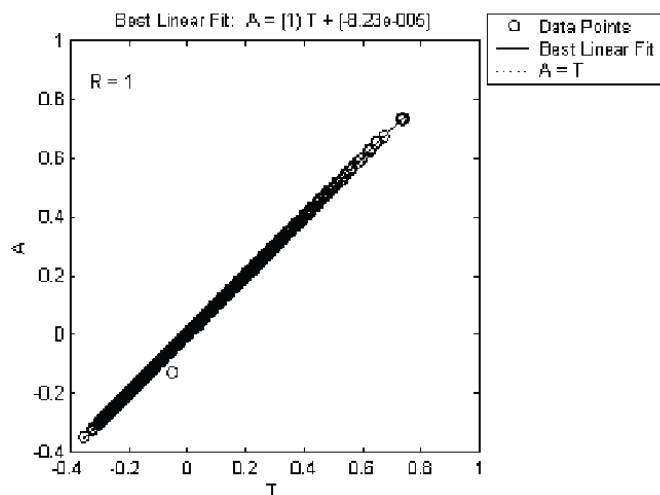


Figura 3.7: Performance della rete neurale proposta da Zitar et al. in [50]. Regressione tra uscita predetta e livello glicemico futuro misurato [50].

questo caso viene effettuata una PCA per individuare gli ingressi significativi, che risultano essere l'ultimo livello glicemico misurato, le dosi di insulina

rapida e media iniettata, i carboidrati ingeriti e il tempo relativo tra l'ultima misura effettuata e l'istante in cui si vuole predire il segnale glicemico.

3.4.5 Reti neurali distinte per modellare la variabilità circadiana del segnale glicemico

Particolarmente interessante risulta il report di Kok, che riassume il lavoro relativo ad un assegno di ricerca, svolto come parte del suo Master of Science [24]. L'obiettivo era determinare se fosse possibile utilizzare una rete neurale per predire accuratamente il livello glicemico. I dati utilizzati sono relativi ad un unico paziente (l'autore stesso) e coprono un periodo di 77 giorni consecutivi. Per ogni giorno vengono registrate 4 misure glicemiche (in corrispondenza ai pasti e prima di coricarsi), la quantità di carboidrati ingeriti durante ogni pasto, le dosi di insulina rapida e lenta somministrate, l'intensità dell'esercizio fisico e il livello di stress. Poichè in diversi momenti della giornata il livello glicemico risente in modo diverso di uno stesso cambiamento di una delle grandezze misurate³, l'autore utilizza 4 reti neurali feedforward diverse⁴, ottimizzate indipendentemente l'una dall'altra per quanto riguarda numero e tipo di ingressi e numero di strati e di neuroni nascosti. Le reti sono tutte allenate con back propagation e cross validation. Le performance sono calcolate su un test set indipendente e vengono forniti i risultati in termini di RMSE, coefficiente di correlazione, pendenza e intercetta della retta di regressione tra target e predizione effettiva della rete. I risultati in termini di RMSE sono confrontabili con quelli riportati in letteratura.

Il lavoro di Baghdadi *et al.* [4] può essere considerato un proseguimento del lavoro di Kok [24], in quanto utilizza lo stesso dataset e allena 4 reti neurali distinte per i vari intervalli della giornata. Le reti utilizzate in questo caso sono però di tipo radial basis function. I risultati ottenuti sono migliori di quelli registrati con le reti feedforward.

Un'ulteriore affinamento dei due lavori appena riassunti è riportato da Zainuddin *et al.* in [48]. Le principali differenze, rispetto a [24] e a [4] sono l'utilizzo della PCA per selezionare gli ingressi veramente significativi e l'utilizzo di reti neurali di tipo wavelet⁵ per la predizione. Gli autori registrano un ulteriore miglioramento dei risultati. In Tabella 3.1 sono riassunti i risultati ottenuti da Kok in [24], da Baghdadi *et al.* in [4] e da Zainuddin *et al.* in [48] in termini di RMSE delle varie reti neurali proposte.

³Per esempio è noto che la sensibilità insulinica varia nell'arco della giornata.

⁴Vengono utilizzate una rete per i valori registrati in mattinata, una per il pomeriggio, un'altra per la sera e un'ultima per la notte. Ogni rete ha in ingresso i valori relativi all'intervallo della giornata a cui essa si riferisce, mentre il target è il livello glicemico misurato nell'intervallo della giornata successivo.

⁵In una wavelet neural network le funzioni d'attivazione dei neuroni degli strati nascosti sono delle wavelet.

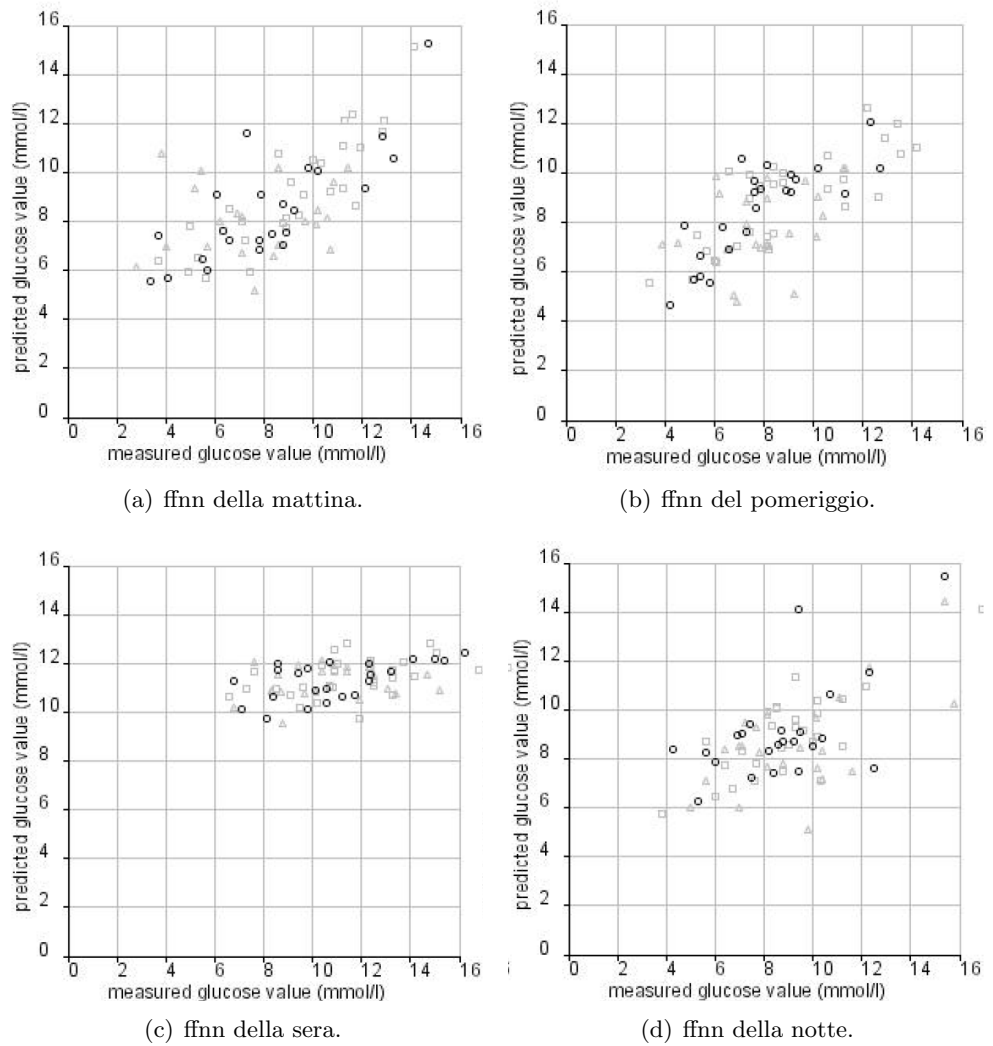


Figura 3.8: Performance delle quattro reti neurali proposta da Kok in [24]. Rette di regressione tra valore glicemico predetto e valore glicemico misurato [24].

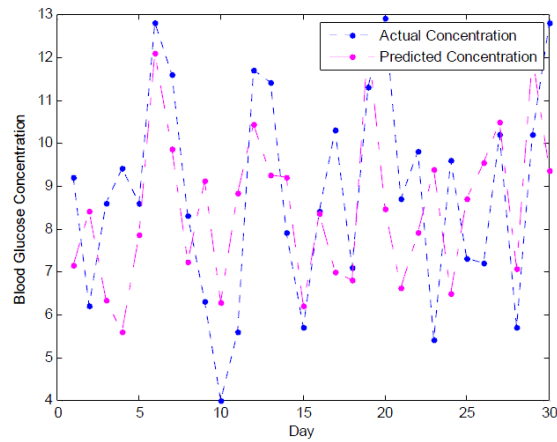


Figura 3.9: Plot delle performance di una delle reti neurali proposte da Zainuddin *et al.* in [48]. Glicemia predetta e misurata a fine mattinata. I risultati sono relativi a una rete neurale con funzioni wavelet gaussiane nello strato nascosto [48].

PERFORMANCE COMPARISON BETWEEN THE PROPOSED SYSTEM AND THE OTHER NEURAL NETWORK MODELS

| | RMSE for Different Intervals | | | |
|----------------------------|------------------------------|-----------|---------|--------|
| | Morning | Afternoon | Evening | Night |
| Kok [18] | 1.8 | 1.8 | 2.1 | 2.2 |
| Baghdadi and Nasrabadi [4] | 0.0826 | 0.0513 | 0.0373 | 0.0118 |
| The Proposed System | 0.0450 | 0.0348 | 0.0330 | 0.0170 |

Tabella 3.1: Confronto tra le performance delle reti neurali proposte da Kok in [24], Baghdadi *et al.* in [4] e da Zainuddin *et al.* in [48] in termini di RMSE dei vari predittori. (Tabella tratta da [48]).

3.5 Reti neurali con ingressi CGM

Visto il recente sviluppo dei sensori CGM sono pochi i lavori in letteratura che affrontano il problema della predizione glicemica, utilizzando reti neurali con in ingresso dati CGM.

3.5.1 Rete neurale con in ingresso misure di glucosio e informazioni su pasti, insulina ed esercizio fisico ricavate dalla serie glicemica stessa

Il lavoro di Eskaf *et al.* [11], che è uno dei primi che si trovano in letteratura, risale ad Aprile 2008. Gli autori utilizzano una rete neurale per predire il valore glicemico futuro, con PH di 30 minuti. I dati utilizzati provengono da un solo paziente, monitorato per 7 giorni consecutivi. Le misure glicemiche sono ottenute ogni 5 minuti e gli autori le definiscono misure ematiche, ma non è specificata la tecnica utilizzata per ottenere glicemie così frequenti e riferite direttamente al glucosio plasmatico e non interstiziale. Il metodo proposto dagli autori è particolarmente originale, in quanto le informazioni su carboidrati ingeriti ed effetto di insulina somministrata ed esercizio fisico non vengono richieste al paziente, ma vengono estratte dalla serie glicemica stessa. L'ipotesi di partenza è che il livello di glucosio nel sangue possa essere simulato come un sistema meccanico oscillatorio, in cui il cibo ingerito assume la forma di un ingresso impulsivo, mentre insulina somministrata ed esercizio fisico agiscono come un fattore di smorzamento. In un tale sistema il livello di glucosio può essere espresso da una funzione sinusoidale di ampiezza proporzionale alla forza dei carboidrati assunti e fase proporzionale allo smorzamento introdotto da insulina ed esercizio fisico. Utilizzando i 12 campioni glicemici relativi all'ultima ora di monitoraggio si calcolano la frequenza del sistema (attraverso DFT) e la forza dei carboidrati ingeriti e il fattore di smorzamento, utilizzando le equazioni matematiche che descrivono l'oscillatore smorzato. La rete usata per la predizione è di tipo feedforward, con un unico strato nascosto di 50 neuroni con funzione di attivazione tangente sigmoideale, ed un nodo d'uscita con funzione d'attivazione lineare. Il training viene eseguito con back-propagation con momento. Gli ingressi della rete sono: le 11 differenze prime del segnale glicemico registrato nell'ultima ora, la forza dei carboidrati assunti e lo smorzamento dovuto a insulina ed esercizio fisico, calcolati come spiegato precedentemente. L'uscita della rete è il fattore di cambiamento del glucosio a 30 minuti dall'ultimo campione registrato. Tale fattore di cambiamento è la differenza tra il valore futuro (tra 30 minuti) e l'ultimo valore misurato, normalizzata all'ultimo valore misurato. Le performance della rete vengono valutate su un test set indipendente dal training set, (di cui però non vengono fornite informazioni). Il RMSE risulta compreso tra 4 e 24.76 mg/dl. Da un plot in cui si confrontano valore predetto e valore reale

(riportato in Figura 3.10), si deduce che la predizione non è effettuata ogni volta che è disponibile un nuovo dato, bensì solo una volta ogni ora. Dopo

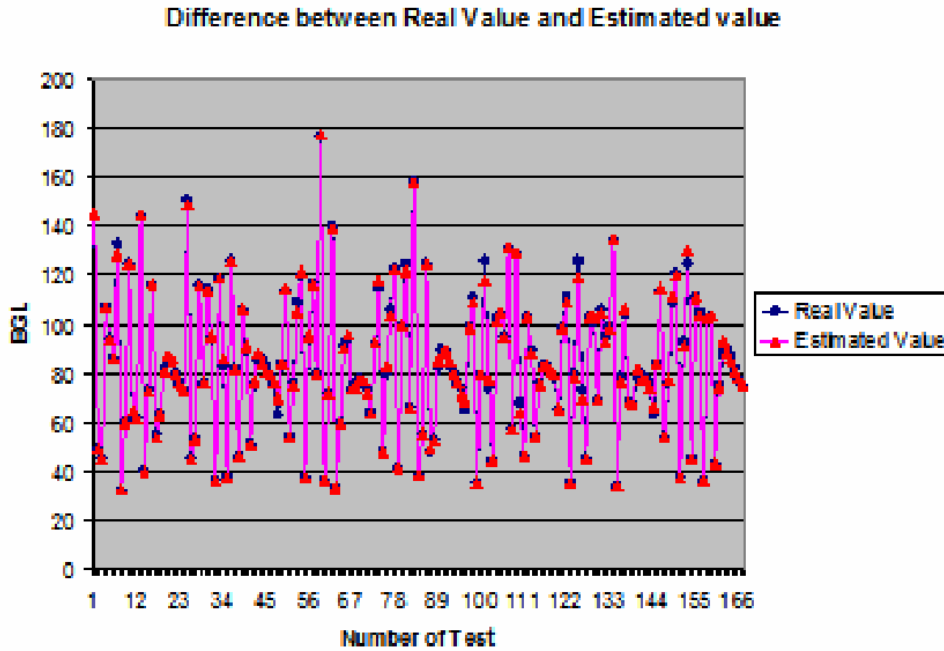


Figura 3.10: Performance della rete neurale di Eskaf *et al.*. Errore tra valore glicemico misurato e valore predetto [11].

aver predetto il valore glicemico a 30 minuti dall'ultimo valore misurato, infatti, si aspetta di raccogliere i dati relativi all'ora successiva e, solo in seguito, si effettua una nuova predizione. Non è chiaro il motivo per cui non venga eseguita una predizione "in continua". Inoltre, benchè risulti molto interessante il metodo con cui gli autori calcolano l'effetto di carboidrati, insulina ed esercizio fisico, sgravando totalmente il paziente dall'onere di doverli registrare, c'è da sottolineare che il modello oscillatorio è sicuramente una semplificazione ed un'approssimazione del complesso meccanismo di regolazione del glucosio. Di conseguenza, anche i valori ottenuti per la forza dei carboidrati e per il fattore di smorzamento sono un'approssimazione dei loro effetti reali.

3.5.2 Rete ricorsiva implementata e testata su pazienti diabetici virtuali

Al 2008 (Luglio) risale anche il lavoro di Sandham *et al.* [38]. Gli autori utilizzano una rete ricorsiva, e testano varie configurazioni che differiscono per il numero di eventi passati in ingresso e per il PH. Gli ingressi della rete sono: l'ora dell'ultimo valore misurato, la quantità di carboidrati ingeriti, le unità di insulina rapida ed intermedia iniettate, gli ultimi livelli glicemici

registrati (uno o più, in base alla configurazione adottata). L'orizzonte di predizione viene settato a 1, 2, 4, oppure 8 ore. In particolare, se il PH è di 4 ore le uscite della rete saranno 3, ovvero le predizioni della glicemia futura a 1, 2 e 4 ore. Con gli altri valori di PH il comportamento è analogo. I dati utilizzati sono simulati sfruttando il software AIDA. Si tratta di un programma educativo interattivo, che, utilizzando un modello compartimentale, simula il sistema glucosio-insulina nei pazienti diabetici di tipo 1. Lo scopo principale di questo software è simulare gli effetti di dieta e dosi di insulina sul livello di glucosio ematico. Sandham *et al.* valutano le performance della loro rete neurale in termini di errore medio e le definiscono confrontabili con quelle fornite dal simulatore interno al software AIDA.

3.5.3 Rete neurale per la predizione del valore glicemico successivo

Il lavoro di Pappada *et al.* [31] è datato Settembre 2008. Gli autori utilizzano una rete feedforward, sviluppata col software NeuroSolution[®], per predire la glicemia futura, con un orizzonte di predizione che varia da 50 a 180 minuti. Il dataset è composto dai dati di 18 pazienti, che hanno indossato un dispositivo CGM per un periodo della durata di almeno 3 e massimo 9 giorni. Le misure glicemiche sono fornite ogni 5 minuti. Oltre ad esse sono stati registrati anche i valori glicemici misurati con SMBG, i carboidrati ingeriti durante i pasti, le dosi di insulina somministrata, la presenza di esercizio fisico o di emozioni, la presenza di sintomi tipici dell'ipo o dell'iperglicemia. La rete utilizzata presenta degli elementi di ritardo all'ingresso, in modo da mantenere memoria di un certo numero di misurazioni passate, ed è configurata in modo da predire il valore glicemico successivo (ovvero quello che si avrà dopo 5 minuti dall'ultimo valore registrato). Di conseguenza, per PH più lunghi viene reiterata la predizione. La rete è allenata con l'algoritmo di back-propagation, fino al raggiungimento di un numero massimo prestabilito di epoche o, in alternativa, finché l'errore di predizione (Mean Square Error, MSE) non è inferiore a 0.1. Il numero di ingressi, il numero di neuroni nascosti e alcuni parametri da utilizzare durante il training vengono ottimizzati con un algoritmo genetico. Gli autori non forniscono però nessuna informazione circa i risultati ottenuti con questa procedura, quindi non è nota la struttura effettiva della rete. I risultati vengono valutati in termini di differenza media assoluta (Mean Absolute Difference, MAD) tra target e uscita della rete, e sono definiti buoni in euglicemia, soddisfacenti in iperglicemia, nonostante la predizione tenda a sottostimare il vero valore glicemico, mentre la rete non sembra in grado di predire le ipoglicemie che risultano sempre sovrastimate. Questo comportamento potrebbe essere dovuto alla scarsità di eventi ipoglicemici presenti nel training set con cui la rete è allenata. Gli autori osservano anche

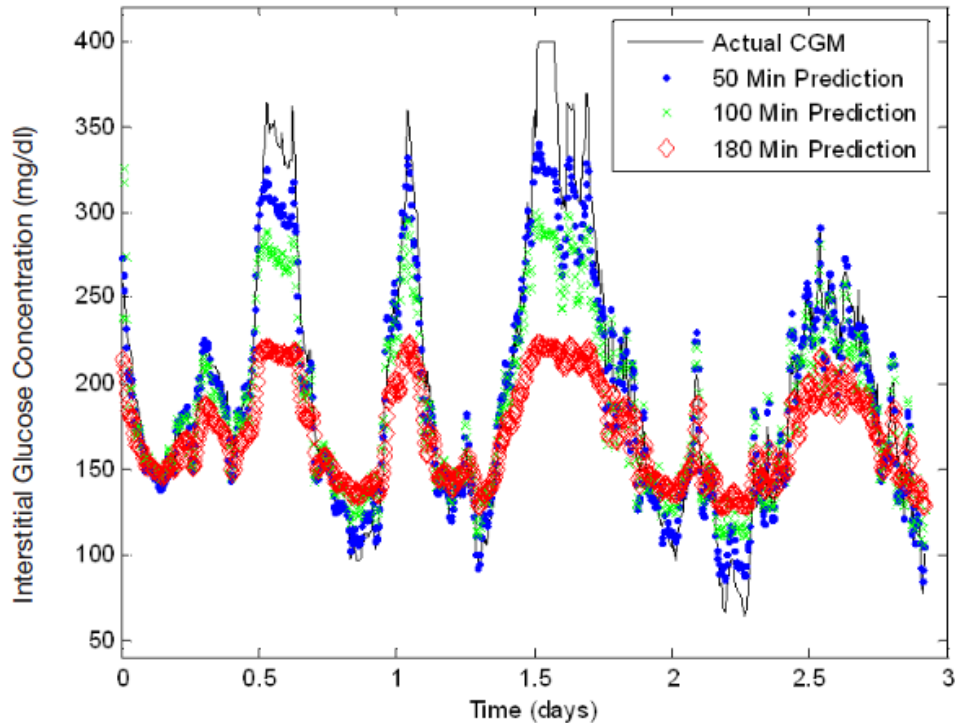


Figura 3.11: Performance della rete neurale di Pappada *et al.* [31].

un graduale peggioramento delle performance, all’aumentare dell’orizzonte di predizione, come si può notare in Figura 3.11.

3.5.4 Reti neurali e kernels per predire il profilo glicemico futuro

Ha un’impostazione profondamente diversa il lavoro di Pereverzyev *et al.* [32]. In questo caso, infatti, non viene semplicemente predetto il valore di glicemia dopo PH minuti rispetto all’ultimo campione registrato, viene invece predetta una funzione che descrive l’andamento glicemico dal momento attuale fino a PH minuti nel futuro. Gli autori utilizzano una struttura a 2 livelli, costituiti da due reti neurali adattative. In particolare, date in ingresso le ultime m misure CGM registrate, la rete “principale” sceglie uno spazio funzionale adeguato all’interno del quale l’altra rete individua la funzione da utilizzare per predire il profilo glicemico futuro. In pratica, la prima rete sceglie un kernel tra i vari proposti dagli autori, a tale kernel è associata una famiglia parametrica di funzioni. Il compito della seconda rete è individuare la funzione da utilizzare all’interno di tale famiglia, settando i parametri ottimi che la individuano. Tra i vari vantaggi di questa modalità c’è il fatto che viene fornito un profilo glicemico predetto, e non semplicemente un valore puntuale. Inoltre, se si vogliono incorporare

nuove informazioni all'interno della rete, basta aggiungere una nuova unità, che verrà allenata con dati adeguati, in modo da saper individuare un kernel e successivamente una funzione specifica che descriva l'influenza della nuova informazione sulla curva glicemica. Gli autori hanno agito proprio in

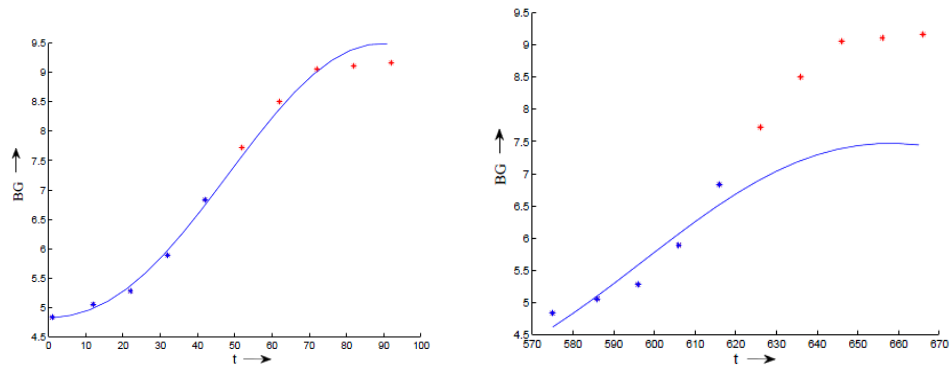


Figura 3.12: Performance del predittore proposto da Pereverzyev *et al.*. La linea continua rappresenta il profilo predetto dalle reti neurali, mentre gli asterischi corrispondono a misure registrate dal sensore CGM. L'orizzonte di predizione è pari a 50 minuti. Nella figura di sinistra la predizione è effettuata tenendo in considerazione le informazioni su pasti e insulina. Nella figura di destra il profilo è generato considerando solo informazioni relative alla glicemia [32].

questo modo per introdurre informazioni su pasti e insuline, migliorando le performance del predittore, come dimostra Figura 3.12. I risultati ottenuti con questa tecnica sono paragonabili a quelli presenti in letteratura, benchè il periodo di campionamento sia più alto di quello generalmente utilizzato (ovvero 10 minuti invece che 5 minuti o 1 minuto).

3.5.5 Rete neurale feedforward per la predizione on-line della glicemia, sfruttando esclusivamente misure CGM

Il lavoro di Pérez-Gandía *et al.* [34], pubblicato a Gennaio 2010, costituisce il punto di partenza del presente elaborato e verrà quindi descritto in dettaglio. La predizione è eseguita attraverso una rete neurale feedforward, che ha in ingresso solamente glicemie relative ai 20 minuti passati e l'istante temporale attuale, e come target la glicemia futura che si registrerà tra PH minuti (vengono testati PH di 15, 30 e 45 minuti). La rete è stata allenata e testata su due dataset distinti.

- *Datasets.* Il primo dataset è costituito da misure glicemiche di 9 pazienti, monitorati in modo non continuativo per 72 ore alla settimana, per 4 settimane. Per ogni paziente sono forniti 6 profili giornalieri completi e 6 profili di 12 ore ciascuno. Le glicemie sono misurate con sensore CGM Guardian[®], che fornisce una misura ogni 5 minuti.

Il secondo dataset è costituito da 6 pazienti, monitorati per 72 ore continuative con sensore CGM FreeStyle NavigatorTM, che fornisce una misura ogni minuto. I dati provenienti da questo sensore sono stati pre-filtrati usando un filtro causale di Kalman prima di essere utilizzati, per limitarne il rumore.

- *Architettura della rete.* La rete è feedforward, totalmente connessa, ha 2 strati nascosti di 10 e 5 neuroni rispettivamente e uno strato d'uscita contenente un unico neurone. Le funzioni d'attivazione sono di tipo tangente sigmoideale negli strati nascosti e lineare nello strato d'uscita. La rete è allenata con l'algoritmo di back-propagation, nella variante proposta da Levenberg-Marquardt, utilizzando la modalità batch. Il training set è composto, in entrambi i casi, da 6 profili giornalieri relativi a 3 soggetti diversi.
- *Ingressi della rete.* La rete ha in ingresso i valori di glucosio registrati negli ultimi 20 minuti e l'istante temporale attuale (espresso numerando in maniera successiva le misurazioni registrate, a partire dalla mezzanotte del giorno in esame). La rete allenata e testata su dataset Guardian[®] ha quindi 5 ingressi (dato che le misure sono fornite ogni 5 minuti), mentre la rete allenata e testata su dataset FreeStyle NavigatorTM ha 20 ingressi.
- *Metrica di valutazione delle performance.* La bontà della predizione è espressa in termini di *RMSE* e *ritardo della predizione*, calcolati su un validation set indipendente dai dati utilizzati per il training. In particolare il ritardo tra profilo target e profilo predetto è calcolato in modo distinto su trend crescenti e decrescenti, in concomitanza all'attraversamento di soglie posizionate al 25%, 50% e 75% della distanza nadir-picco per trend positivi, e della distanza picco nadir per trend negativi. Prima di eseguire questo calcolo entrambi i profili (target e predetto) vengono smoothati con filtro passa-basso di Butterworth.

I risultati ottenuti vengono confrontati con quelli ricavati col modello AR(1) di Sparacino *et al.* [39]. Il RMSE è leggermente inferiore nei profili predetti con FFNN, mentre i ritardi sono lievemente maggiori. Vale la pena notare che i ritardi sono significativamente maggiori in corrispondenza dei trend negativi. Questo potrebbe essere dovuto sia all'andamento del segnale glicemico stesso, che risulta più lento nelle fasi di discesa, sia alla minor presenza di ipoglicemie nel training set. Inoltre, i risultati ottenuti su dataset Guardian[®] sono migliori dei risultati ottenuti su dataset FreeStyle NavigatorTM. I motivi potrebbero essere il maggior rumore sovrapposto ai dati provenienti dal sensore FreeStyle NavigatorTM (dovuto, principalmente, alla più alta frequenza di campionamento), ma anche al fatto che l'architettura della rete e l'ampiezza del training set sono mantenute invariate con

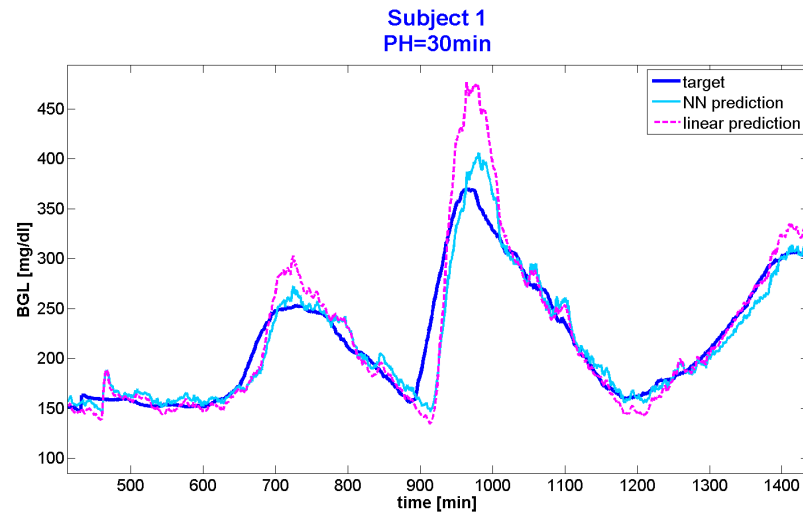


Figura 3.13: Performance della rete neurale di Pérez-Gandía *et al.* su profilo monitorato con sensore FreeStyle NavigatorTM. Profilo target in blu, profilo predetto dalla rete neurale in azzurro, profilo predetto dal modello lineare tempo variante in magenta tratteggiato [34].

entrambi i dataset. Però la rete che utilizza dati provenienti da FreeStyle NavigatorTM ha 21 ingressi, mentre la rete che utilizza dati Guardian[®] ha solo 6 ingressi, ciò si riflette in ben 150 pesi in più da stimare durante il training della prima rete.

3.6 Conclusioni

In questo capitolo abbiamo visto alcune tecniche proposte in letteratura per la predizione della glicemia futura. Il grande interesse dimostrato in ambito scientifico verso questi temi è giustificato dal fatto che un'accurata predizione della glicemia futura può introdurre notevoli benefici nella terapia convenzionale dei pazienti diabetici, limitando sia le complicanze a breve termine, sia quelle a lungo termine, legate a questa patologia. L'avvento dei sensori CGM apre nuove frontiere, poichè rende disponibile un'enorme quantità di misurazioni glicemiche, che possono essere utilizzate per affinare gli algoritmi di predizione già esistenti e, soprattutto, per svilupparne di nuovi, che sappiano sfruttare appieno la ricchezza del dataset reso disponibile.

Le reti neurali sono dei promettenti modelli, in quanto possono integrare informazioni provenienti da domini completamente diversi, ovvero misurazioni glicemiche passate, quantità di carboidrati ingeriti e dosi di insulina iniettata.

In questa tesi, partendo dalla rete neurale proposta da Pérez-Gandía *et al.* in [34], sono stati sviluppati 2 nuovi algoritmi basati su reti neurali. Nel

prossimo capitolo saranno descritte le varie modifiche apportate alla rete di partenza e le nuove soluzioni testate. Saranno inoltre presentate le tecniche utilizzate per ottimizzare la struttura delle nuove reti neurali e verranno definiti gli indici considerati per quantificarne le performance e scegliere la soluzione migliore tra le varie analizzate.

Capitolo 4

Approcci alla predizione della glicemia basati su reti neurali: possibili nuove architetture e criteri di valutazione

Come visto nel capitolo precedente, le reti neurali utilizzate in letteratura per la predizione del livello glicemico presentano varie strutture e diversi tipi di ingressi. Poichè non esistono né una regola generale, né basi teoriche per la scelta di una particolare rete, nel presente lavoro sono state testate varie architetture e varie configurazioni. Verranno di seguito elencate le possibili modifiche da apportare alla rete di Pérez-Gandía *et al.* [34]. Successivamente saranno descritti il metodo utilizzato per la scelta della rete “ottima” e le metriche usate per esprimere la bontà del predittore.

4.1 Architetture candidate

Le modifiche testate riguardano il numero di strati e di neuroni della rete, il numero e il tipo di ingressi, l’architettura della rete stessa e il target della predizione. Naturalmente, questo test non è esaustivo, né potrebbe esserlo, non esistendo una procedura standard da seguire.

4.1.1 Architettura della rete neurale

L’architettura della rete è molto importante, poichè essa influenza la capacità della rete neurale di generalizzare. Per una semplice rete neurale feedforward, allenata con l’algoritmo di back-propagation, le possibili variabili sono il numero di strati nascosti e il numero di neuroni in ogni strato. Se

il numero totale di nodi è troppo piccolo la rete non è generalmente in grado di rappresentare adeguatamente il training set e l'errore di predizione è alto. Un numero di nodi eccessivo, d'altra parte, può portare a overfitting dei dati, che risulta comunque limitato se l'allenamento della rete viene fermato per cross validation.

Numero di strati nascosti

Il *teorema di approssimazione universale* (pag. 34) afferma che una rete neurale con un solo strato nascosto può approssimare accuratamente qualsiasi funzione, a patto che abbia un numero sufficiente di neuroni nascosti con funzione di attivazione non lineare. Per alcune funzioni, però, il numero di neuroni necessario potrebbe essere molto grande o, addirittura, infinito. In questi casi una rete neurale con due strati nascosti solitamente è molto più adeguata.

La rete neurale utilizzata in [34] è costituita da due strati nascosti. Nel presente elaborato sono state testate reti sia con uno che due strati nascosti. I risultati ottenuti hanno dimostrato che un solo strato nascosto, di dimensioni accettabili, è sufficiente, (come era stato notato anche da Kok in [24]), e due strati nascosti non migliorano le performance della rete, ma accrescono il numero di parametri da stimare durante l'allenamento, rallentando questa fase e portando a risultati meno accurati per tutti i pesi.

Numero di neuroni

Esistono varie tecniche per determinare il numero di unità appropriate per una rete neurale. Tra le più utilizzate ci sono i metodi definiti *growing* (crescita) e *pruning* (potatura). Essi, rispettivamente, aumentano e diminuiscono il numero di neuroni e di connessioni dinamicamente, durante l'allenamento della rete. Il metodo utilizzato nel presente elaborato è invece una tecnica statica. Esso consiste nell'allenamento successivo di reti con un numero crescente di neuroni, e nella registrazione delle performance di ognuna. La rete scelta sarà quella più piccola che fornisce un miglioramento significativo delle performance rispetto alle architetture più semplici.

Numero, tipo e forma degli ingressi

La selezione degli ingressi è uno degli aspetti più critici nella creazione di una rete neurale. Affinchè la rete possa imparare da essi, deve esserci correlazione tra ingressi e uscita. Se gli ingressi selezionati sono troppi, però, si rischia di ottenere una rete che non generalizza adeguatamente.

Sono stati testati vari ingressi, che influenzano il livello glicemico, come la quantità di carboidrati ingeriti e le dosi di insulina somministrata. Inoltre, è stato oggetto di indagine anche il numero di misure passate in ingresso alla rete. Anche questo dato risulta critico, soprattutto in una rete feedforward

(utilizzata per predire serie temporali). Le reti feedforward, infatti, non hanno memoria, e l'unico modo per dotarle di conoscenze sul passato è inserire tra gli ingressi non solo l'ultima misura, ma anche un certo numero di misure precedenti.

Per velocizzare e ottimizzare il training della rete è inoltre consigliato scalare gli ingressi in modo che varino approssimativamente nell'intervallo $[0, 1]$ (se si utilizzano funzioni d'attivazione sigmoidali), o nell'intervallo $[-1, 1]$ (se si utilizzano funzioni d'attivazione di tipo tangente sigmoidale). In tal modo tutti gli ingressi acquistano inizialmente la stessa importanza e, inoltre, variano nel range in cui le funzioni d'attivazione dei neuroni nascosti si comportano linearmente.

La scala glicemica, però, non è simmetrica e il valore euglicemico è molto più vicino alla soglia dell'ipoglicemia che a quella dell'iperglicemia, come si può osservare in Figura 4.1. A ciò si aggiunge il fatto che le glicemie dei pazienti diabetici sono spesso polarizzate nella zona iperglicemica e presentano, occasionalmente, anche valori molto alti. Quando si mappano linearmente i dati registrati dal sensore CGM nell'intervallo $[-1, 1]$ si trova un grandissimo numero di dati negativi. Ciò potrebbe parzialmente compromettere il funzionamento della rete, poichè il range ottimo in cui dovrebbero variare gli ingressi non viene adeguatamente utilizzato. Una mappatura che riporta i dati a media nulla e standard deviation unitaria risulta più adeguata, anche se l'insieme di valori in ingresso alla rete non è comunque simmetricamente distribuito attorno allo zero. Sono state utilizzate per la mappatura dei valori glicemici in ingresso anche funzioni sigmoidali, nonché la funzione rischio, proposta da Kovatchev *et al.* in [27], che ha la peculiarità di simmetrizzare la scala glicemica. Questi tentativi non hanno comunque portato ad un miglioramento dei risultati ottenuti.

Reti neurali ricorsive

Per superare il problema connesso all'assenza di memoria delle reti feedforward e rendere le performance della rete meno dipendenti dal numero di misure glicemiche passate utilizzate in ingresso si può ricorrere ad una rete ricorsiva. Grazie al feedback questa rete, infatti, è in grado di ricordare teoricamente tutta la storia passata degli ingressi.

Sono state testate varie reti ricorsive, sia con uno che con due strati nascosti, con feedback sia solo dall'uscita, sia dall'uscita di ogni strato di neuroni. I risultati ottenuti non sono stati soddisfacenti, soprattutto perchè le reti spesso non arrivavano a convergenza. Si è quindi preferito mantenere e investigare l'architettura feedforward. Le reti ricorsive restano comunque un terreno da approfondire in maniera più dettagliata, utilizzando per il training algoritmi più idonei ed efficienti della *back-propagation through time*¹.

¹L'algoritmo di back-propagation through time è l'unico reso disponibile nel Neural Network Toolbox™ di Matlab®7.10.0 per l'allenamento delle reti neurali ricorsive.

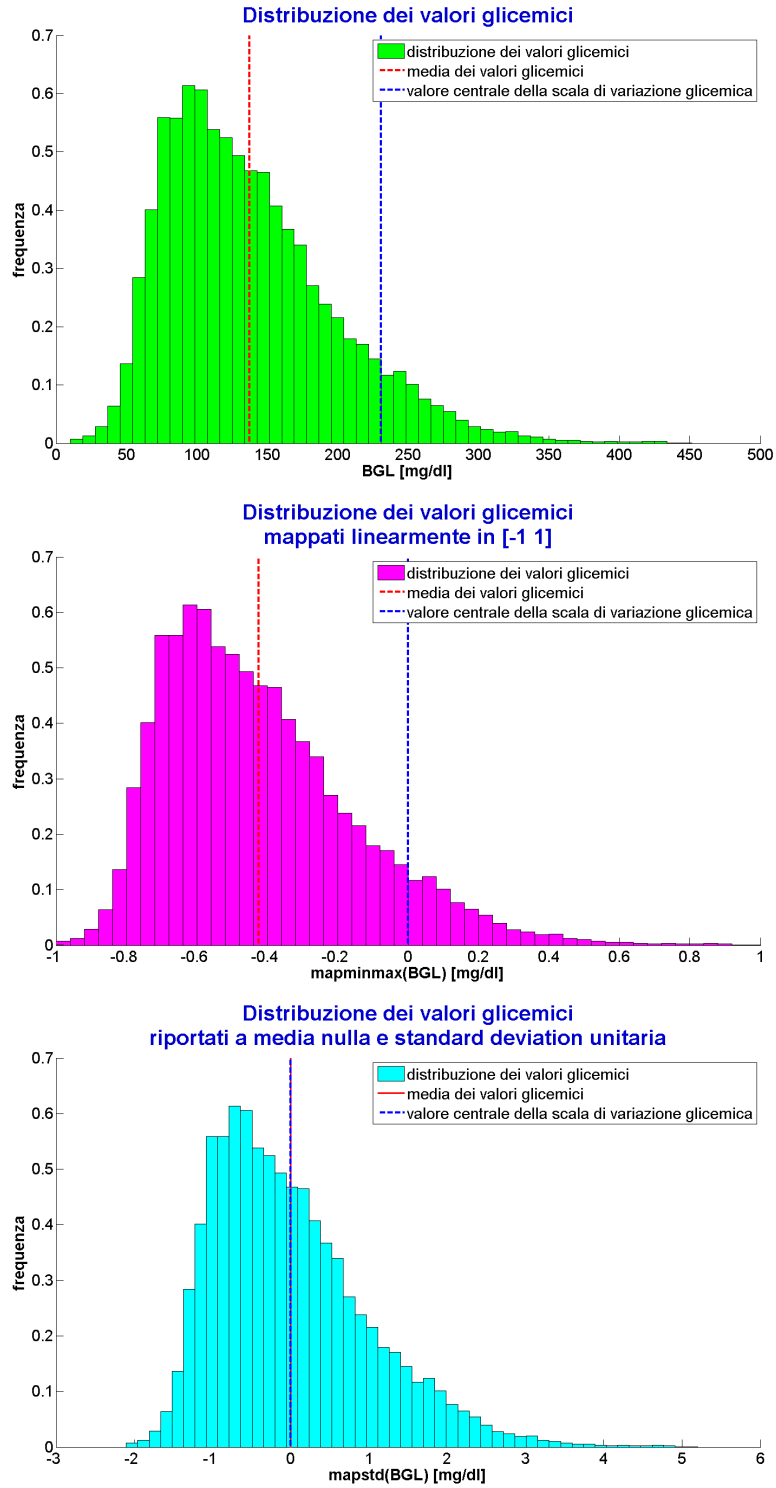


Figura 4.1: Asimmetria della distribuzione dei valori glicemici.

4.1.2 Target della predizione

L'obiettivo della rete neurale è la predizione del livello glicemico futuro, a PH minuti dall'ultimo valore registrato. Questo risultato può essere ottenuto predicendo direttamente il valore futuro, ma anche in altri modi di seguito elencati.

Errore del modello lineare

Come suggerito da Pijenburg in [33], in molte situazioni pratiche si hanno conoscenze sufficientemente complete sulla parte lineare di un sistema non lineare. Queste conoscenze possono essere sfruttate, utilizzando un modello lineare in parallelo a una rete neurale, sollevando così la rete neurale dall'onere di dover modellare anche la parte lineare del sistema e lasciandole il solo compito di descrivere le non linearità del processo.

Le fluttuazioni glicemiche registrate nell'arco della giornata seguono un andamento non lineare, ma ben approssimabile con modelli lineari. È quindi sensato utilizzare la rete neurale in parallelo al modello lineare, in modo che essa, in pratica, predica solamente l'errore del modello lineare, in cui sono condensate tutte le dinamiche non lineari del sistema. Una rete così strutturata si è dimostrata non inferiore a una rete che predice direttamente il livello glicemico.

Predizione ad un passo reiterata

Quando vengono utilizzati modelli AR, solitamente, si predice il valore immediatamente futuro e si reitera il procedimento, utilizzando come ingressi del modello le predizioni passate, fino a predire il valore a PH minuti. Un approccio simile è adottato anche da Tresp *et al.* [42].

È quindi stata progettata una rete che, utilizzando solamente misure glicemiche passate, veniva allenata per predire il valore immediatamente futuro (quindi con PH=1 minuto). Successivamente, in fase di simulazione, la predizione a un minuto veniva reindirizzata in ingresso alla rete per ottenere la predizione successiva (ovvero a 2 minuti), e il procedimento veniva reiterato PH volte. Questa rete si è però dimostrata totalmente incapace di predire la glicemia futura. Infatti, con PH di 1 minuto, la miglior scelta per la rete era predire come valore successivo l'ultimo valore visto in ingresso. In tal modo, però, la rete non coglie il trend del segnale glicemico e, reiterando PH volte il procedimento, si ottiene una predizione con un ritardo pari all'orizzonte di predizione stesso.

Classificazione

Nel caso l'obiettivo della predizione non sia la conoscenza dell'esatto valore glicemico, ma sia sufficiente ottenere un'informazione approssimativa,

si può pensare di dividere la scala glicemica in varie classi e predire non più il valore futuro puntuale, bensì la classe a cui apparterrà il valore glicemico futuro. Ci si riconduce, in tal modo, a un problema di classificazione. Poiché l'impostazione di una rete neurale per la classificazione è profondamente diversa da quella adottata per la predizione di serie temporali, questo approccio non verrà indagato.

4.2 Scelta della struttura migliore

Per scegliere la struttura della rete e il tipo di ingressi sono state confrontate le performance dei vari modelli. Le varie reti sono state allenate e testate sul solo training set (da cui era già stato escluso il validation set), per un numero predefinito di epoche. In particolare, vista la natura stocastica delle reti neurali, è stato utilizzato un metodo detto *k-fold cross validation*. Ovvero, come schematizzato in Figura 4.2, il training set è stato

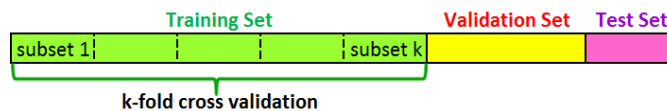


Figura 4.2: *Divisione del dataset per l'applicazione di k-fold cross validation.*

suddiviso in k sottoinsiemi disgiunti, contenenti ognuno lo stesso numero di coppie ingresso-uscita. Nel nostro caso sono stati utilizzati 4 sottoinsiemi, contenenti ognuno 7 serie temporali della durata di 3 giorni (4246 campioni) ciascuna. Si allena quindi la rete su $k - 1$ sottoinsiemi e la si testa sulle rimanenti serie temporali. Questo procedimento viene ripetuto k volte, utilizzando come “test set” ogni volta un sottoinsieme diverso. Le performance dell'architettura sotto analisi sono la media delle performance ottenute nei k test. Questo procedimento viene ripetuto per ogni architettura da indagare. Solitamente si parte con una rete con un solo strato nascosto ed un solo neurone, si aumenta esponenzialmente il numero di neuroni nascosti, per avere un'idea dell'ordine di grandezza della quantità di neuroni da usare e, successivamente, si infittisce l'analisi nell'intorno del numero di neuroni che è risultato migliore. Si procede poi allo stesso modo, aggiungendo alla rete un ulteriore strato nascosto.

Dopo aver scelto la struttura migliore della rete è stato analizzato, seguendo lo stesso procedimento, il tipo e il numero di ingressi, mantenendo inalterata la struttura scelta. Sono quindi stati aggiunti solo gli ingressi che apportavano un miglioramento effettivo al profilo predetto.

4.3 Metriche di valutazione delle performance della rete

Le performance della predizione sono state espresse con vari indici, ovvero il *Root Mean Square Error* (RMSE), J [43], e il *guadagno temporale* (Temporal Gain, TG). Ognuno di essi ha delle utili proprietà e dei limiti. Il giudizio finale, quindi, deve tenere conto dei risultati ottenuti per i vari indici, ma non può prescindere dalla valutazione visiva obiettiva del profilo predetto.

4.3.1 Regolarità del profilo predetto: J

L'indice J è definito come [43]

$$J = \frac{\text{ESOD}_{\text{norm}}}{\text{TG}_{\text{norm}}^2} \quad (4.1)$$

dove

$$\text{ESOD}_{\text{norm}} = \frac{\text{ESOD}_{\text{predizione}}}{\text{ESOD}_{\text{CGM}}} \quad (4.2)$$

ovvero l'energia delle derivate seconde (ESOD, Energy of Second Order Derivative) normalizzata è il rapporto tra l'energia delle derivate seconde del segnale predetto e l'energia delle derivate seconde del segnale target, e

$$\text{TG}_{\text{norm}} = \frac{\text{PH} - \text{ritardo medio}}{\text{PH}} \quad (4.3)$$

ovvero il guadagno temporale normalizzato è il rapporto tra il guadagno temporale medio e l'orizzonte di predizione.

Questo indice è particolarmente adeguato a fornire informazioni sull'irregolarità del segnale predetto, rispetto al profilo target, e tanto più assume valori prossimi a 0 tanto più è buona la predizione.

4.3.2 Aderenza del profilo predetto al target: RMSE

Il RMSE è uno degli indici più utilizzati in letteratura per esprimere la bontà di un predittore e, solitamente, è la quantità che viene minimizzata durante l'allenamento di una rete neurale. Quest'indice esprime la distanza tra il profilo predetto e il profilo target ed è definito da

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{y}_{\text{predetto}}^{(i)} - \mathbf{y}_{\text{CGM}}^{(i)})^2} \quad (4.4)$$

dove $\mathbf{y}_{\text{predetto}}^{(i)}$ indica l' i -esimo campione del segnale predetto, mentre $\mathbf{y}_{\text{CGM}}^{(i)}$ è l' i -esimo campione del segnale target e N è la lunghezza della serie temporale (ovvero il numero totale di campioni). Anche in questo caso, più

è piccolo il valore del RMSE, più è buona la predizione. Uno dei difetti principali di questo indice è il fatto che esso non tiene minimamente conto della regolarità del profilo predetto, quindi, un segnale che oscilla ad altissima frequenza attorno al valore target può avere un RMSE comunque molto basso, risultando apparentemente migliore di un segnale che segue perfettamente l'andamento del target, ma con un piccolo bias o un ritardo ininfluente. La Figura 4.3 illustra quanto appena affermato.

4.3.3 Guadagno Temporale

Il guadagno temporale è l'anticipo con cui si riesce a predire il segnale target. Idealmente esso dovrebbe corrispondere al PH, in pratica la predizione è quasi sempre affetta da un certo ritardo, di conseguenza il guadagno temporale risulta

$$TG = PH - \text{ritardo} \quad (4.5)$$

Il guadagno temporale può essere calcolato in vari modi, di seguito vengono descritte le due formule del guadagno temporale utilizzate nel presente elaborato.

Guadagno temporale medio

È definito come

$$TG_{\text{medio}} = PH - \text{ritardo}_{\text{medio}} \quad (4.6)$$

con il ritardo medio definito come lo shift che minimizza il RMSE. Il vantaggio di questa definizione è la facilità di calcolo e il fatto che non ci siano parametri da settare (a parte il ritardo massimo ammesso). Gli svantaggi principali sono quelli già citati per il RMSE, ed il fatto che una predizione potrebbe risultare meno ritardata non perchè effettivamente lo sia, ma semplicemente perchè il RMSE rimane sempre molto alto. Per calcolare questo parametro è stato eseguito un pre-filtraggio del profilo target e del profilo predetto utilizzando un filtro di Butterworth, in modo da limitare le oscillazioni ad alta frequenza presenti in alcuni segnali. In presenza di troppo rumore, infatti, il guadagno temporale risultava poco veritiero.

Guadagno temporale in corrispondenza dell'attraversamento di soglie critiche

È definito come

$$TG_{\text{ipoglicemia}} = PH - \text{ritardo}_{\text{ipoglicemia}} \quad (4.7)$$

$$TG_{\text{iperlicemia}} = PH - \text{ritardo}_{\text{iperlicemia}} \quad (4.8)$$

dove $\text{ritardo}_{\text{ipoglicemia}}$ è il ritardo con cui il segnale predetto attraversa la soglia dell'ipoglicemia (70 mg/dl) rispetto al segnale target, mentre $\text{ritardo}_{\text{iperlicemia}}$

4.3 Metriche di valutazione delle performance della rete

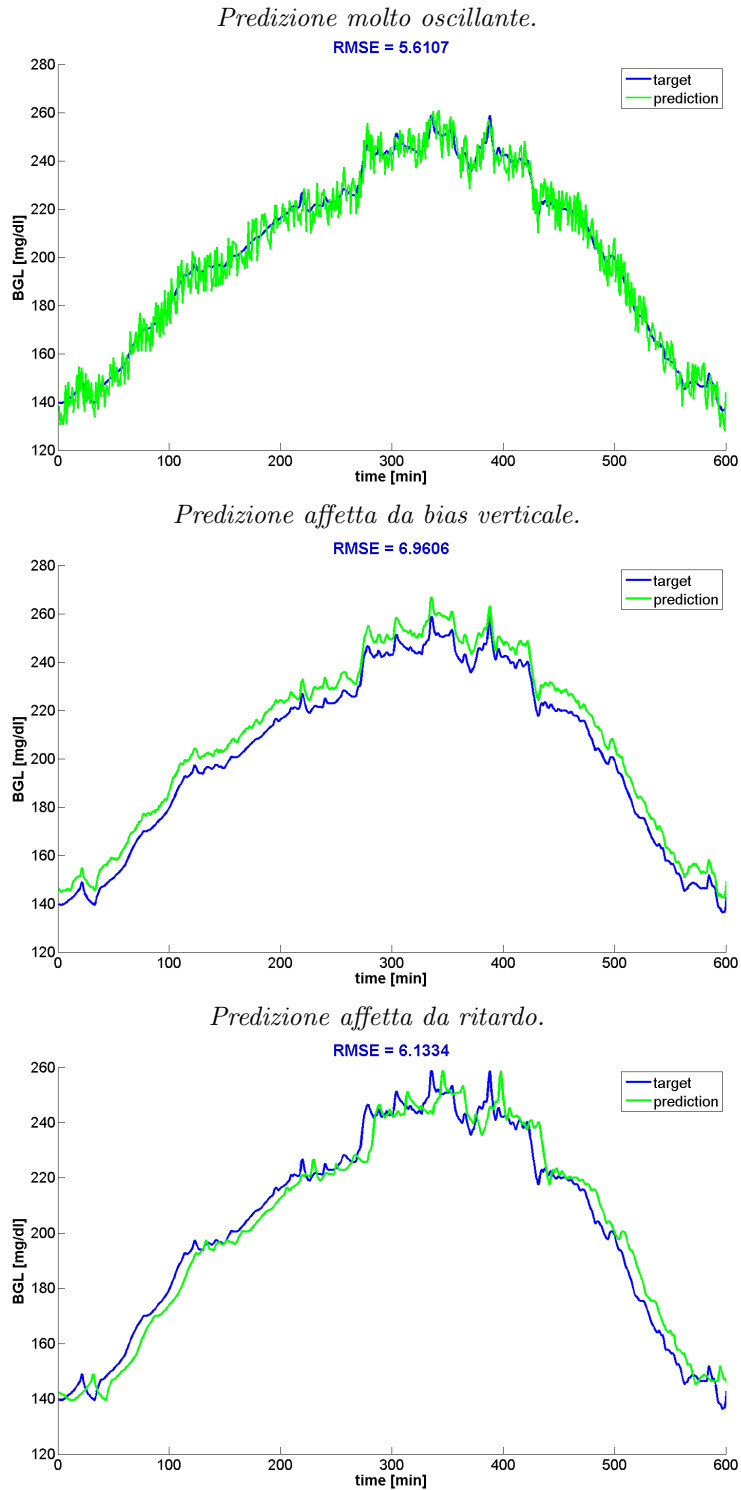


Figura 4.3: Difetti del RMSE. La predizione con il RMSE minore è fisiologicamente non accettabile e visivamente peggiore delle altre due.

è il ritardo con cui la predizione attraversa la soglia iperglicemica (180 mg/dl) rispetto al target.

Questa definizione del guadagno temporale è particolarmente utile nel caso la predizione sia mirata a individuare future crisi ipo e iperglicemiche. I difetti connessi a questo indice sono il maggior numero di parametri decisionali da settare e il fatto che non tiene conto né dei falsi attraversamenti, né degli attraversamenti non predetti.

4.3.4 Obiettivo della predizione

La valutazione delle performance di un predittore non può prescindere dall'obiettivo per cui il predittore è utilizzato. Nel nostro caso si possono individuare due obiettivi principali:

Aderenza del segnale predetto al profilo reale . In questo caso si è interessati ad ottenere un'uscita il più possibile simile al segnale target, nella maggior parte degli istanti temporali predetti.

Predizione dell'attraversamento di soglie critiche . Il comportamento del segnale predetto e il ritardo nella predizione, lontano dalle soglie critiche, sono di scarso interesse. L'importante è predire in maniera tempestiva un'eventuale crisi ipo o iperglicemica futura, in modo da poter intervenire per evitarla o limitarne la durata e gli effetti.

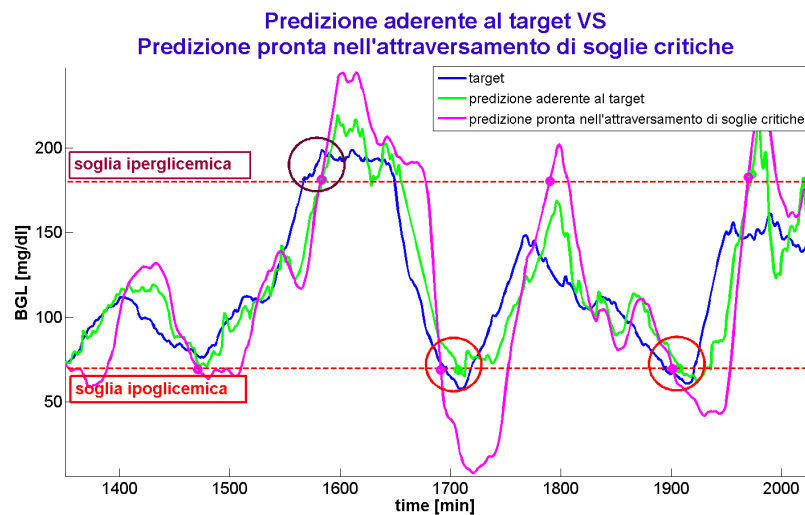


Figura 4.4: Confronto tra una predizione aderente al target (profilo verde) e una predizione pronta in corrispondenza dell'attraversamento di soglie critiche (profilo magenta).

I concetti appena introdotti sono rappresentati in Figura 4.4, dove possiamo notare che la predizione verde, sebbene sia in media più simile e vicina

al profilo target, attraversa le soglie critiche dell'ipoglicemia (cerchiate in rosso) in ritardo rispetto alla predizione magenta. Anche in corrispondenza della soglia iperglicemica (cerchiata in marrone) il profilo magenta anticipa leggermente quello verde.

4.4 Conclusioni

In questo capitolo sono state descritte varie reti neurali utilizzabili per la predizione della glicemia futura. È inoltre stato presentato l'algoritmo utilizzato per determinare l'architettura migliore tra quelle indagate e le metriche sfruttate per valutare le performance delle reti neurali. Infine è stato notato che i risultati ottenuti con un determinato modello devono sempre essere valutati in base all'obiettivo della predizione. Inoltre, scopi diversi possono portare a prediligere algoritmi diversi.

Nel prossimo capitolo verrà presentata la procedura che ha condotto allo sviluppo di due nuovi algoritmi basati su reti neurali, utilizzati per predire la glicemia futura. In particolare, verranno descritti tutti gli ingressi testati, sia quelli effettivamente utilizzati nella configurazione definitiva, sia quelli scartati poichè non apportavano miglioramenti significativi. Infine si accennerà ad alcune potenzialità e ad alcuni difetti delle reti neurali proposte.

Capitolo 5

Due nuovi approcci basati su reti neurali per predire la glicemia futura

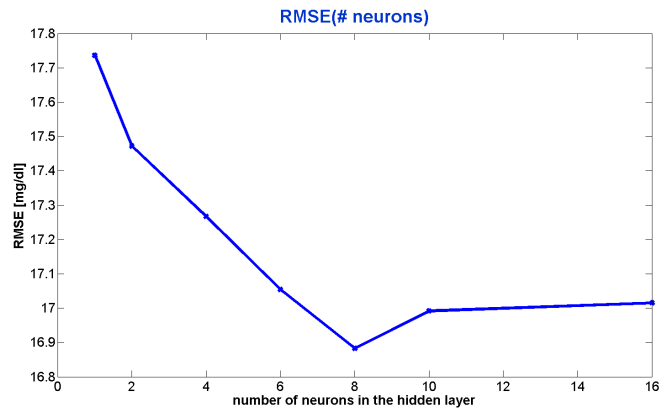
In questo capitolo verranno descritti nel dettaglio i principali passi che hanno portato, a partire dalla rete neurale proposta da Pérez-Gandía *et al.* in [34], allo sviluppo di nuove reti feedforward per la predizione della glicemia futura. Le reti proposte sono state testate unicamente su dati simulati [9] e su dati provenienti da sensore FreeStyle NavigatorTM, che fornisce una misurazione glicemica al minuto. L'orizzonte di predizione utilizzato per ottimizzare la struttura delle nuove reti è stato settato a 30 minuti. Successivamente, mantenendo invariata la struttura scelta, sono state testate le performance delle reti sviluppate anche con altri orizzonti di predizione.

5.1 Approccio 1: rete neurale feedforward per la predizione del livello glicemico

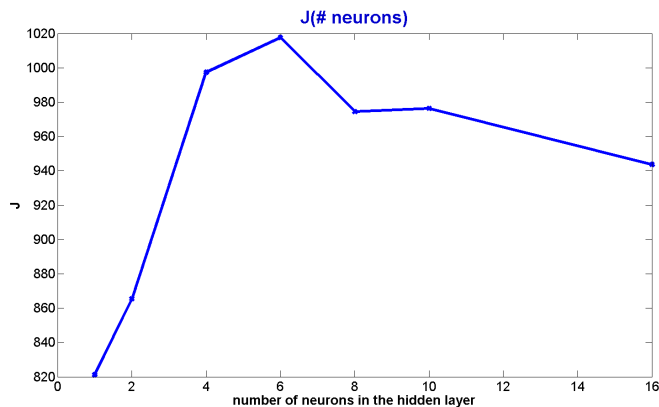
La rete neurale proposta in [34] è una rete feedforward con 2 strati nascosti, composti rispettivamente da 10 e 5 neuroni, con un neurone d'uscita e un numero d'ingressi tale da fornire alla rete una memoria di 20 minuti (ciò equivale a 20 ingressi se il periodo di campionamento del sensore CGM è pari a 1 minuto, come accade col FreeStyle NavigatorTM). I neuroni degli strati nascosti hanno funzione d'attivazione tangente sigmoideale, mentre il neurone d'uscita è caratterizzato da funzione d'attivazione lineare. Gli ingressi e il target presentati alla rete durante l'allenamento vengono automaticamente mappati linearmente nell'intervallo $[-1, 1]$ dall'algoritmo di training stesso. Il training è eseguito attraverso back-propagation, in modalità batch, e viene fermato per cross validation, in modo da evitare l'overfitting dei dati.

5.1.1 Semplificazione dell'architettura della rete

La rete appena descritta, benchè abbia una struttura non eccessivamente ingombrante, presenta 281 parametri incogniti, tra pesi e biases. Ciò implica, in fase di training, l'ottimizzazione di una funzione di 281 variabili, e può portare a stime poco accurate. Il primo passo è stato, quindi, verificare se strutture più semplici fossero in grado di fornire predizioni di qualità non inferiore. Attraverso la procedura di k-fold cross validation, descritta nella Sezione 4.2 (pag. 82), si è visto che una rete con un unico strato nascosto di 8 neuroni dà risultati non inferiori a quelli ottenuti con reti con un maggior numero di neuroni e/o di strati nascosti, come dimostrano Figura 5.1, Figura 5.2 e Figura 5.3. Per contro, i parametri da stimare in fase di training sono 185, ovvero 96 in meno rispetto all'architettura precedente.



(a) RMSE in funzione del numero di neuroni nascosti.



(b) J in funzione del numero di neuroni nascosti.

Figura 5.1: Andamento del RMSE e di J all'aumentare del numero di neuroni nascosti della rete neurale.

In Figura 5.1(a) notiamo che il RMSE indica una netta superiorità della

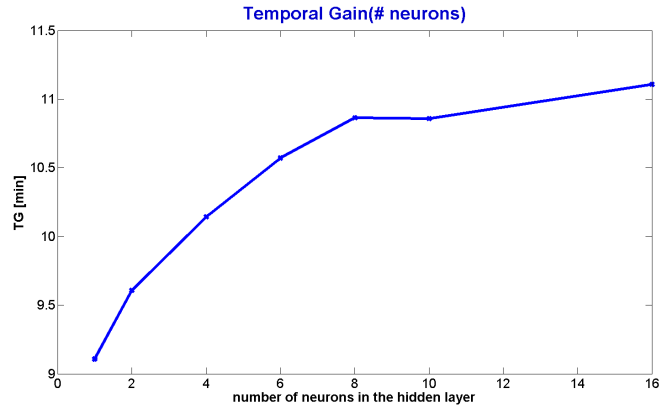
rete neurale con 8 neuroni nello strato nascosto. L'andamento di J , riportato in Figura 5.1(b), invece, porterebbe a scegliere reti neurali con un numero di neuroni minimo, poichè esse si comportano in maniera molto simile ad un modello lineare. Il profilo predetto da queste reti, però, benchè regolare, è troppo ritardato per essere utile.

In Figura 5.2(a) e in Figura 5.2(b) si può osservare che il guadagno temporale medio e il guadagno temporale in corrispondenza dell'attraversamento della soglia ipoglicemica mostrano che reti con un numero di neuroni nascosti superiore a 8 non hanno performance significativamente migliori di ffn-8-1. Il guadagno temporale relativo all'iperglicemia, riportato in Figura 5.2(c), presenta un punto di massimo in corrispondenza della rete con 8 neuroni, seguito da una marcata discesa per la rete con 10 neuroni. Il suo valore torna successivamente a crescere. Nonostante ciò, poichè gli altri indici mostrano la superiorità di ffn-8-1 rispetto alle altre reti, è stata scelta comunque questa struttura.

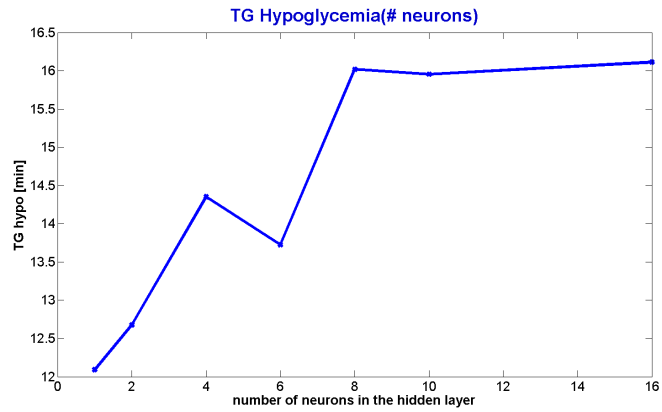
Per quanto riguarda l'analisi visiva dei profili predetti da reti neurali con un diverso numero di neuroni nello strato nascosto, in Figura 5.3(a) osserviamo che la rete neurale con un unico neurone (pannello in alto a sinistra) produce una predizione molto simile a quella di un modello lineare, ma con un ritardo superiore. La rete neurale con 4 neuroni (pannello in alto a destra) ha un ritardo non trascurabile. La rete neurale con 16 neuroni (pannello in basso a destra) produce una predizione più irregolare delle altre. Quindi, la rete neurale con 8 neuroni (pannello in basso a sinistra) risulta, anche visivamente, leggermente migliore. In Figura 5.3(b) sono rappresentati, su uno stesso grafico, i profili predetti da reti neurali con 1, 4, 8 e 16 neuroni nello strato nascosto. Il profilo predetto dalla rete neurale con 8 neuroni (in verde) ha un ritardo leggermente inferiore rispetto ai profili arancio e nero, predetti dalle reti con 1 e 4 neuroni rispettivamente, ed un ritardo paragonabile con quello del profilo magenta, predetto dalla rete con 16 neuroni. Il profilo magenta risulta però più irregolare. Quindi, anche da questa analisi, ffn-8-1 risulta la rete da preferire.

È quindi stata adottata questa configurazione e tutte le modifiche future sono state testate su questa rete. Solamente alla fine, dopo aver determinato gli ingressi che risultavano effettivamente utili alla rete, è stata nuovamente eseguita l'ottimizzazione della struttura. I risultati ottenuti hanno confermato che, anche con i nuovi ingressi, la rete costituita da un solo strato nascosto di 8 neuroni ha performance non inferiori ad architetture più elaborate, come si può notare in Figura 5.10, in Figura 5.11 e in Figura 5.12, alla fine di questa sezione (pag 104, pag 106, e pag 107 rispettivamente).

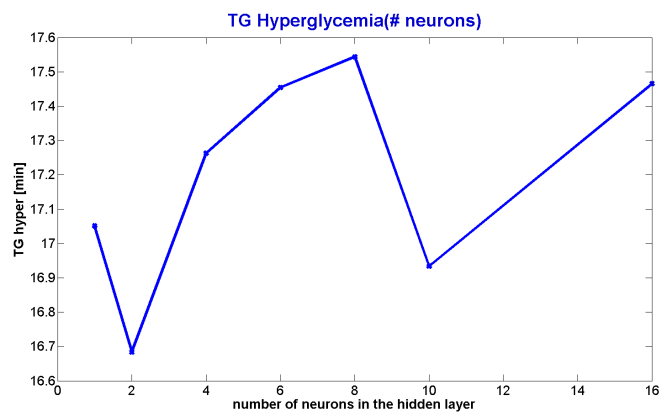
CAPITOLO 5. Due nuovi approcci basati su reti neurali per predire la glicemia futura



(a) Guadagno temporale medio.



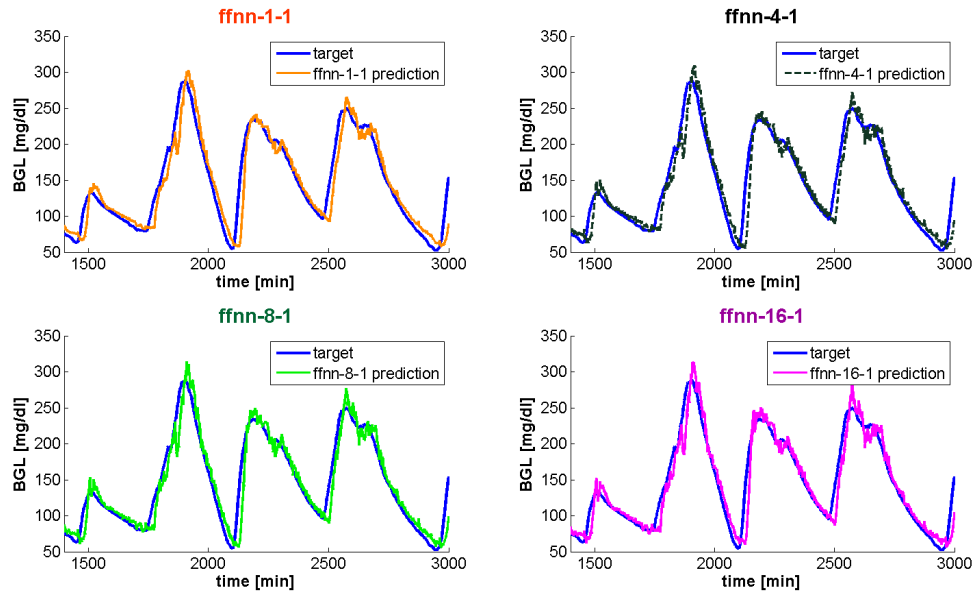
(b) Guadagno temporale relativo alla predizione di ipoglicemie.



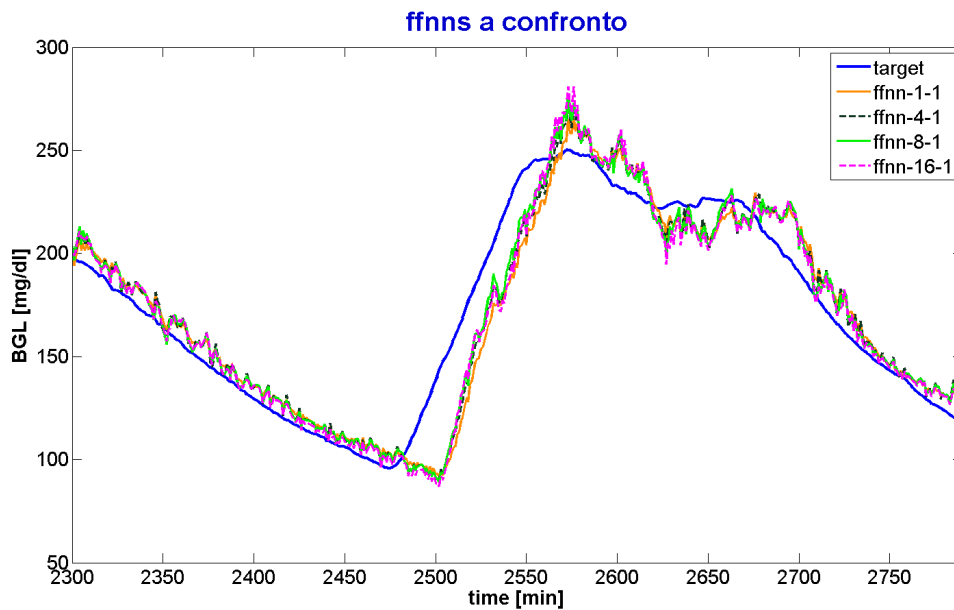
(c) Guadagno temporale relativo alla predizione di iperglicemie.

Figura 5.2: Andamento del guadagno temporale all'aumentare del numero di neuroni nascosti della rete neurale.

5.1 *Approccio 1: rete neurale feedforward per la predizione del livello glicemico*



(a) Profili predetti da reti neurali con diversi numeri di neuroni nello strato nascosto.



(b) Profili sovrapposti, predetti da reti neurali con un diverso numero di neuroni nello strato nascosto.

Figura 5.3: *Confronto tra profili predetti da reti neurali con un diverso numero di neuroni nello strato nascosto.*

5.1.2 Nuovi ingressi considerati nella rete

Informazioni sull'andamento passato del segnale

Una rete neurale feedforward è priva di memoria, risulta quindi particolarmente critico il numero di campioni precedenti della serie temporale che le vengono presentati in ingresso, poichè da essi dipende l'abilità della rete di apprendere relazioni tra i dati passati e il comportamento futuro del segnale. Per questo motivo, il primo passo è stato testare il numero di ingressi ottimo della rete. Si è visto che la storia relativa agli ultimi 15 minuti è sufficiente a predire il valore glicemico futuro, con PH=30 minuti. Una memoria più lunga non migliora i risultati e, in molti casi, li peggiora. Inoltre, il dato relativo all'istante temporale può essere eliminato, senza notare differenze nel comportamento della rete.

Per velocizzare la procedura di training, e ottimizzare la distribuzione degli ingressi, è opportuno mapparli "manualmente" in un piccolo intervallo centrato attorno allo 0 (in modo che la tangente sigmoidale, ovvero la funzione d'attivazione dei neuroni nascosti della rete, abbia comportamento quasi lineare in questo intervallo). Come già visto in precedenza, una mappatura lineare nell'intervallo $[-1, 1]$ non produce risultati ottimi, poichè, a causa di sporadiche misurazioni glicemiche molto alte, i valori mappati risultano essere quasi tutti negativi (circa il 90% degli ingressi mappati è minore di 0). Viene quindi effettuata una mappatura che riporta le glicemie registrate a media nulla e standard deviation pari a 0.5. In questo modo si ottiene una distribuzione più equa di ingressi e uscite della rete.

Il segnale glicemico, come la maggior parte dei segnali biologici, varia molto lentamente. Di conseguenza, i 15 valori relativi alle ultime 15 misurazioni sono tutti simili tra loro. Ciò potrebbe confondere la rete, poichè essa non riesce a pesare adeguatamente il significato dei vari ingressi. Inoltre, la glicemia non cambia in maniera casuale, ma presenta dei trend definiti, alternando tratti in salita a tratti in discesa. Oltre al valore puntuale che essa assume, è particolarmente interessante la pendenza registrata nell'ultimo intervallo temporale. L'informazione sulla pendenza è contenuta nelle differenze prime del segnale. Si è quindi proceduto fornendo in ingresso alla rete solamente l'ultimo valore glicemico registrato e le 15 differenze prime, relative agli ultimi 16 valori misurati. Anche il target è stato modificato, sostituendo al valore glicemico predetto a PH minuti l'offset, a PH minuti, rispetto all'ultimo valore registrato. Ovvero, il nuovo target della rete è

$$\text{dBGL}_{\text{target}}(\mathbf{t}) = \text{BGL}(\mathbf{t} + \mathbf{PH}) - \text{BGL}(\mathbf{t}) \quad (5.1)$$

Il passaggio alle derivate del segnale ha due vantaggi:

1. Le differenze prime del segnale glicemico sono abbastanza scorrelate tra loro, mentre il segnale è altamente correlato, proprio in virtù della sua natura smooth. La rete neurale predilige ingressi scorrelati, in

5.1 Approccio 1: rete neurale feedforward per la predizione del livello glicemico

modo da riuscire a pesarli in maniera diversa, in base alla loro importanza. Gli ingressi scorrelati, inoltre, portano una quantità maggiore di informazione, eliminando la ridondanza.

2. Le derivate prime del segnale glicemico sono distribuite simmetricamente attorno allo zero, in modo quasi gaussiano, (come si può osservare in Figura 5.4). Di conseguenza i problemi legati alla mappatura di ingressi e uscite della rete vengono automaticamente eliminati.

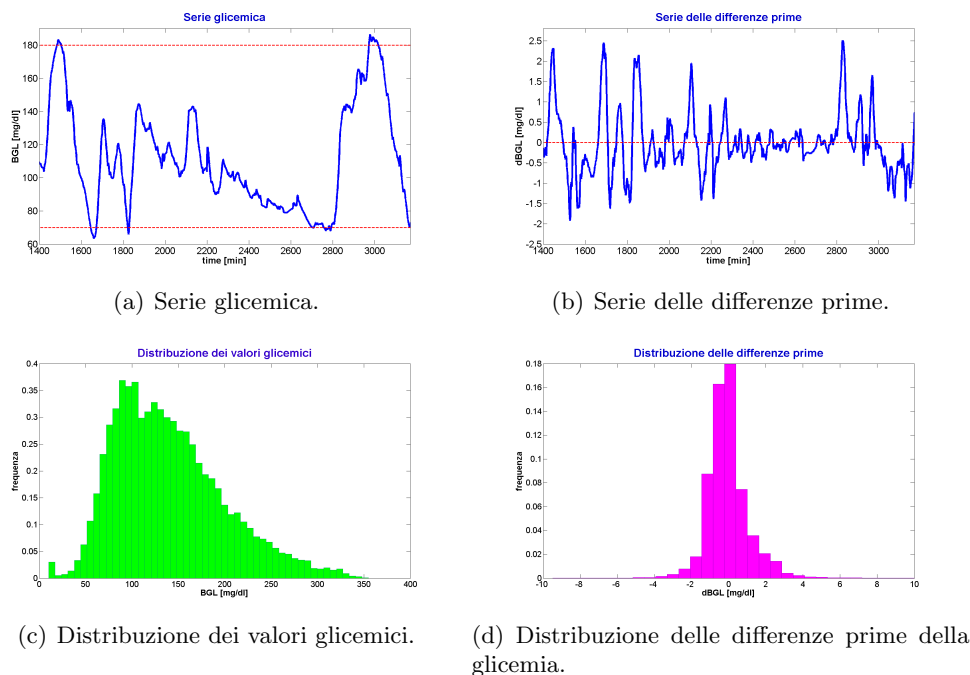


Figura 5.4: Confronto tra la distribuzione dei valori glicemici (5.4(c)) e la distribuzione delle loro differenze prime (5.4(d)), ottenute a partire, rispettivamente, dal segnale glicemico (5.4(a)) e dalla sua derivata (5.4(b)).

Oltre ai vantaggi, però, le derivate prime portano anche dei problemi, tra i quali i principali sono la grande sensibilità al rumore di misura e l'amplificazione del rumore stesso, intrinseca all'operazione di derivazione. Ciò si riflette in una irregolarità del profilo predetto non trascurabile. Per limitare le oscillazioni ad altissima frequenza, dovute all'amplificazione del rumore, l'informazione portata dalle ultime 15 differenze prime del segnale è stata condensata nel loro valore medio. Questa scelta è giustificata anche dall'osservazione del comportamento della rete con vari vettori d'ingresso: essa tende a seguire l'ultimo dato visto e non è in grado di discriminare l'importanza dei 15 valori passati, in base alla loro distanza temporale dall'istante attuale. La predizione ottenuta, fornendo in ingresso alla rete solamente l'ultimo valore glicemico registrato, e la media delle ultime 15

differenze prime, risulta molto più regolare. Il prezzo da pagare è la tolleranza di un ritardo leggermente superiore, dovuto al fatto che, mediando il valore delle ultime 15 differenze prime, forniamo in ingresso alla rete un'informazione sul trend glicemico ritardata rispetto a quella che si avrebbe utilizzando tutti i valori puntuali. La Figura 5.5 confronta l'uscita della rete con in ingresso le 15 differenze prime del segnale (profilo verde), con l'uscita della rete che ha in ingresso la media delle ultime 15 differenze prime (profilo magenta). La predizione della rete che ha in ingresso le 15 differenze prime risulta molto irregolare. Il profilo magenta, prodotto dalla rete che ha in ingresso la media delle differenze prime, presenta molte meno oscillazioni, ma è visibilmente ritardato rispetto a quello verde.

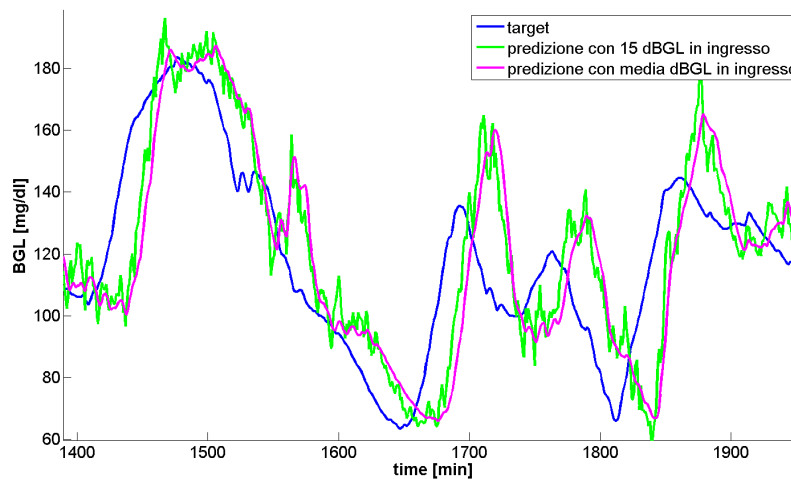


Figura 5.5: Confronto tra l'uscita della rete con 15 differenze prime in ingresso e l'uscita della rete con la media delle differenze prime in ingresso.

Informazioni riguardanti i carboidrati ingeriti

Come è noto dalla fisiologia, la quantità di carboidrati ingeriti è strettamente correlata con il livello glicemico. Sarebbe perciò opportuno rendere disponibili alla rete informazioni riguardanti i pasti. Bisogna, quindi, decidere quale sia la forma migliore per riassumere quest'informazione e renderla utilizzabile dalla rete neurale. Dato che la rete feedforward non ha memoria non avrebbe senso fornirle in ingresso la quantità di carboidrati ingeriti solamente in concomitanza al pasto e, successivamente, azzerare l'ingresso. L'effetto dei carboidrati, infatti, non è immediato e, soprattutto, non si esaurisce appena termina il pasto. Si potrebbe pensare di dividere la quantità di carboidrati ingeriti per il numero di minuti durante i quali essi influenzeranno la glicemia e, ad ogni istante, fornire in ingresso alla rete solo l'informazione riguardante una frazione del pasto totale. Questa solu-

zione è profondamente sbagliata, poichè l'effetto del pasto non è equamente distribuito durante tutto l'intervallo necessario per il suo assorbimento. Si è quindi deciso di utilizzare la *rate of appearance* (ra) dei carboidrati ingeriti, ricavata grazie ad un generatore [9] al quale vengono fornite in ingresso informazioni sui pasti e su alcuni parametri fisiologici del soggetto. Il modello utilizzato per generare la rate of appearance è basato su equazioni differenziali, che descrivono il transito di glucosio attraverso il tratto gastrointestinale, la sua produzione endogena e la sua utilizzazione sia insulino-dipendente che insulino-indipendente, nonché la cinetica del glucosio, durante il passaggio dal plasma ai tessuti. Il modello è stato identificato su dati di pazienti diabetici ospedalizzati, ed ogni pasto è stato identificato separatamente. Inoltre, per cercare di tenere conto della variabilità circadiana di alcuni parametri, essi vengono modificati anche nel generatore, in base al pasto in esame.

Operativamente, per simulare la rate of appearance dei carboidrati ingeriti, ad ogni istante temporale si fornisce in ingresso alla funzione generatrice un vettore di stato, che riassume la storia passata del sistema (per quanto riguarda i nutrienti assunti), e la quantità di carboidrati ingeriti nell'istante stesso (questo valore sarà quasi sempre pari a 0). La funzione genera, quindi, la predizione della rate of appearance del pasto, dall'istante attuale, con l'orizzonte temporale richiesto.

La rate of appearance dei carboidrati può essere generata on-line. È infatti sufficiente fornire alla funzione la quantità di carboidrati assunti. Per tener conto dell'informazione sui pasti sono stati aggiunti ulteriori 4 ingressi alla rete neurale:

- Il valore della rate of appearance dei carboidrati in $t + PH$, ovvero in corrispondenza dell'istante temporale in cui si desidera predire il livello di glucosio;
- Le 3 differenze prime della rate of appearance, equispaziate nell'intervallo $[t, t + PH]$.

In tal modo si fornisce in ingresso alla rete un'informazione su come ci si aspetta che sia l'andamento della rate of appearance futura e la sua variazione nell'intervallo temporale che va dall'istante attuale all'orizzonte di predizione desiderato. In Figura 5.6 è rappresentata la curva della rate of appearance dei carboidrati ingeriti nell'arco di una giornata.

Grazie all'introduzione di questi nuovi ingressi le performance della rete migliorano sensibilmente, soprattutto in corrispondenza dei trend glicemici crescenti. L'aumento di glucosio nel sangue e, quindi, nell'interstizio, è dovuto proprio al pasto ed è difficilmente predicibile dai campioni glicemici passati, nel caso manchi quest'informazione.

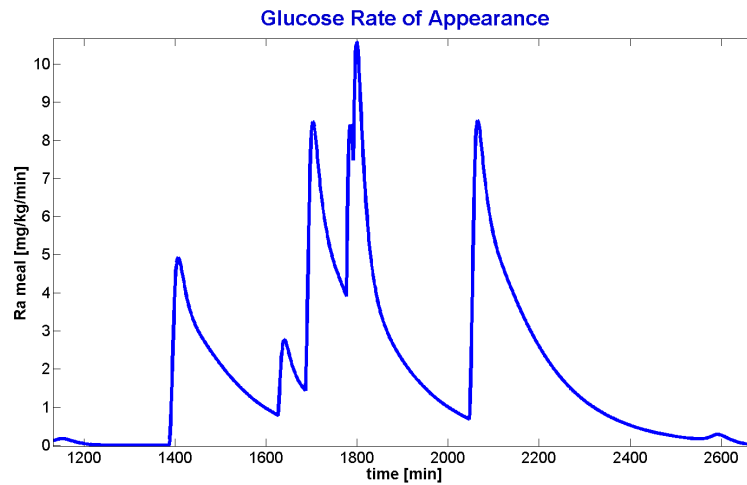


Figura 5.6: Rate of appearance dei carboidrati ingeriti nell'arco di una giornata.

Informazioni riguardanti le dosi di insulina somministrata

Per sopperire alla deficienza insulinica e mantenere la glicemia il più possibile entro il range euglicemico i pazienti diabetici insulino dipendenti vengono trattati con iniezioni subcutanee di insulina. Nel tentativo di mimare il comportamento di un pancreas sano, in concomitanza ai pasti vengono iniettati boli di analoghi rapidi dell'ormone, la cui azione inizia circa 30 minuti dopo l'iniezione e si esaurisce approssimativamente nell'arco di 3 ore. Inoltre, per garantire un livello basale di insulina in circolo, molti diabetici ricorrono ad una iniezione giornaliera di un analogo lento dell'insulina, che ha effetti ipoglicemizanti che si protraggono per circa 24 ore. In alternativa, è possibile utilizzare anche microinfusori, che rilasciano costantemente, goccia a goccia, insulina rapida.

È chiaro che il livello di glucosio nel sangue risente fortemente delle dosi di insulina iniettata. Perciò sarebbe importante fornire quest'informazione in ingresso alla rete neurale. Si ripropone, anche per l'insulina, così come per i carboidrati, il problema di quale sia la forma migliore per sintetizzare le informazioni disponibili sui boli somministrati e fornirle al predittore. L'effetto dell'insulina non è istantaneo e, soprattutto, non è linearmente distribuito durante tutto il suo periodo d'azione. Appare quindi sensato utilizzare, come già fatto per i carboidrati, la *rate of appearance* dell'insulina iniettata. Il simulatore [9] descrive la cinetica dell'insulina e l'assorbimento subcutaneo attraverso un modello bicompartimentale, i cui parametri sono stati identificati su dati di 11 soggetti diabetici di tipo 1 ospedalizzati. Il procedimento è analogo a quello seguito per generare la rate of appearance dei carboidrati. Ad ogni istante temporale si forniscono in ingresso alla funzione un vettore di stato che riassume la storia del sistema,

e la quantità di insulina iniettata. In uscita si ottiene la predizione della concentrazione di insulina plasmatica nell'intervallo $[t_k, t_k + T]$, ovvero dall'istante attuale t_k con l'orizzonte temporale T richiesto. La funzione sa gestire la generazione di profili insulinici sia in pazienti che utilizzano terapia multi-iniettiva (ovvero analoghi dell'insulina sia ad azione rapida, sia ad azione lenta), sia in pazienti che utilizzano microinfusore (ovvero solo analoghi ad azione rapida, ma somministrati in parte in modo continuo e in parte attraverso boli in concomitanza dei pasti). In pratica, ad ogni istante temporale vengono simulati due profili, uno relativo alla concentrazione di insulina rapida, l'altro relativo alla concentrazione di insulina lenta. Il profilo totale di insulina plasmatica è la somma dei due.

Dopo aver generato la rate of appearance dell'insulina iniettata, sono state testate varie configurazioni di ingressi, che differivano principalmente per i ritardi applicati al profilo generato. Sono state inizialmente considerate la rate of appearance attuale e 3 campioni delle sue derivate passate distribuite negli ultimi PH minuti, e la rate of appearance futura relativa a $t + PH$ e 3 campioni delle sue derivate equispaziate nell'intervallo $[t, t + PH]$. Successivamente, sono state considerate anche la rate of appearance ritardata di 60 minuti rispetto all'istante di predizione, (ovvero, con $PH=30$ minuti, la rete ha in ingresso la rate of appearance registrata 30 minuti prima dell'istante attuale), e la rate of appearance anticipata di 30 minuti rispetto all'istante di predizione, (ovvero, con $PH=30$ minuti, essa corrisponde alla Ra che si registrerà dopo 60 minuti rispetto all'istante attuale). In tal modo si è cercato di tener conto di diversi valori per il ritardo d'azione dell'insulina. In nessun caso, però, si è ottenuto un miglioramento significativo dei risultati. Di conseguenza si è deciso di non includere le informazioni riguardanti le dosi di insulina tra gli ingressi.

L'incapacità della rete neurale di utilizzare adeguatamente le informazioni relative all'insulina sono probabilmente strettamente correlate al meccanismo d'azione dell'insulina stessa. Per esempio, l'effetto quantitativo dell'ormone sul livello glicemico dipende dalla sensibilità/resistenza insulinica, un parametro strettamente legato all'indice di massa corporea. La sensibilità insulinica varia sia in base al soggetto, sia, in uno stesso soggetto, in base al momento della giornata, presentando un andamento circadiano. Il ritardo d'azione dell'insulina è un ulteriore parametro non trascurabile. Esso presenta variabilità interpersonale, ma anche intrapersonale, dipende infatti dalla zona corporea in cui viene iniettato il bolo di insulina, dalla quantità di insulina residua in circolo, dal momento della giornata e da molti altri fattori di cui si può difficilmente tenere conto. Infine, un altro elemento di rilievo è il fatto che i boli di insulina vengono iniettati in concomitanza dei pasti. Di conseguenza, il livello glicemico inizialmente aumenta e, solo successivamente, quando l'effetto dei carboidrati inizia ad esaurirsi e l'insulina raggiunge il suo picco d'azione, il suo effetto ipoglicemizzante riporta il valore di glucosio nel sangue entro il range euglicemico. La rete neurale

associa erroneamente la rate of appearance dell'insulina all'aumento della glicemia, che effettivamente si osserva appena viene iniettato l'analogo dell'ormone. Per questo motivo la nuova informazione non solo non è sfruttata adeguatamente, ma è male interpretata, col rischio di peggiorare la qualità della predizione finale.

In Figura 5.7 e in Figura 5.8 è evidente il problema connesso all'interpretazione dell'insulina da parte della rete neurale. In Figura 5.7 sono rappresentate la curva glicemica, la rate of appearance dei carboidrati ingeriti e la rate of appearance dell'insulina iniettata, tutte relative allo stesso istante temporale. È evidente come l'andamento della Ra dell'insulina (curva verde tratteggiata) sia molto simile all'andamento della Ra dei carboidrati (curva rossa). Inoltre, l'aumento della concentrazione insulinica nel sangue coincide con l'aumento glicemico (a causa della concomitanza con i pasti). In Figura 5.8 sono rappresentate la curva glicemica e la rate of appearance dell'insulina, relativa a 2 ore dopo. Grazie all'introduzione di questo ritardo il secondo e il terzo picco insulinico (cerchiati in verde) corrispondono a trend di discesa glicemica; il primo e il quarto picco (cerchiati in rosso), però, non risultano ancora sufficientemente ritardati da correlare adeguatamente con la diminuzione del livello di glucosio nel sangue. Risulta evidente la difficoltà incontrata nell'inserire l'informazione dell'insulina tra gli ingressi della rete, evitando che essa sia interpretata erroneamente.

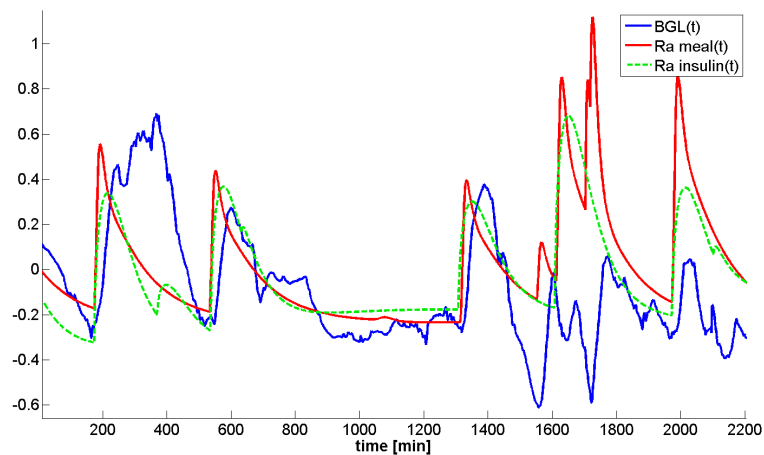


Figura 5.7: Plot di $BGL(t)$, $Ra\ meal(t)$ e $Ra\ insulin(t)$. I valori sono stati scalati in modo da rendere le curve confrontabili.

5.1 Approccio 1: rete neurale feedforward per la predizione del livello glicemico

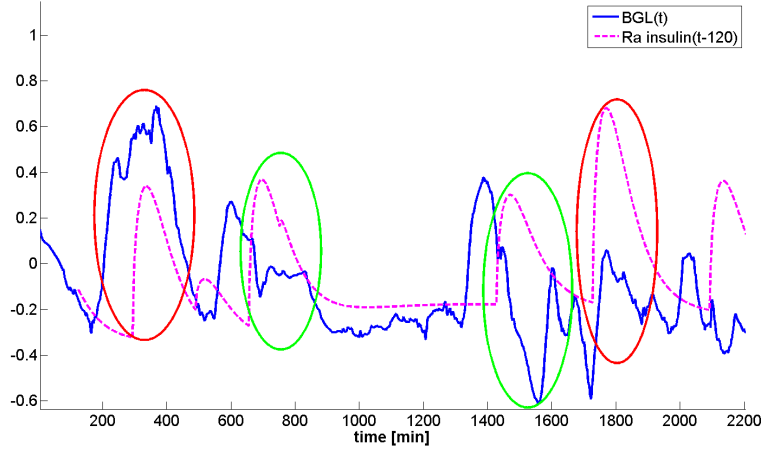


Figura 5.8: Plot di $BGL(t)$ e $Ra\ insulin(t-120)$. I valori sono stati scalati in modo da rendere le curve confrontabili.

Deviazione standard del segnale glicemico e delle sue differenze prime

La *deviazione standard* di una variabile aleatorio \mathbf{X} è statisticamente definita come

$$sd(\mathbf{X}) = \sqrt{E[(\mathbf{X} - E[\mathbf{X}])^2]} \quad (5.2)$$

dove $E[\mathbf{X}]$ è l'aspettazione di \mathbf{X} . Poichè nella realtà solitamente non è nota la distribuzione di probabilità di \mathbf{X} , la deviazione standard è operativamente stimata, in modo campionario, utilizzando la formula

$$\hat{sd}(\mathbf{X}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (5.3)$$

oppure la formula

$$\hat{sd}(\mathbf{X}) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (5.4)$$

dove

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (5.5)$$

è la stima del valor medio di \mathbf{X} e N è il numero di realizzazioni disponibili di \mathbf{X} . Questo indice misura la dispersione di un insieme di dati rispetto alla loro media, ed è tanto più alto quanto più i dati differiscono dal valor medio.

In ingresso alla rete viene fornito il valore medio delle differenze prime del segnale glicemico negli ultimi 15 minuti, perdendo in tal modo la variabilità del trend stesso. La deviazione standard delle differenze prime degli ultimi 15 minuti rappresenta proprio quest'informazione. È stata quindi aggiunta questa nuova grandezza tra gli ingressi della rete. Inoltre, poichè si utilizza per la predizione solamente il livello di glucosio attuale, si è pensato di riassumere la storia recente dei valori glicemici calcolandone la deviazione standard relativa agli ultimi 15 minuti. Queste nuove informazioni non hanno portato a nessun miglioramento delle capacità predittive della rete e, quindi, non sono state aggiunte agli ingressi effettivamente utilizzati.

Area sotto la curva della rate of appearance

L'*area sotto la curva* (Area Under the Curve, AUC), relativa alla concentrazione plasmatica di una certa sostanza, rappresenta una misura dell'esposizione dell'organismo alla sostanza stessa. Quindi, le AUCs, relative alla rate of appearance dei carboidrati ingeriti e alla rate of appearance dell'insulina somministrata, riassumono la storia e l'effetto dei relativi segnali nell'intervallo di tempo in cui vengono calcolate. Potrebbero, quindi, essere un parametro più adeguato da inserire tra gli ingressi della rete, rispetto al valore puntuale della rate of appearance. Sono state considerate l'AUC relativa alla Ra del pasto e l'AUC relativa alla Ra dell'insulina, calcolate ad ogni istante t_k considerando il tratto di curva compreso nell'intervallo $[t_k, t_k + PH]$. Questo nuovo segnale, che rappresenta l'integrale definito tra i due punti della curva rispetto al quale è calcolato, è molto simile alla curva stessa, come si può notare in Figura 5.9, dove si vede chiaramente, in entrambi i plot, come l'AUC risulti semplicemente una versione leggermente ritardata dell'Ra. Poichè questo ingresso non ha migliorato le performance della rete, né sostituendolo, né affiancandolo alla rate of appearance, non è stato introdotto tra gli ingressi della rete neurale.

5.1 *Approccio 1: rete neurale feedforward per la predizione del livello glicemico*

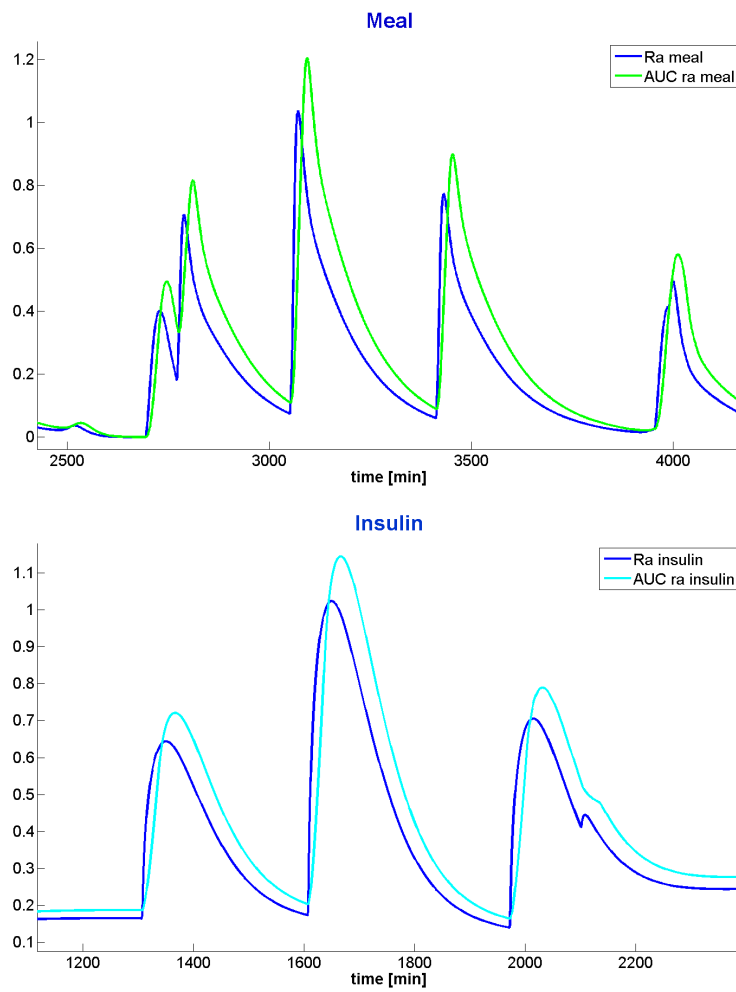
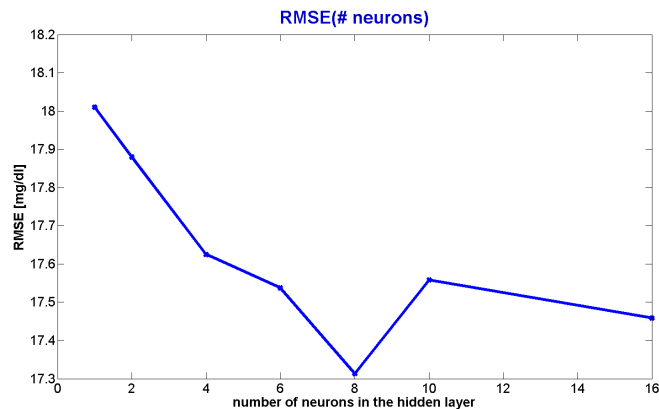


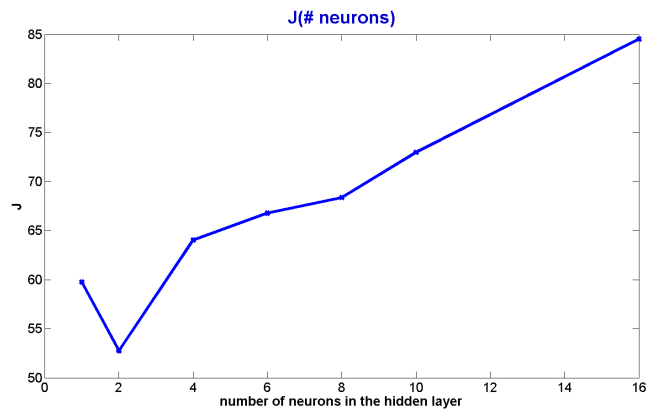
Figura 5.9: *Plot dell’Ra e dell’AUC a confronto. Le grandezze sono state scalate per essere rese confrontabili.*

5.1.3 Ri-ottimizzazione del numero di neuroni nascosti

Come affermato in precedenza, dopo aver scelto gli ingressi effettivamente significativi per la rete neurale, è stata eseguita nuovamente l'ottimizzazione della struttura della rete stessa. L'analisi visiva e gli indici utilizzati per descrivere le performance della rete hanno confermato che una rete con un unico strato con 8 neuroni nascosti è da preferire sia ad architetture più semplici, sia ad architetture più elaborate.



(a) RMSE in funzione del numero di neuroni nascosti.



(b) J in funzione del numero di neuroni nascosti.

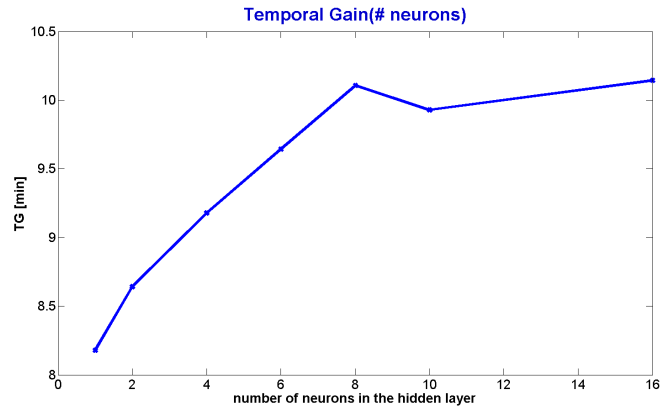
Figura 5.10: Andamento del RMSE e di J all'aumentare del numero di neuroni nascosti.

L'andamento del RMSE, riportato in Figura 5.10(a), porta a scegliere la rete con 8 neuroni nascosti. I valori di J, riportati in Figura 5.10(b), come osservato anche in precedenza, sono più bassi per reti con pochi neuroni, poichè esse producono profili molto simili a spezzoni lineari. Si può comunque notare che per architetture con più di 8 neuroni J peggiora sensibilmente rispetto al valore ottenuto in corrispondenza a ffnn-8-1.

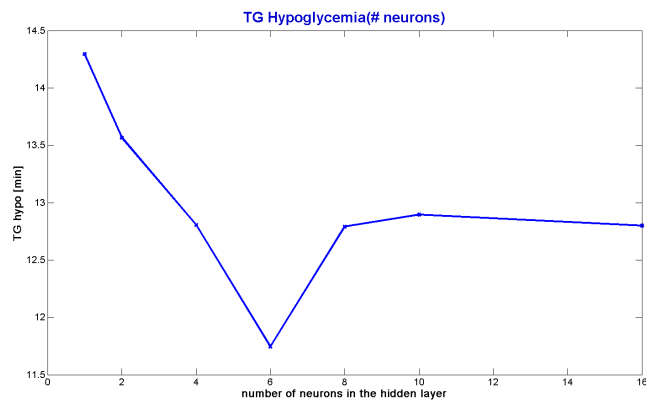
Il guadagno temporale è più difficile da interpretare. Il TG medio raggiunge il suo valore massimo in corrispondenza della rete con 8 neuroni e poi si stabilizza, come si può osservare in Figura 5.11(a). Il TG relativo alle ipoglicemie risulta molto alto per reti con pochi neuroni, ma, in realtà, le predizioni ottenute con questi semplici modelli sono caratterizzate da molti falsi attraversamenti della soglia ipoglicemica. Questo indice, in corrispondenza di *ffnn-8-1* raggiunge un secondo punto di massimo e successivamente si stabilizza a tale valore. Il suo andamento in funzione del numero di neuroni nascosti è riportato in Figura 5.11(b). Il TG relativo alle iperglicemie, riportato in Figura 5.11(c), ha un massimo locale in corrispondenza di *ffnn-8-1*.

In Figura 5.12 sono riportati i profili predetti da reti neurali con un numero crescente di neuroni nascosti. Osservando i plots sovrapposti si nota che il profilo verde (corrispondente alla predizione della rete neurale con 8 neuroni nascosti) risulta in anticipo, sia nei trend in salita che in quelli in discesa, rispetto ai profili arancione e nero (corrispondenti alle predizioni della rete neurale con un unico neurone e alla predizione della rete neurale con 4 neuroni). Il profilo magenta, relativo alla predizione della rete con 16 neuroni nascosti, ha un ritardo confrontabile col profilo verde, ma è più irregolare.

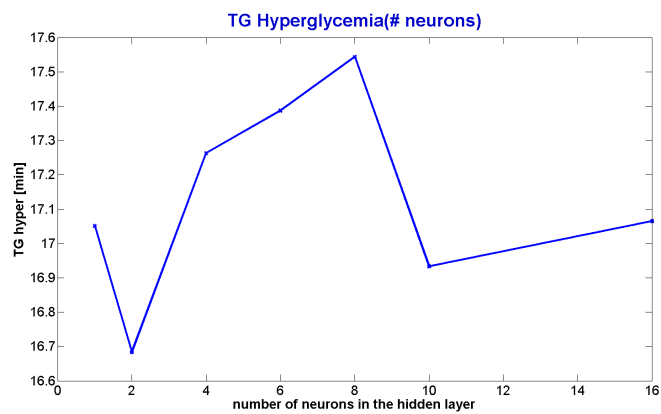
CAPITOLO 5. Due nuovi approcci basati su reti neurali per predire la glicemia futura



(a) Guadagno temporale medio.



(b) Guadagno temporale relativo alla predizione di ipoglicemie.



(c) Guadagno temporale relativo alla predizione di iperglicemie.

Figura 5.11: Andamento del guadagno temporale all'aumentare del numero di neuroni nascosti.

5.1 *Approccio 1: rete neurale feedforward per la predizione del livello glicemico*

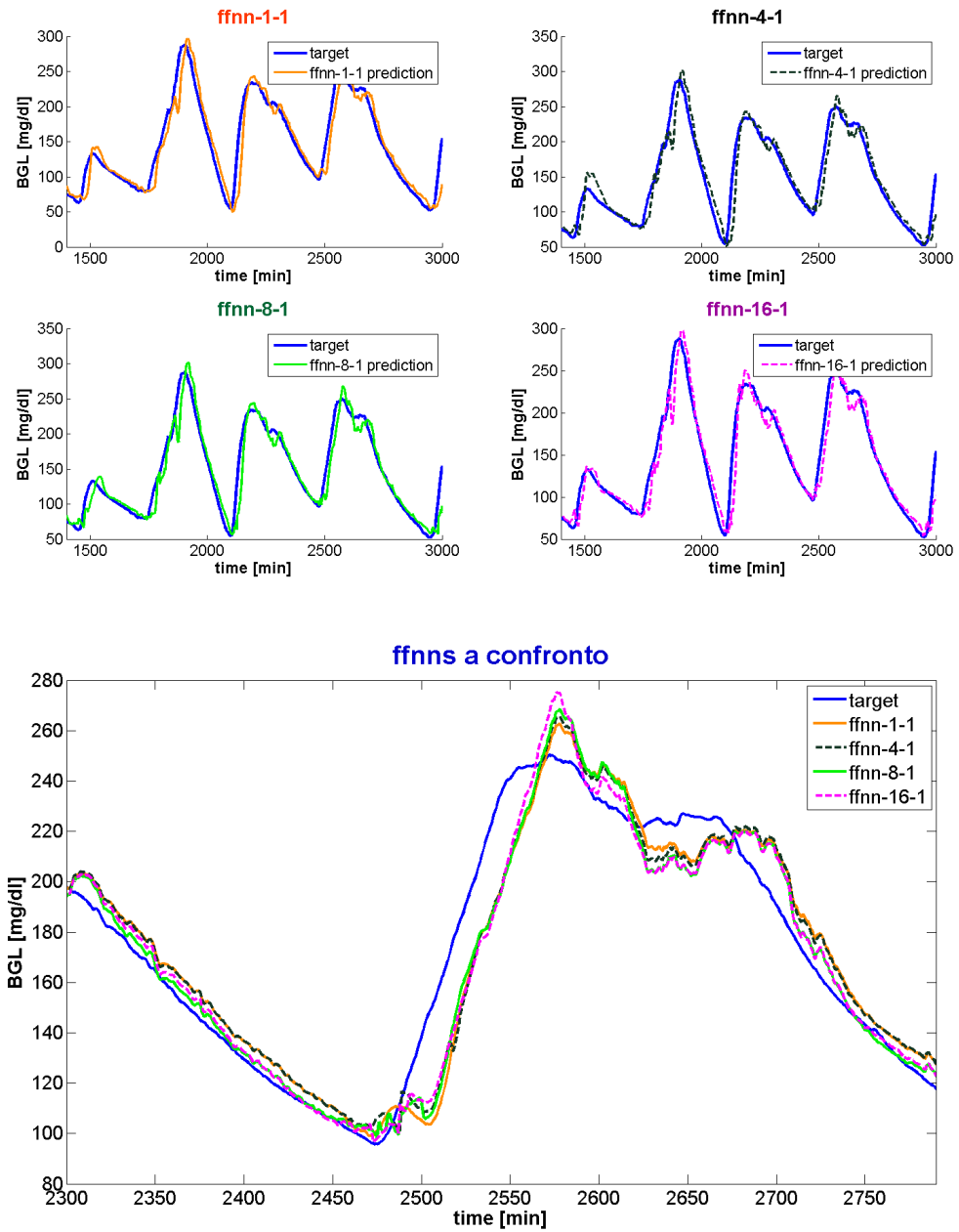


Figura 5.12: *Profili predetti da reti neurali con un diverso numero di neuroni nascosti.*

5.2 Approccio 2: rete neurale feedforward per correggere l'errore del modello lineare

Il segnale glicemico, benchè presenti dinamiche non lineari, ha un andamento che può essere ben descritto anche con un modello lineare tempo variante, come dimostrato da Sparacino *et al.* in [39]. La rete neurale, nel tentativo di predire la concentrazione di glucosio futura, deve quindi modellare sia la parte lineare, sia quella non lineare del segnale. La si potrebbe sollevare dall'onere di modellare l'andamento lineare, che può essere predetto adeguatamente utilizzando per l'appunto modelli lineari, e concentrarsi solo sulla predizione della parte non lineare del segnale.

Si è quindi pensato di spostare l'attenzione dalla predizione della serie glicemica alla predizione della serie dell'errore del modello lineare, in cui sono teoricamente racchiuse tutte e sole le dinamiche non lineari del segnale.

L'uscita predetta dal modello lineare a PH minuti è data da

$$\widehat{BGL}(t + PH) = a_1 \cdot (t + PH) + a_2 \quad (5.6)$$

dove a_1 e a_2 sono, rispettivamente, il coefficiente angolare e il termine noto della retta di regressione che interpola le ultime misure registrate dal sensore. I parametri del modello vengono stimati ad ogni istante temporale, utilizzando la tecnica dei minimi quadrati ricorsivi pesati. In particolare, il valore ottimo del *forgetting factor* μ viene scelto minimizzando l'indice J [43]

$$J = \frac{ESOD_{\text{norm}}}{TG_{\text{norm}}^2} \quad (5.7)$$

L'errore commesso dal modello lineare all'istante t , relativo alla predizione glicemica in $t + PH$, potrà essere calcolato solamente nell'istante $t + PH$, quando il vero valore glicemico potrà essere confrontato con quello predetto PH minuti prima e vale

$$e_{\text{LIN}}(t+PH) = \widehat{BGL}_{\text{LIN}}(t+PH) - BGL_{\text{CGM}}(t+PH) \quad (5.8)$$

dove $\widehat{BGL}_{\text{LIN}}(t+PH)$ rappresenta la predizione del modello lineare, effettuata nell'istante t con orizzonte di predizione PH , mentre $BGL_{\text{CGM}}(t+PH)$ rappresenta il valore che sarà effettivamente registrato dal sensore, tra PH minuti (tale valore, che è il target finale della predizione, è disponibile già all'istante t chiaramente solo in fase di training. Può quindi essere utilizzato per calcolare l'errore del predittore lineare). Per correggere la predizione glicemica del modello lineare è sufficiente sottrarre il suo errore, predetto dalla rete neurale. Naturalmente anche la predizione della rete neurale sarà affetta da errore, quindi non è detto che il valore di glucosio ottenuto correggendo il modello lineare sia effettivamente quello che sarà osservato dopo PH minuti.

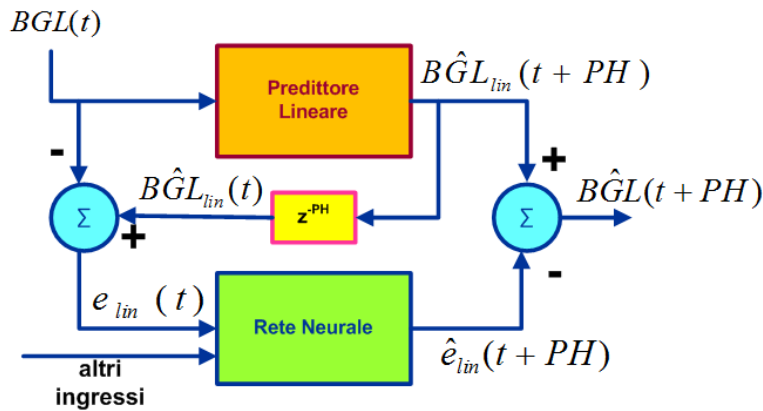


Figura 5.13: Schematizzazione del sistema realizzato nell'approccio 2 per predire la glicemia futura.

In Figura 5.13 è schematizzato il sistema di predizione della glicemia messo a punto nell'approccio 2. Il valore glicemico predetto è ottenuto combinando l'uscita del modello lineare con l'uscita della rete neurale che ne predice l'errore.

La struttura e gli ingressi della rete sono stati ottimizzati attraverso la procedura descritta nella Sezione 4.2 (pag. 82). La rete migliore è risultata essere nuovamente una rete con un solo strato di 8 neuroni nascosti, con funzione d'attivazione tangente sigmoideale, ed un neurone d'uscita con funzione d'attivazione lineare. Gli ingressi della rete sono:

- l'errore del modello lineare, all'istante attuale, calcolato come in (5.8);
- la media delle differenze prime dell'errore, negli ultimi 15 minuti;
- la rate of appearance dei carboidrati ingeriti, all'istante in cui si vuole predire l'errore (ovvero a $t + PH$) e le sue 3 derivate, equispaziate nell'intervallo $[t, t + PH]$;
- il livello glicemico registrato dal sensore all'istante attuale;
- la media delle differenze prime della concentrazione di glucosio negli ultimi 15 minuti.

L'uscita della rete è la differenza tra l'errore che commetterà il modello lineare in $t + PH$ e l'errore attuale. L'inserimento del livello glicemico e della media delle sue derivate tra gli ingressi hanno portato un miglioramento visibile della predizione. In loro assenza, infatti, la rete neurale non riusciva a modellare adeguatamente l'errore lineare. Anche in questo caso, informazioni relative all'insulina, all'area sotto la curva della rate of appearance di carboidrati e insulina, alla deviazione standard dell'errore del modello lineare e delle sue derivate, e alla deviazione standard del livello glicemico

e delle sue derivate non hanno migliorato significativamente le performance della rete. Di conseguenza non sono state inserite tra gli ingressi.

5.2.1 Potenzialità e problemi della rete per la correzione dell'errore del modello lineare

L'uscita effettiva della rete che predice direttamente il livello glicemico è la variazione glicemica che si avrà, tra **PH** minuti, rispetto all'ultimo valore registrato dal sensore. Per ottenere il valore di glucosio effettivo è sufficiente sommare alla predizione attuale il valore attuale registrato dal sensore.

$$dBGL(t) = BGL(t+PH) - BGL(t) \quad (5.9)$$

Quindi

$$\widehat{BGL}(t+PH) = d\widehat{BGL}(t) + BGL(t) \quad (5.10)$$

dove $d\widehat{BGL}(t)$ e $\widehat{BGL}(t+PH)$ rappresentano i valori predetti e, quindi, affetti da errore di predizione, mentre $BGL(t)$ è il valore effettivamente registrato dal sensore all'istante attuale.

Analogamente, l'uscita effettiva della rete che predice l'errore del modello lineare è la differenza tra l'errore della predizione lineare attuale (relativa all'istante $t + PH$) e l'errore relativo all'istante attuale t (calcolato tenendo in memoria la predizione effettuata **PH** minuti prima).

$$de_{LIN}(t) = e_{LIN}(t+PH) - e_{LIN}(t) \quad (5.11)$$

con $e_{LIN}(t+PH)$ calcolato come in (5.8):

$$e_{LIN}(t+PH) = \widehat{BGL}_{LIN}(t+PH) - BGL_{CGM}(t+PH)$$

Per ottenere l'errore effettivo che si sta commettendo col modello lineare è sufficiente sommare all'uscita della rete l'errore relativo all'istante attuale.

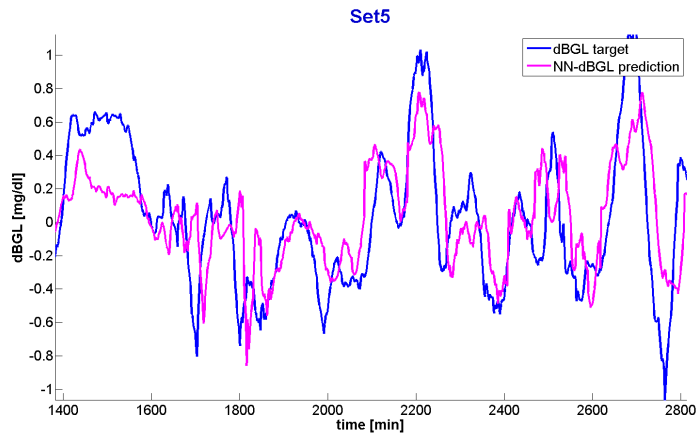
$$\widehat{e}_{LIN}(t+PH) = d\widehat{e}_{LIN}(t) + e_{LIN}(t) \quad (5.12)$$

dove $\widehat{e}_{LIN}(t+PH)$ è l'errore predetto e $d\widehat{e}_{LIN}(t+PH)$ è l'uscita della rete, e sono entrambi affetti da errore di predizione, mentre $e_{LIN}(t)$ è il vero errore della predizione del modello lineare relativa a **PH** minuti prima. Successivamente, per ottenere la predizione glicemica corretta bisogna sottrarre l'errore così ottenuto al valore predetto dal modello lineare.

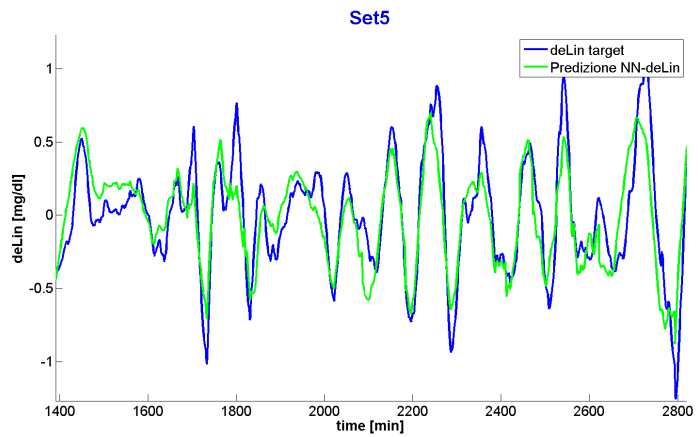
$$\begin{aligned} \widehat{BGL}_{CGM}(t+PH) &= \widehat{BGL}_{LIN}(t+PH) - \widehat{e}_{LIN}(t+PH) \\ &= \widehat{BGL}_{LIN}(t+PH) - d\widehat{e}_{LIN}(t) - e_{LIN}(t) \end{aligned} \quad (5.13)$$

In Figura 5.14 sono raffigurati l'uscita effettiva della rete che predice direttamente il livello glicemico (Figura 5.14(a)), l'uscita effettiva della rete

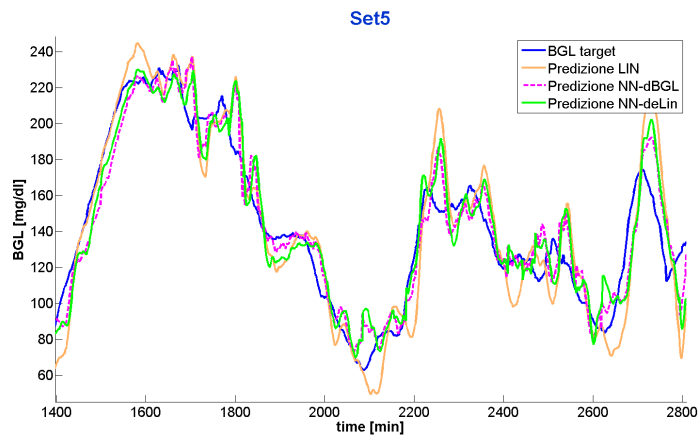
5.2 Approccio 2: rete neurale feedforward per correggere l'errore del modello lineare



(a) Uscita effettiva della rete che predice direttamente il livello glicemico.



(b) Uscita effettiva della rete che predice l'errore del modello lineare.



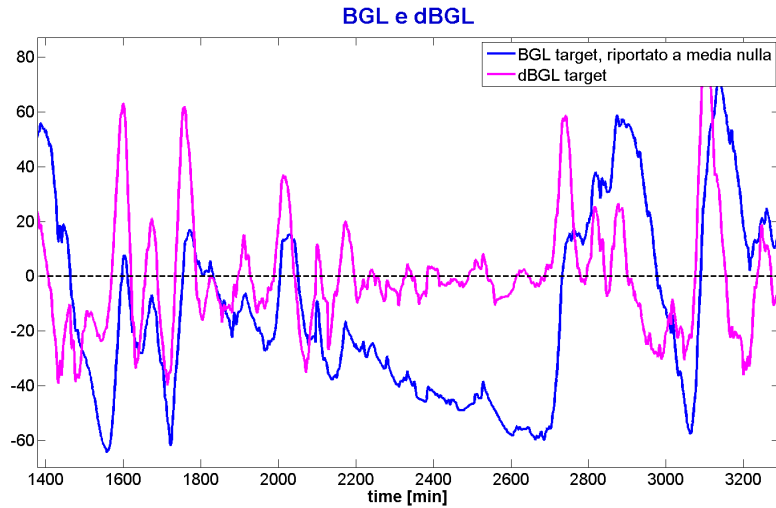
(c) Livello glicemico effettivo.

Figura 5.14: Confronto tra le performance della rete che predice direttamente il livello glicemico e le performance della rete che predice l'errore del modello lineare.

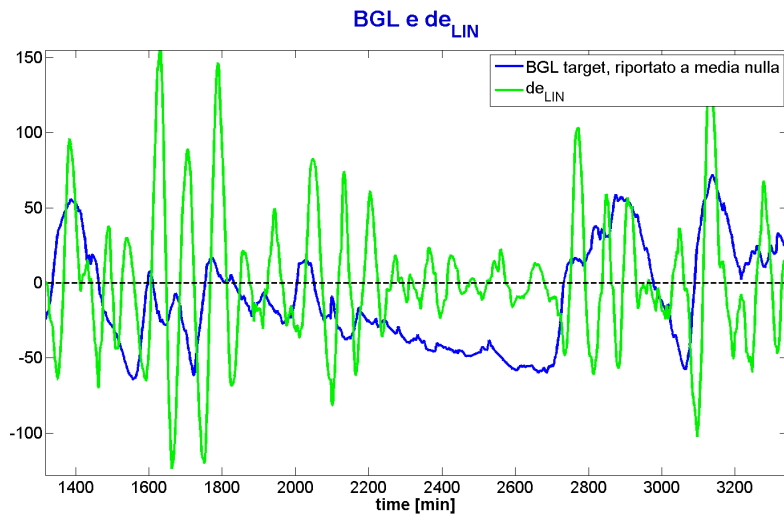
che predice l'errore del modello lineare (Figura 5.14(b)) e il confronto tra il livello di glucosio predetto utilizzando le due reti (Figura 5.14(c)). Come si può notare, l'uscita effettiva della rete che predice l'errore del modello lineare (approccio 2) segue abbastanza fedelmente l'andamento del segnale target. La bontà della predizione del livello di glucosio effettivo è però paragonabile a quella ottenuta predicendo direttamente l'offset glicemico. Ciò è dovuto al fatto che l'errore del modello lineare e le sue differenze prime (l'uscita effettiva della rete può essere considerata una differenza prima calcolata con passo PH minuti) assumono valori anche molto elevati, in modulo, rispetto al valore glicemico target da predire. Di conseguenza, un piccolo errore di predizione pesa molto nel calcolo del valore glicemico finale. La rete che predice direttamente il valore glicemico (approccio 1) commette errori di predizione maggiori. Ciò nonostante, poichè il segnale glicemico è regolare, le sue differenze prime calcolate con passo PH minuti (ovvero l'uscita effettiva di questa rete) hanno un valore piccolo, in modulo, rispetto alla glicemia target. Quindi, gli errori di predizione hanno un peso limitato. Perciò, sebbene questa rete produca un'uscita effettiva di qualità inferiore rispetto all'altra, la predizione del livello glicemico futuro (che è la quantità a cui siamo effettivamente interessati) ottenuta con le due reti non presenta differenze significative. Il fenomeno appena descritto è rappresentato in Figura 5.15. In particolare, in Figura 5.15(a) si può notare che il valore target della rete che predice l'offset glicemico futuro (curva magenta) ha un'ampiezza limitata rispetto al reale valore glicemico (curva blu), che è il target finale che si desidera stimare. Quindi, gli errori di predizione non sono particolarmente influenti sulla predizione glicemica finale, ottenuta sommando l'offset predetto (affetto da errore) all'ultimo valore registrato dal sensore (privo di errori). Dalla Figura 5.15(b), invece, si nota che per la rete neurale che corregge la predizione del modello lineare il valore target (curva verde) ha un'ampiezza paragonabile e spesso superiore al valore glicemico da predire (curva blu). Di conseguenza, piccoli errori di predizione dell'offset dell'errore deteriorano molto la predizione finale del livello glicemico.

Per sfruttare appieno le potenzialità della rete neurale che predice l'errore del modello lineare, si potrebbe minimizzare, in fase di training, una funzione obiettivo che penalizzi maggiormente gli errori di predizione in corrispondenza a valori target di modulo elevato. Per far ciò è però necessario utilizzare, per il training della rete, procedure diverse dall'algoritmo di back-propagation. Una possibile soluzione sono gli algoritmi stocastici introdotti in Sezione 2.4.4 (pag. 48). Poichè tali algoritmi non sono implementati nel Neural Network Toolbox™ di Matlab® 7.10.0, queste tecniche attualmente non sono ancora state investigate.

5.3 Approccio 2: rete neurale feedforward per correggere l'errore del modello lineare



(a) Confronto tra segnale glicemico reale, riportato a media nulla, e uscita target della rete che predice direttamente la glicemia (approccio 1).



(b) Confronto tra segnale glicemico reale, riportato a media nulla, e uscita target della rete che predice l'errore del modello lineare (approccio 2).

Figura 5.15: Confronto tra livello glicemico da predire e target effettivo delle reti neurali.

5.3 Conclusioni

In questo capitolo sono stati descritti due possibili approcci, basati su reti neurali, per predire la glicemia futura. Il secondo metodo presentato è molto promettente, ma per sfruttarne appieno le potenzialità è necessario investigare metodi alternativi per il training della rete neurale, minimizzando una funzione costo che pesi in maniera diversa gli errori di predizione, penalizzando maggiormente quelli commessi in corrispondenza di picchi più ampi del segnale target. Questi valori, infatti, corrispondono alle situazioni in cui il modello lineare produce una predizione notevolmente differente dal segnale glicemico. Di conseguenza questi sono i valori che necessitano più degli altri di essere corretti dalla rete neurale. In caso ciò non avvenga essi deteriorano notevolmente la bontà dell'intera predizione.

Nei prossimi capitoli saranno riportati e commentati i risultati ottenuti implementando e testando le due reti neurali proposte su un dataset simulato e, successivamente, su un dataset reale.

Capitolo 6

Implementazione e risultati su dati simulati

Le due reti descritte nel capitolo precedente sono state inizialmente implementate e testate su un dataset simulato e i risultati così ottenuti sono stati confrontati con quelli ottenuti con la rete proposta da Pérez-Gandía *et al.* in [34] e con quelli ottenuti col modello lineare tempo variante descritto da Sparacino *et al.* in [39].

6.1 Creazione del dataset

Per creare il dataset sono stati simulati i profili glicemici in 11 giorni consecutivi, di 20 pazienti virtuali affetti da diabete di tipo 1. Per ognuno dei 20 profili così creati sono stati selezionati 3 spezzoni della durata di 3 giorni ciascuno. Successivamente, i 60 spezzoni ottenuti in questo modo sono stati divisi in maniera random in training e validation set, composti, rispettivamente, da 28 e 12 serie temporali, e test set, costituito da 20 serie temporali.

6.1.1 Modello per la simulazione dei profili glicemici

Il modello utilizzato per simulare la glicemia di pazienti diabetici di tipo 1 è il modello descritto da Dalla Man *et al.* in [9], che risulta essere il miglior simulatore stato dell'arte. La riproduzione del funzionamento del sistema glucosio-insulina è stata costruita e validata sui dati relativi a 204 soggetti sani, studiati mediante pasto con triplo tracciante, una complessa procedura sperimentale che permette di ottenere i profili dei principali flussi di glucosio ed insulina in maniera quasi modello-indipendente. Usando concentrazioni e flussi è stato creato un modello del sistema glucosio-insulina con il metodo di decomposizione in sottosistemi e ricorrendo alla strategia delle funzioni forzanti, che minimizzano le incertezze strutturali nella defi-

nizione dei modelli dei singoli processi. L'identificazione del modello è stata eseguita nel soggetto medio, quindi, per poterlo utilizzare per descrivere la popolazione diabetica di tipo 1, ne è stata parzialmente modificata la struttura, rimuovendo il sotto-modello che descrive la secrezione insulinica e sostituendolo con un modulo per l'infusione subcutanea di analoghi dell'ormone. I parametri relativi a questo nuovo sotto modello sono stati identificati sui dati di 11 pazienti diabetici di tipo 1, monitorati per 3 giorni consecutivi durante una degenza in ospedale. Inoltre, affinché la simulazione sia in grado di fornire dati analoghi a quelli registrati da un sensore CGM, è stato aggiunto un modello della cinetica del glucosio sottocutaneo.

Il simulatore finale è costituito da 13 equazioni differenziali non lineari, che descrivono la cinetica del glucosio nel plasma e nei tessuti, il transito del glucosio nel tratto gastrointestinale, la produzione endogena di glucosio e la sua utilizzazione nei tessuti insulino dipendenti e in quelli insulino indipendenti (cervello e eritrociti), l'infusione subcutanea di insulina e la sua cinetica, l'esercizio fisico e, infine, la cinetica plasma-interstizio, (per poter simulare le misure di glucosio che sarebbero registrate da un sensore subcutaneo per il continuous glucose monitoring). I parametri del modello riproducono la variabilità inter e intraindividuale che si riscontra nella realtà e seguono dei trend circadiani. Inoltre, per simulare profili registrati da un sensore CGM il più veritieri possibile sono stati modellati anche l'errore di misura e gli errori di calibrazione.

6.1.2 Difetti e limiti dei dati simulati

Nonostante sia presente un modello dell'errore di misura e dell'errore di calibrazione del sensore, i profili simulati risultano comunque molto regolari e puliti rispetto a quelli reali. Per limitare questo problema è stato aggiunto alle serie temporali un segnale di rumore descritto attraverso un modello AR(1)

$$n(\mathbf{k}) = 0.95n(\mathbf{k} - \mathbf{1}) + w \quad (6.1)$$

con $w \sim N(\mathbf{0}, \mathbf{2})$, ovvero segnale gaussiano a media nulla e varianza pari a 2. La Figura 6.1 rappresenta un profilo glicemico simulato (in blu) e il profilo risultante dopo l'aggiunta del rumore, modellato come nell'equazione 6.1 (in verde). Non è stato utilizzato rumore puramente gaussiano, poiché i profili, dopo l'aggiunta di tale rumore, risultavano poco simili a quelli reali e, inoltre, la rete neurale riusciva ad ignorare la presenza del rumore stesso (cosa che non accade, invece, con i dati reali).

Un ulteriore inconveniente che si riscontra nei profili simulati è la quasi totale assenza di ipo e iperglicemie, che sono invece abbondanti nei profili reali. Quindi, per cercare di limitare tale problema, è stata variata in maniera random, in corrispondenza di alcuni pasti, la quantità di carboidrati tipicamente ingeriti. Inoltre sono state casualmente variate anche alcune

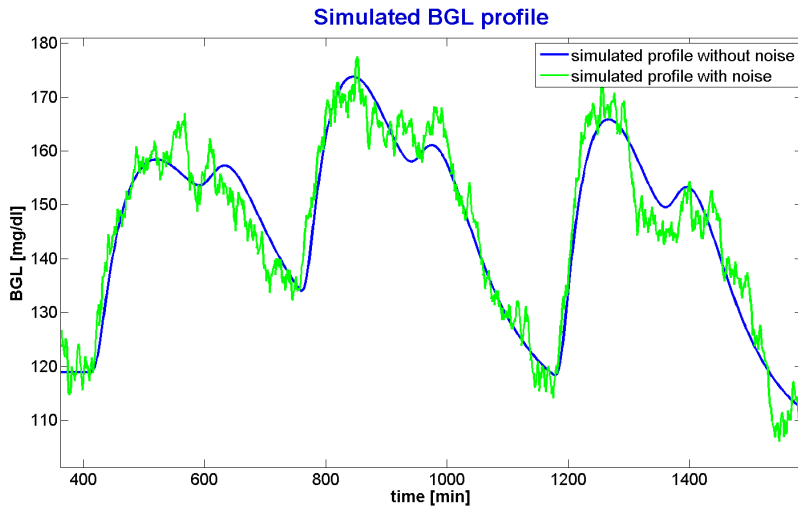


Figura 6.1: Profilo glicemico di un paziente virtuale, prima e dopo l'aggiunta di rumore.

dosi di insulina, normalmente calcolate in modo da controbilanciare perfettamente l'aumento glicemico prodotto dai carboidrati assunti. In questo modo si è ottenuto un numero accettabile di attraversamenti delle soglie ipo e iperglicemiche.

Vale la pena sottolineare che le serie temporali così ottenute presentano comunque un comportamento diverso da quelle reali, in quanto sono molto più regolari e gli aumenti glicemici sono molto rapidi e ripidi e rispecchiano fedelmente l'andamento della rate of appearance dei carboidrati ingeriti. Mancano inoltre tutte quelle variazioni glicemiche, difficilmente spiegabili e spesso imprevedibili, determinate dall'interazione di molteplici fattori, che si verificano in un soggetto reale.

6.2 Risultati: predizione del profilo glicemico

In Figura 6.2 sono rappresentati due profili glicemici simulati e i relativi segnali predetti dalle due reti proposte e dal modello lineare. Per comodità i 4 predittori verranno in seguito indicati con i seguenti nomi:

NN dBGL è la nuova rete neurale che predice la differenza tra il livello glicemico attuale e il livello glicemico che si registrerà tra PH minuti (approccio 1 della Sezione 5.1, pag. 89);

NN deLIN è la nuova rete neurale che predice l'errore commesso dal modello lineare (approccio 2 della Sezione 5.2, pag. 108);

NN0 è la rete neurale proposta da Pérez-Gandía *et al.* in [34];

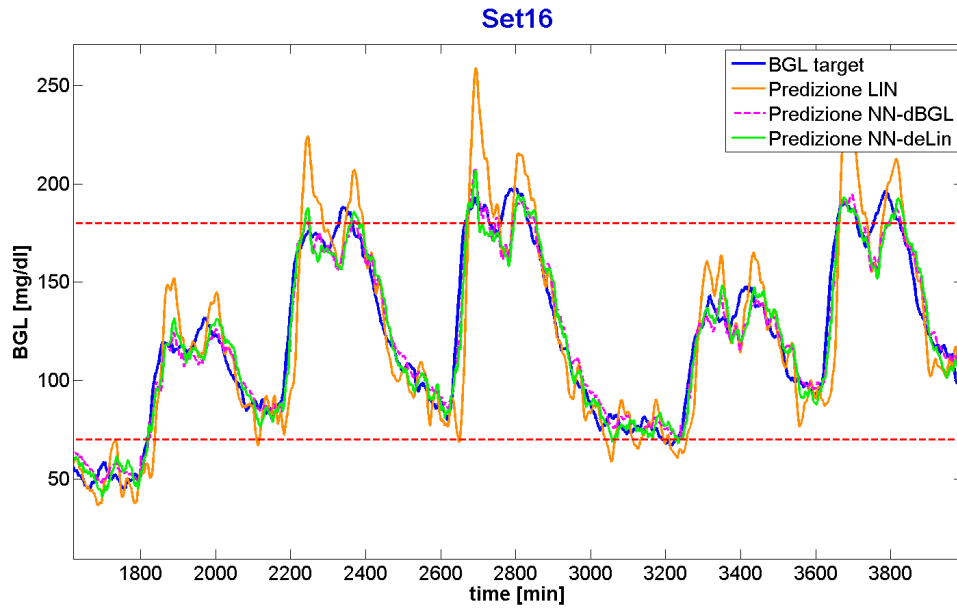
LIN è il modello lineare tempo-variante descritto da Sparacino *et al.* in [39].

In entrambi i casi la predizione ottenuta con le due reti neurali è migliore di quella fornita dal modello lineare. In particolare nei tratti in cui la glicemia aumenta le reti neurali colgono tempestivamente il cambiamento di trend del segnale, mentre il modello lineare accumula un notevole ritardo iniziale. Per quanto riguarda gli attraversamenti della soglia iperglicemica notiamo che i tre predittori hanno performance molto simili, anche se il modello lineare tende a predire più falsi attraversamenti rispetto alle reti neurali, come si può vedere nel pannello inferiore. Nei tratti in cui la glicemia diminuisce i tre modelli si comportano in modo analogo e in corrispondenza dell'attraversamento della soglia ipoglicemica il modello lineare ha un guadagno temporale superiore a quello delle reti neurali. Questa difficoltà incontrata dalle reti neurali nel predire le crisi ipoglicemiche può essere attribuita sia alla scarsità di eventi ipoglicemici presenti nel training set, sia agli ingressi delle reti neurali stesse. Entrambe le reti neurali hanno in ingresso informazioni relative alla rate of appearance dei carboidrati ingeriti e alle sue derivate. Ciò permette loro di predire in maniera molto accurata il segnale glicemico in corrispondenza dei trend in salita, generalmente dovuti proprio all'assunzione di cibo. Per quanto riguarda i trend in discesa, invece, le reti neurali devono effettuare la predizione affidandosi esclusivamente ai valori glicemici passati e alle loro derivate. Non c'è un segnale in ingresso che presenti un'alta correlazione con la diminuzione della glicemia, come invece accade quando essa aumenta. Risulta quindi più difficile ottenere una predizione accurata e un alto guadagno temporale in corrispondenza degli eventi ipoglicemici.

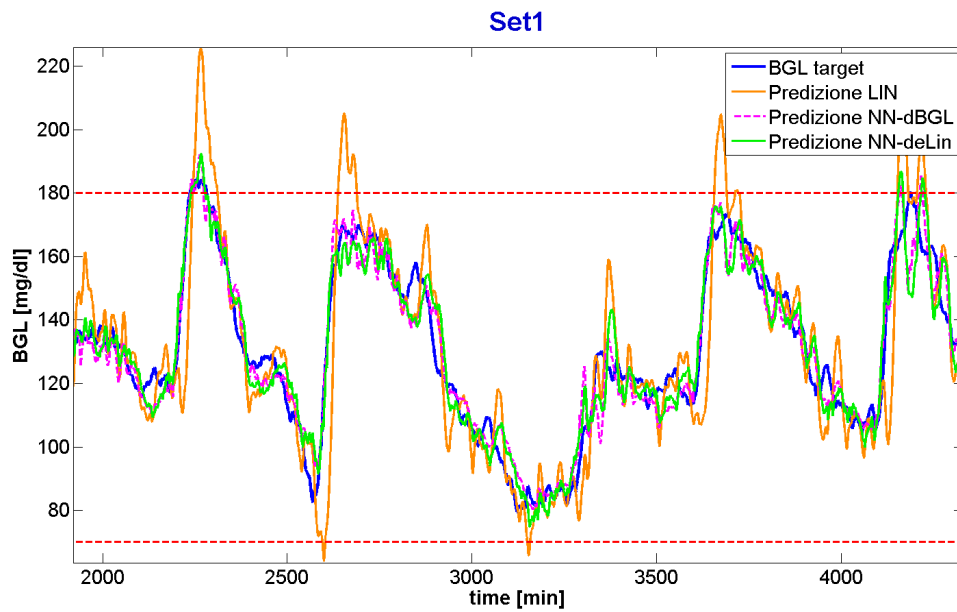
Nelle Figure 6.3, 6.4 , 6.5 sono riassunte, attraverso dei boxplots e delle tabelle, le performance dei 4 predittori sull'intero test set (composto da 20 serie temporali della durata di 3 giorni ciascuna).

J: (Figura 6.3). La predizione ottenuta sia con le reti neurali sviluppate, sia con il modello lineare, è molto regolare, come dimostrano i bassi valori dell'indice J. La rete proposta da Pérez-Gandía *et al.* genera, invece, una predizione irregolare con oscillazioni ad alta frequenza non accettabili dal punto di vista fisiologico. Bisogna sottolineare che il confronto in termini di J tra modello lineare e reti neurali favorisce il modello lineare. Il forgetting factor μ di tale modello viene infatti ottimizzato proprio minimizzando l'indice J, di conseguenza il predittore lineare scelto è il modello che, tra tutti quelli possibili, dà i migliori risultati in termini di J.

RMSE: (Figura 6.4). In termini di root mean square error le performance delle 3 reti neurali sono molto simili e nettamente superiori a quelle del modello lineare. Le reti neurali risultano però avvantaggiate, in quanto il RMSE è proprio la grandezza che viene minimizzata durante



(a) Soggetto simulato 16.



(b) Soggetto simulato 1.

Figura 6.2: Predizione di serie glicemiche simulate. In blu è rappresentato il profilo misurato dal sensore, in arancio la predizione ottenuta col modello lineare, in magenta la predizione fornita da NN dBGL e in verde la predizione ottenuta con NN deLin. I profili predetti con le due reti neurali risultano molto simili.

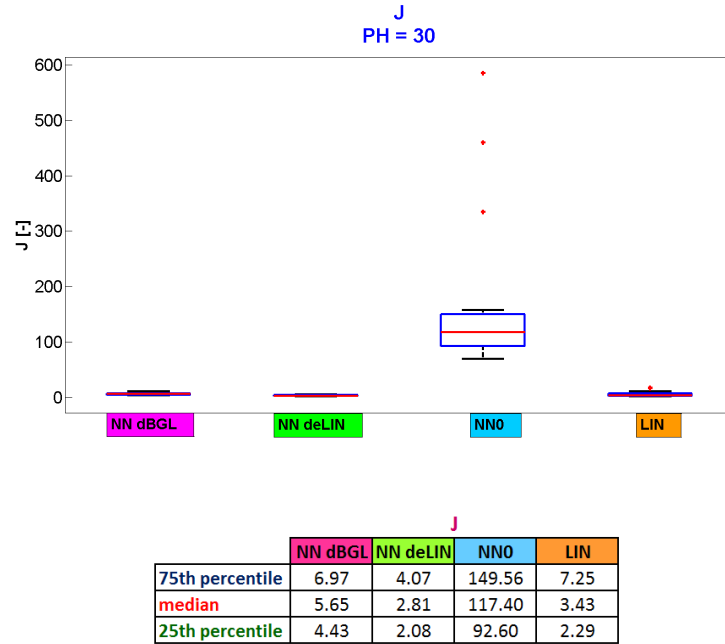


Figura 6.3: Boxplot e tabella riassuntiva con i valori dell'indice J , per i 4 algoritmi di predizione considerati.

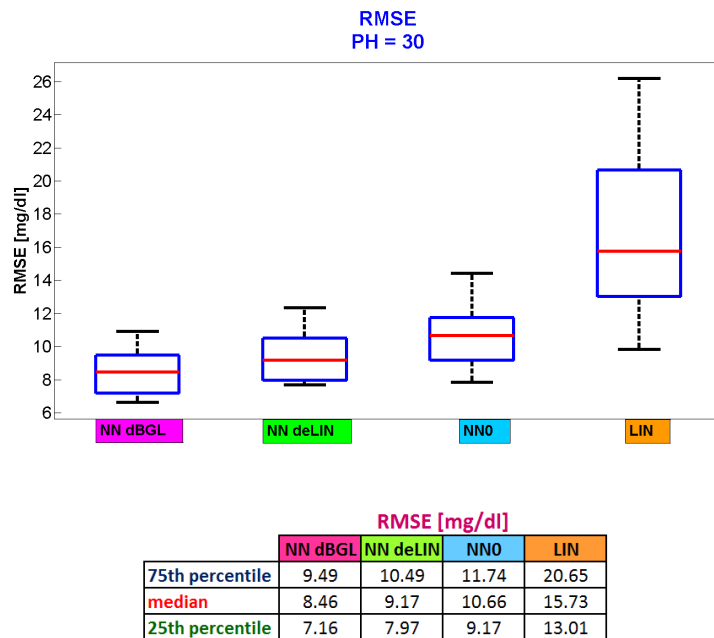


Figura 6.4: Boxplot e tabella riassuntiva con i valori del RMSE, per i 4 algoritmi di predizione considerati.

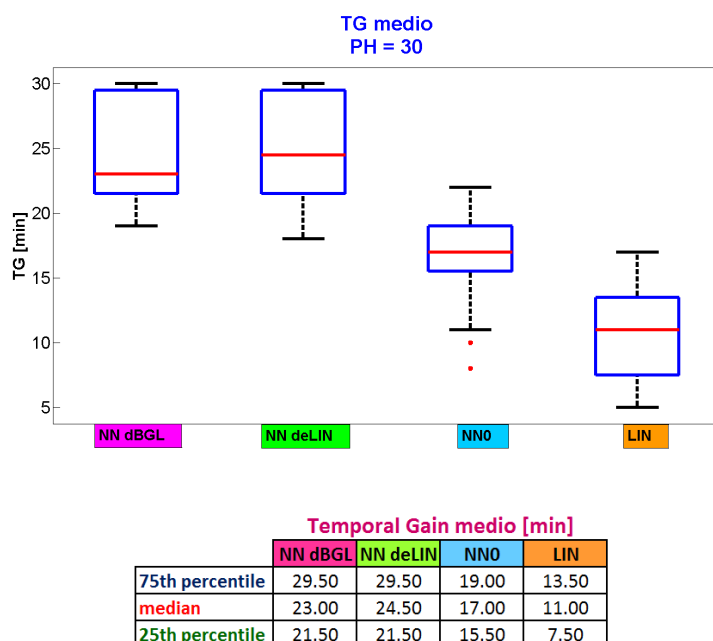


Figura 6.5: Boxplot e tabella riassuntiva con i valori del TG medio, per i 4 algoritmi di predizione considerati.

il loro training. Di conseguenza, analogamente a quanto detto per il modello lineare parlando di J, la rete neurale scelta è quella che, tra tutte le configurazioni possibili, dà i migliori risultati per quanto riguarda il RMSE.

TG medio: (Figura 6.5). Il TG medio, ovvero il guadagno temporale medio del predittore, indica una netta superiorità delle due nuove reti neurali rispetto al modello lineare e rispetto a NNO.

6.3 Risultati: attraversamento di soglie critiche

Come già notato in precedenza, il predittore che fornisce risultati migliori nel complesso può avere performance peggiori rispetto ad altri predittori, se si focalizza l'attenzione esclusivamente sull'attraversamento di soglie critiche. Nonostante visivamente le performance delle 2 nuove reti neurali e del modello lineare risultino molto simili in corrispondenza dei trend decrescenti, in verità il modello lineare predice in modo più tempestivo gli attraversamenti delle soglie ipoglicemiche, come dimostra la Figura 6.6. Per quanto riguarda le iperglicemie, invece, osservando la Figura 6.7 si può notare che le 2 reti neurali permettono di ottenere un guadagno temporale leggermente maggiore di quello ottenuto col modello lineare. La differenza è però minima rispetto a quella che ci si potrebbe aspettare dai plot delle

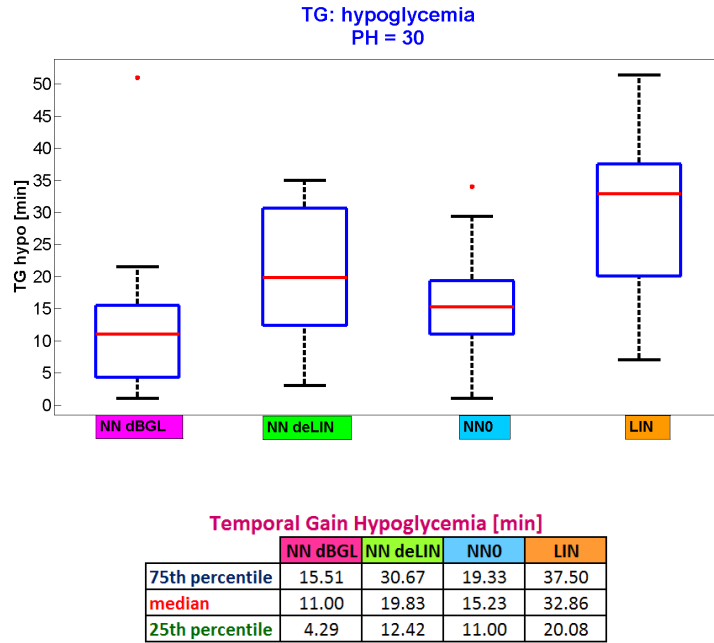


Figura 6.6: Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le ipoglicemie, con i 4 algoritmi di predizione considerati.

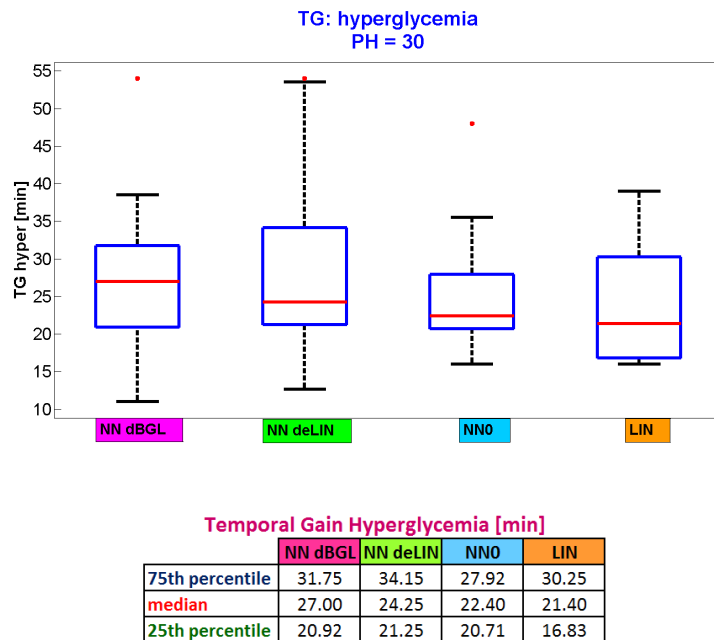


Figura 6.7: Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le iperglicemie, con i 4 algoritmi di predizione considerati.

serie temporali predette a confronto. Benchè il modello lineare colga con un notevole ritardo, rispetto alle reti neurali, l'inizio di un trend glicemico crescente, riesce successivamente a recuperare questo ritardo e a predire con un anticipo adeguato l'attraversamento della soglia iperglicemica. L'analisi del guadagno temporale nell'attraversamento delle soglie di ipo e iperglicemia non può però prescindere dall'analisi della *Tabella di verità* (o *Confusion Matrix*), in cui vengono riassunte la percentuale di eventi correttamente predetti e la percentuale di eventi predetti in modo errato.

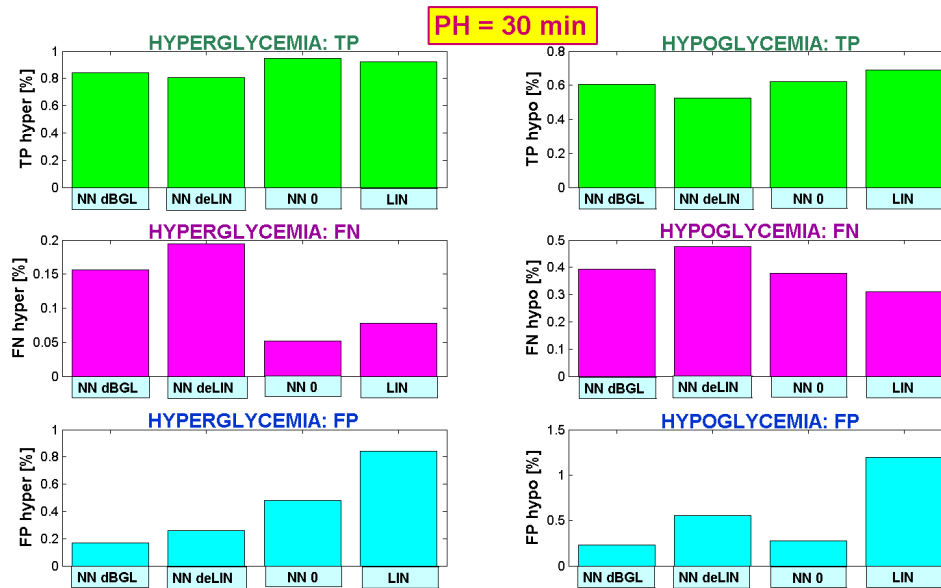


Figura 6.8: Confusion Matrix relativa all'attraversamento delle soglie ipoglicemiche (colonna di destra) e iperglicemiche (colonna di sinistra).

In Figura 6.8 sono rappresentati sia per l'ipoglicemia (colonna di destra) che per l'iperglicemia (colonna di sinistra) le percentuali di:

- *Veri Positivi* (True Positive, TP), ovvero gli attraversamenti di soglia correttamente predetti in un intervallo di tempo pari a $[t-PH, t+PH]$, dove t indica l'istante in cui effettivamente il segnale glicemico attraversa la soglia critica;
- *Falsi Negativi* (False Negative, FN), ovvero gli attraversamenti di soglia che il predittore non predice nell'intervallo $[t-PH, t+PH]$. Questo evento si verifica sia quando il predittore erroneamente non predice l'attraversamento della soglia critica, sia quando lo predice con un ritardo o un anticipo superiore a PH minuti. Se la predizione è ritardata più di PH minuti risulta totalmente inutile (in quanto viene predetto un evento che in realtà è già avvenuto). Un anticipo superiore a $2 \cdot PH$

minuti, d'altro canto, è poco credibile e solitamente è da imputare a errori di predizione.

- *Falsi Positivi* (False Positive, FP), ovvero tutti gli attraversamenti di soglia che il modello predice, ma che non esistono nella serie temporale reale. Un eccessivo numero di falsi positivi rende poco credibile la predizione degli attraversamenti di soglia, poichè spesso essi in realtà non si verificheranno.

Notiamo che il modello lineare e la NN0 predicono correttamente un numero leggermente maggiore di eventi sia ipo che iperglicemici, come dimostrano le percentuali di veri positivi e di falsi negativi. Il modello lineare, però, predice anche un'alta percentuale di falsi attraversamenti. In particolare per quanto riguarda le ipoglicemie, oltre la metà di quelle predette dal modello lineare in verità non si verificano (si può notare nel pannello in basso a destra di Figura 6.8 che la percentuale di falsi positivi predetti dal modello lineare è superiore al 100%).

Da questa analisi è evidente che ognuno dei modelli utilizzati ha dei pregi e dei difetti e non è possibile individuare un predittore che dia risultati nettamente migliori di tutti gli altri con tutti gli indici analizzati.

6.4 Risultati: performance dei predittori al variare del PH

All'aumentare del PH diventa sempre più difficile predire accuratamente il segnale glicemico, a causa del grande numero di fattori che lo potrebbero influenzare nell'intervallo che va dall'istante attuale all'istante in cui si desidera effettuare la predizione. Per completare l'analisi del comportamento delle 2 nuove reti neurali sviluppate, ne sono state testate le performance considerando vari orizzonti di predizione. In particolare, partendo da PH=10 minuti si è arrivati, muovendosi con un passo pari a 5 minuti, fino a PH=60 minuti. Indubbiamente per ogni PH bisognerebbe riottimizzare la struttura e gli ingressi della rete. Ciò avrebbe però richiesto una grande mole di lavoro aggiuntivo, quindi è stata mantenuta invariata la struttura di entrambe le reti neurali e ci si è limitati a modificarne il target e la rate of appearance dei carboidrati ingeriti (in modo tale da fornire in ingresso alla rete il valore assunto dalla Ra in concomitanza all'istante in cui si vuole effettuare la predizione, e in modo che le sue 3 derivate siano equispaziate in $[t, t+PH]$). I risultati ottenuti sono stati confrontati con quelli forniti da NN0 e dal modello lineare.

In Figura 6.9 e in Figura 6.10 sono rappresentati gli andamenti di J e del RMSE in funzione del PH. Mentre il valore di J resta pressochè invariato (a parte per NN0 che produce una predizione molto irregolare), il RMSE, invece, peggiora gradualmente all'aumentare del PH. Per quanto

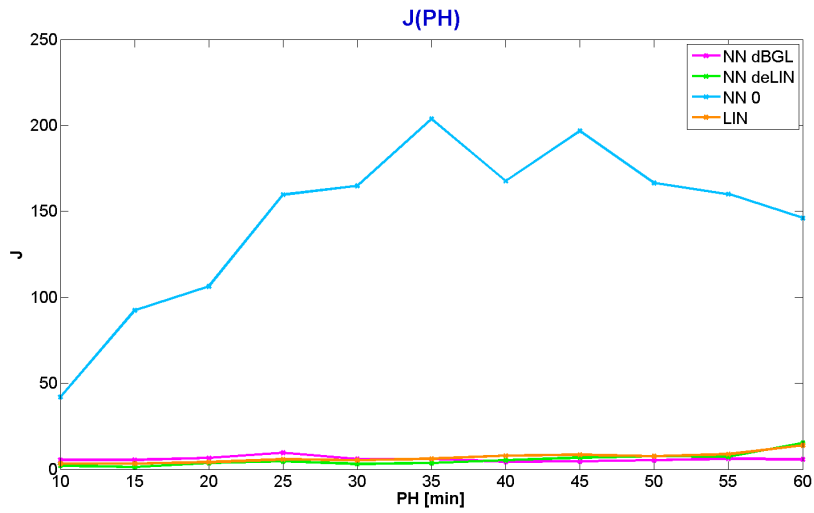


Figura 6.9: Andamento di J al variare del PH .

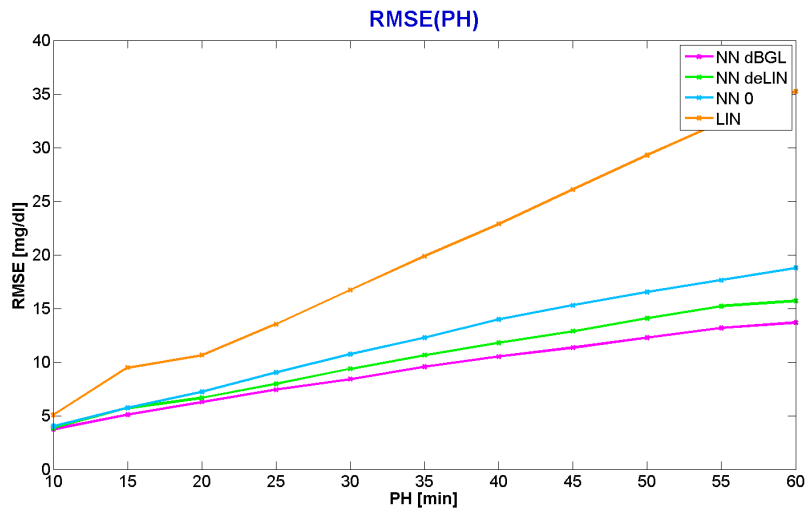
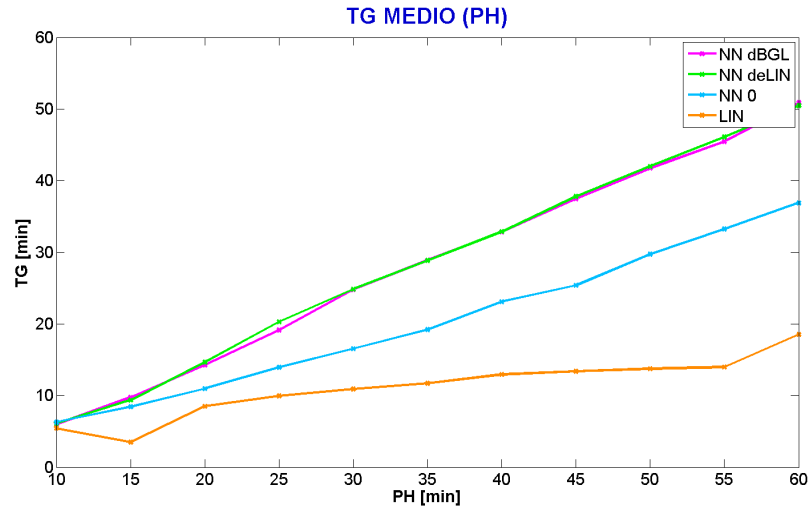
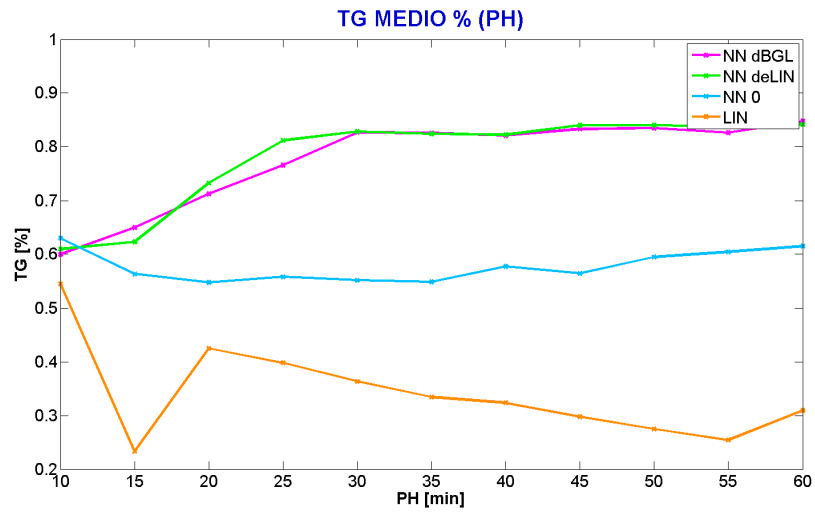


Figura 6.10: Andamento del RMSE al variare del PH .



(a) Guadagno temporale effettivo.



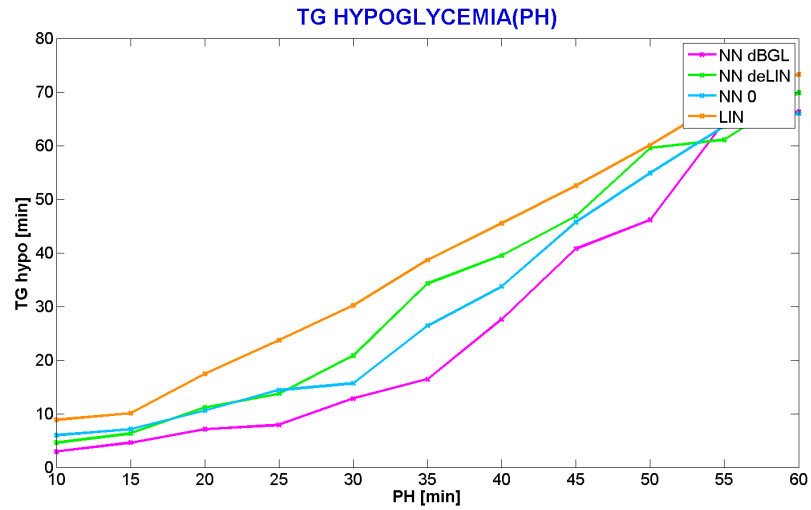
(b) Guadagno temporale percentuale.

Figura 6.11: Guadagno temporale medio al variare del PH.

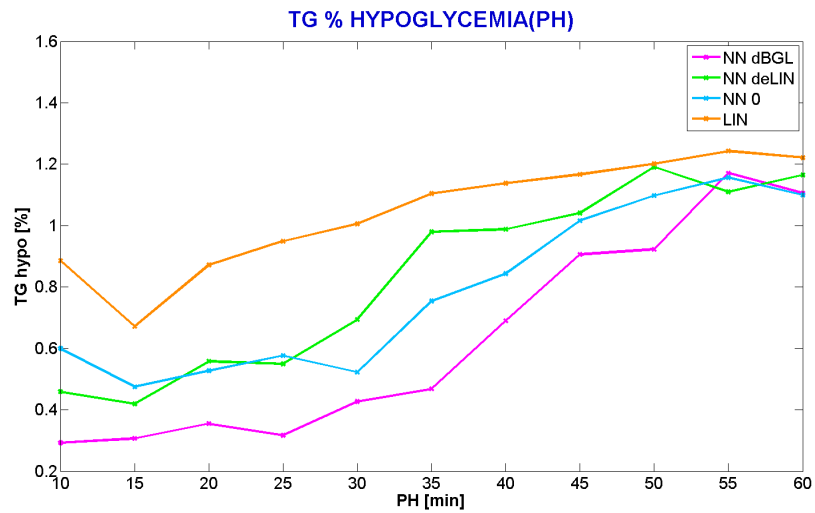
riguarda il guadagno temporale medio, come già notato con $PH=30$ minuti, le performance delle due nuove reti neurali sono nettamente superiori a quelle del modello lineare e migliori di quelle di NN0. Dal grafico del guadagno temporale percentuale (ovvero il guadagno temporale normalizzato all'orizzonte di predizione considerato), riportato in Figura 6.11(b), si nota che questo indice non è particolarmente influenzato dall'orizzonte di predizione. Infatti, per ogni predittore, il valore percentuale resta pressochè invariato per i vari PHs.

Per quanto riguarda la predizione delle crisi ipoglicemiche il guadagno temporale ottenuto col modello lineare è chiaramente superiore a quello registrato con le reti neurali (Figura 6.12). Se, però, analizziamo la colonna sinistra di Figura 6.13 notiamo che il numero di falsi attraversamenti predetti dal modello lineare è inaccettabile. Dal riquadro in basso a destra, infatti, si vede che i falsi negativi raggiungono un valore che sfiora il 300% in corrispondenza di $PH=60$ minuti, ovvero ogni 4 crisi ipoglicemiche predette dal modello lineare 3, in verità, non si verificheranno. Di conseguenza, anche se la predizione è tempestiva, non è assolutamente affidabile. Le reti neurali, invece, hanno un guadagno temporale minore, ma predicono un numero di falsi attraversamenti ipoglicemici nettamente inferiore. Osservando l'andamento dei veri positivi si nota che all'aumentare del PH la rete che predice direttamente il livello glicemico (NN dBGL) riesce ad individuare un numero via via inferiore di ipoglicemie. Il degrado delle performance è particolarmente marcato per PH superiori ai 30 minuti, con i quali vengono predette meno della metà delle crisi ipoglicemiche che si verificheranno. La rete che corregge l'errore del modello lineare, invece, riesce ad individuare un numero alto di ipoglicemie, mantenendo accettabilmente basso il numero di falsi attraversamenti predetti. Osservando il guadagno temporale relativo alla predizione delle ipoglicemie si può avere l'impressione che, per orizzonti di predizione più elevati, i vari modelli si comportino in maniera migliore, dato che il TG diventa quasi pari al PH. Attraverso un'analisi più accurata dei profili predetti si nota che le ipoglicemie sono effettivamente individuate in modo tempestivo, ma ne vengono predette poche rispetto a quelle che in verità si verificano. Ovvero, mentre con un orizzonte temporale più basso eventi difficili da individuare vengono predetti con un ritardo elevato, ma inferiore al PH, e peggiorano il TG finale (che è la media dei TG relativi ad ogni ipoglicemia correttamente predetta), al crescere del PH questi eventi non vengono più individuati, oppure vengono predetti con un ritardo superiore al PH e non se ne tiene conto nel calcolo del TG.

Il comportamento dei vari modelli nella predizione delle iperglicemie è differente da quello osservato per le ipoglicemie. Per orizzonti di predizione minori o uguali a 30 minuti i 4 modelli hanno performance simili. All'aumentare del PH, però, le 2 nuove reti neurali proposte mantengono performance costanti, mentre si nota un netto peggioramento del comportamento di NN0 e del modello lineare. Per questi ultimi 2 predittori si osservano



(a) Guadagno temporale effettivo.



(b) Guadagno temporale percentuale.

Figura 6.12: Guadagno temporale relativo alla predizione delle ipoglicemie, al variare del PH.

6.4 Risultati: performance dei predittori al variare del PH

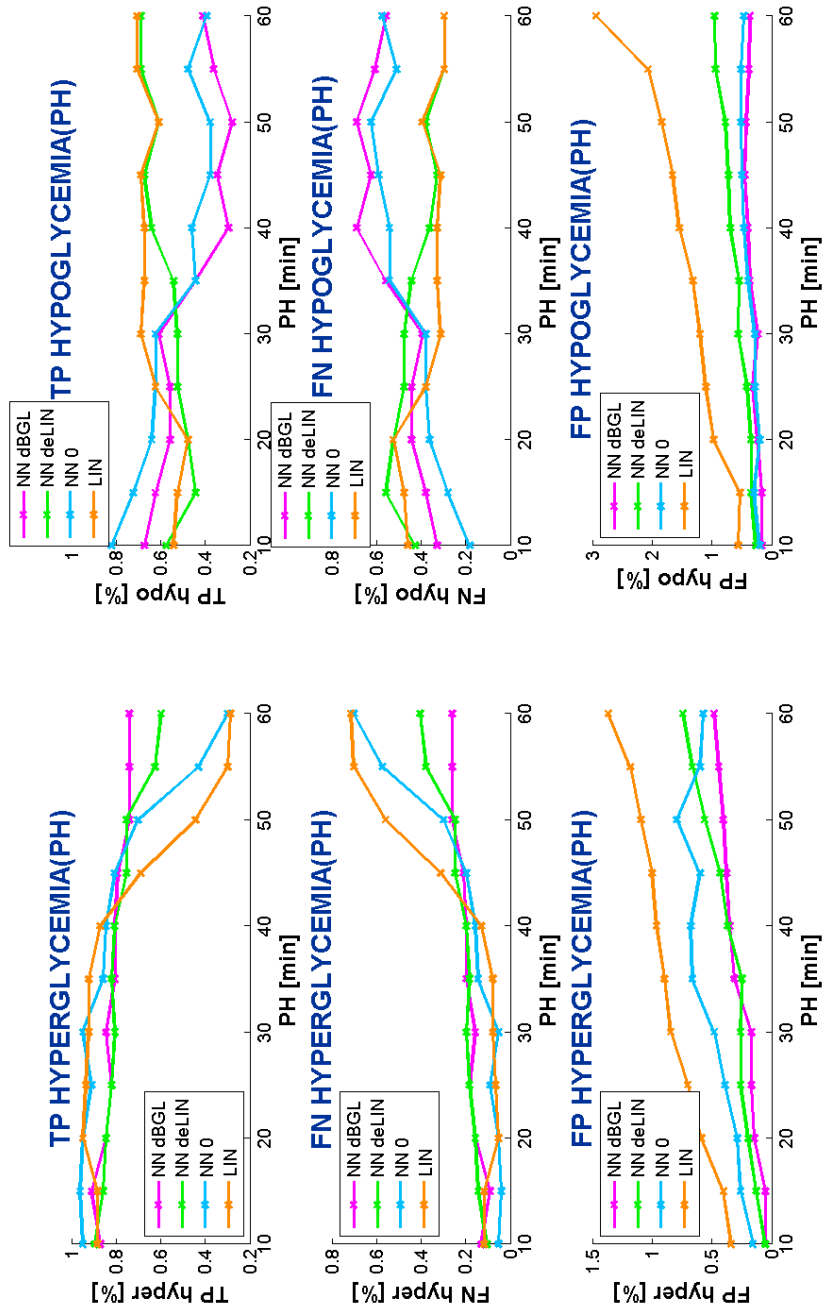
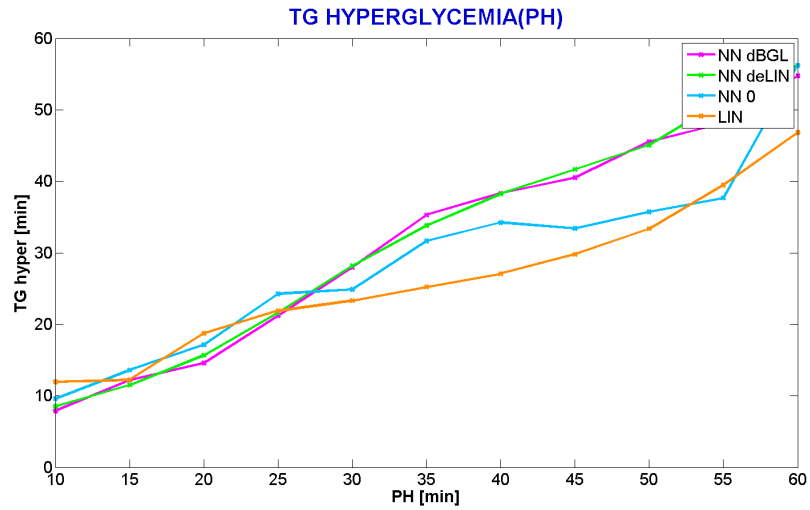
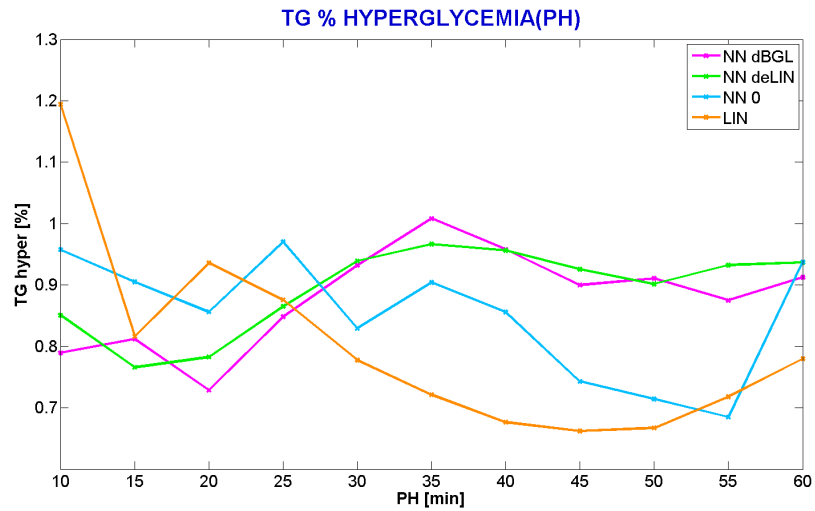


Figura 6.13: Confusion Matrix al variare del PH



(a) Guadagno temporale effettivo.



(b) Guadagno temporale percentuale.

Figura 6.14: Guadagno temporale relativo alla predizione delle iperglicemie, al variare del PH.

sia un degrado del guadagno temporale, sia un crollo del numero di eventi iperglicemici correttamente individuati, come dimostra il grafico che descrive l'andamento dei veri positivi in funzione del PH (Figura 6.13, riquadro in alto a sinistra). Per PH elevati il modello lineare oltre a individuare poche crisi iperglicemiche, predice un alto numero di falsi attraversamenti della soglia iperglicemica, come dimostra la percentuale di falsi positivi superiore al 100%. Da un'analisi complessiva dei vari indici la rete neurale che corregge l'errore del predittore lineare, allenata su dati simulati, risulta superiore agli altri modelli nell'individuare l'attraversamento di soglie critiche. Ciò è particolarmente apprezzabile in Figura 6.13. Per quanto riguarda la predizione delle ipoglicemie la rete mantiene un numero di veri positivi simile a quello del modello lineare e nettamente superiore a quello di NN dBGL e a quello di NN0, mentre i falsi positivi risultano di molto inferiori a quelli che caratterizzano il modello lineare e solo lievemente superiori a quelli delle altre 2 reti. Nel predire le iperglicemie la percentuale di veri positivi è confrontabile con quella di NN dBGL e, per PH elevati, chiaramente superiore a quella di NN0 e del modello lineare. La percentuale di falsi positivi è analoga a quella registrata con le altre 2 reti neurali e marcatamente inferiore a quella del modello lineare.

Questa rete neurale sembra, quindi, unire i pregi delle reti neurali a quelli del modello lineare

6.5 Conclusioni

I risultati ottenuti dalle 2 nuove reti neurali proposte, testate con dati simulati, sono molto incorruggianti, in particolare per quanto riguarda la predizione generale del profilo glicemico e per quanto riguarda la capacità di individuare tempestivamente crisi iperglicemiche future. Le due reti neurali risultano però avvantaggiate nel predire le iperglicemie, rispetto a NN0 e al modello lineare. Tra i loro ingressi, infatti, compare la rate of appearance futura dei carboidrati ingeriti. Il generatore utilizzato per creare l'andamento della rate of appearance è lo stesso generatore incorporato all'interno della funzione che simula le curve glicemiche dei soggetti diabetici virtuali, in base ad alcuni loro parametri fisiologici, ai pasti e alla quantità di insulina iniettata. Di conseguenza la curva glicemica e la curva che descrive la rate of appearance dei carboidrati hanno un andamento molto simile e seguono la stessa dinamica temporale, come si può notare in Figura 6.15. In un paziente reale questo fenomeno non è così marcato, poichè moltissimi fattori, oltre alla quantità di carboidrati ingeriti, concorrono alla determinazione della glicemia.

Inoltre i profili glicemici dei pazienti virtuali sono generati utilizzando funzioni matematiche. Le reti neurali feedforward sono particolarmente adatte ad individuare la funzione che associa i vettori in ingresso a quelli

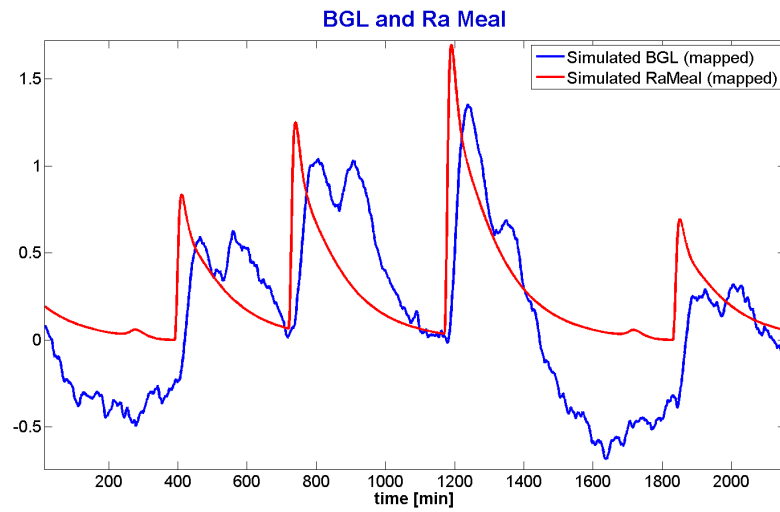


Figura 6.15: *Profilo glicemico e rate of appearance dei carboidrati in un paziente virtuale. Le curve sono state normalizzate per essere confrontabili.*

in uscita che vengono loro presentati in fase di training, anche quando tale funzione è complessa e difficilmente descrivibile analiticamente. In un soggetto reale, però, l'andamento glicemico non può essere perfettamente modellato attraverso funzioni matematiche. Questo renderà più difficile e meno precisa la predizione del livello di glucosio ematico.

Nel prossimo capitolo saranno presentati e commentati i risultati ottenuti con un dataset reale.

Capitolo 7

Implementazione e risultati su dati reali

Consapevoli dei limiti intrinseci ai dati simulati, e delle condizioni favorevoli per le reti neurali allenate e testate su di essi, è stato indagato il comportamento delle 2 nuove reti anche su un dataset reale e i risultati ottenuti sono stati confrontati con quelli forniti dal modello lineare tempo variante di Sparacino *et al.* [39] e con quelli ottenuti con la rete neurale di Pérez-Gandía *et al.* [34]. Inoltre, per poter affermare l'esistenza di differenze significative tra le performance dei vari modelli, i risultati relativi ad ogni indice sono stati confrontati attraverso il *Wilcoxon rank-sum test*, un test statistico non parametrico che può essere utilizzato in alternativa al *t*-test quando i dati in esame non sono distribuiti in modo gaussiano. Questa procedura testa l'ipotesi nulla, che i dati contenuti in 2 vettori \mathbf{x} e \mathbf{y} siano campioni indipendenti di una stessa distribuzione continua, con mediana coincidente, contro l'ipotesi che la loro mediana sia differente. Il livello di significatività adottato in tutti i test è pari al 5%¹.

7.1 Creazione del dataset

Il dataset utilizzato per allenare e testare i vari predittori è composto dai profili glicemici di 29 soggetti diabetici di tipo 1, monitorati con sensore CGM FreeStyle NavigatorTM. Per generare il training e il validation set sono state selezionate, rispettivamente, 28 e 12 serie temporali, della durata di 3 giorni ciascuna, in modo da riprodurre le condizioni in cui erano state allenate le reti neurali coi dati simulati. Per il test set sono stati selezionati 35 spezzoni, di durata superiore alle 24 ore.

¹Ciò implica che l'ipotesi nulla viene confermata se la probabilità di osservare 2 vettori appartenenti alla stessa distribuzione e che differiscono in mediana almeno tanto quanto quelli in esame è superiore a 0.05.

Per ognuno dei 29 soggetti, oltre alle misure glicemiche, il dataset contiene anche l'orario dei pasti e le quantità di carboidrati ingeriti. È stato quindi possibile generare la rate of appearance dei carboidrati assunti con i vari pasti e utilizzarla, assieme alle sue derivate, come ingresso delle reti neurali.

A differenza di quanto proposto da Pérez-Gandía *et al.* non è stato eseguito nessun prefiltraggio dei dati. Si è preferito utilizzare le serie temporali "grezze", benchè in alcuni casi fossero molto rumorose, poichè una rete allenata su dati "puliti" presenta un comportamento inaccettabile in fase di simulazione, se riceve in ingresso dati minimamente corrotti da rumore.

7.2 Risultati: predizione del profilo glicemico

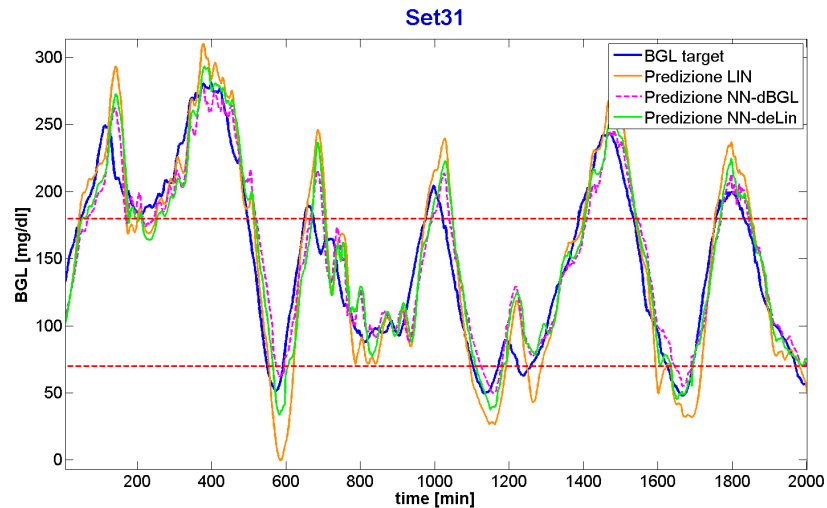


Figura 7.1: Plot del profilo glicemico del paziente 31, e delle relative predizioni.

In Figura 7.1 e in Figura 7.2 possiamo osservare 2 serie temporali e le predizioni che si riferiscono ad esse. In entrambi i casi notiamo che l'uscita delle due reti neurali è più aderente al profilo reale, se confrontata con l'uscita del modello lineare, soprattutto quando si verificano inversioni del trend glicemico. Le reti neurali colgono in anticipo, rispetto al modello lineare, l'inizio dell'aumento del livello di glucosio ematico grazie alla presenza, tra i loro ingressi, della rate of appearance dei carboidrati. Successivamente però, come già osservato con i dati simulati, il modello lineare recupera il ritardo iniziale e predice l'attraversamento della soglia iperglicemica generalmente in anticipo rispetto alle reti neurali. Per quanto riguarda i tratti in cui la glicemia diminuisce, le performance dei tre predittori sono molto simili e, nuovamente, il modello lineare riesce a predire più tempestivamente l'attraversamento della soglia ipoglicemica.

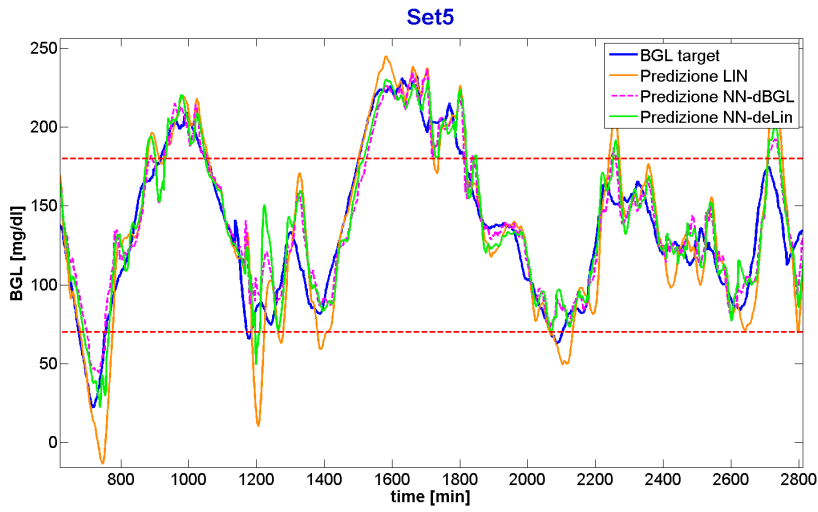


Figura 7.2: Plot del profilo glicemico del paziente 5, e delle relative predizioni.

In Figura 7.3, in Figura 7.4 e in Figura 7.5 sono riportate le statistiche che riassumono i valori di J , del RMSE e del guadagno temporale ottenuti con i vari predittori.

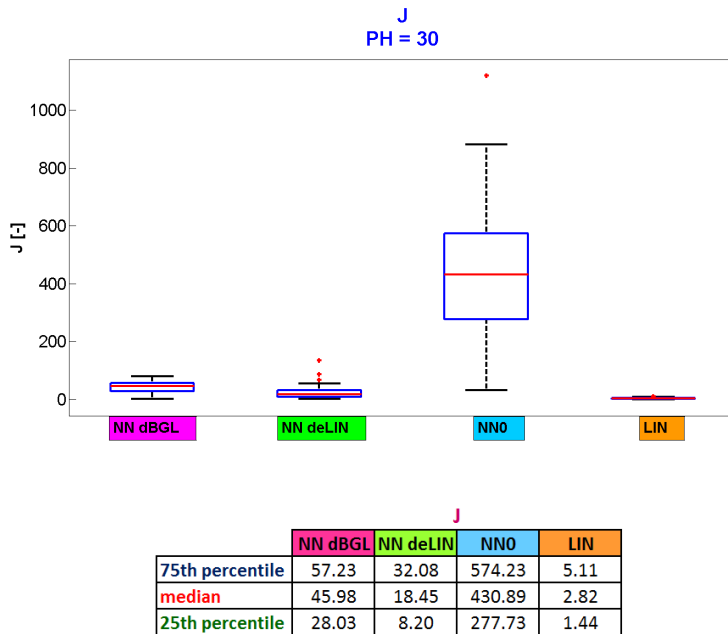


Figura 7.3: Boxplot e tabella riassuntiva con i valori dell'indice J , per i 4 algoritmi di predizione considerati.

J: (Figura 7.3). Quest'indice mostra chiaramente come le 2 nuove reti

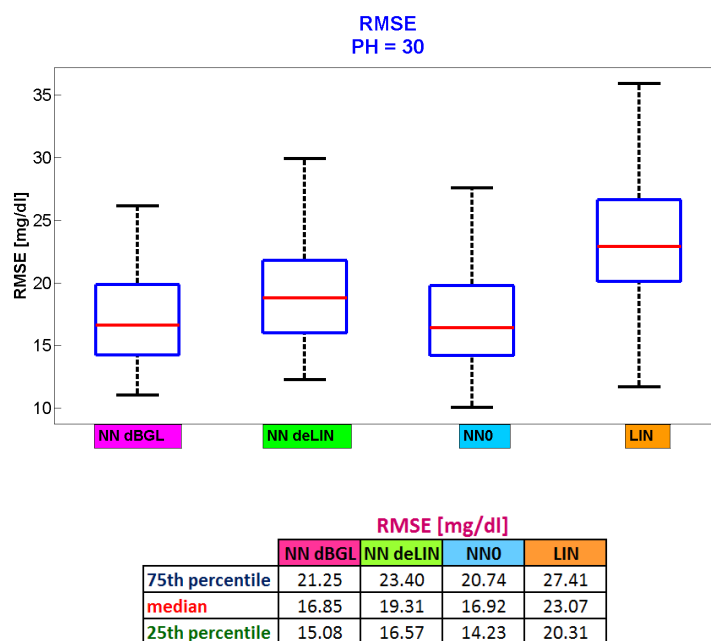


Figura 7.4: Boxplot e tabella riassuntiva con i valori del RMSE, per i 4 algoritmi di predizione considerati.

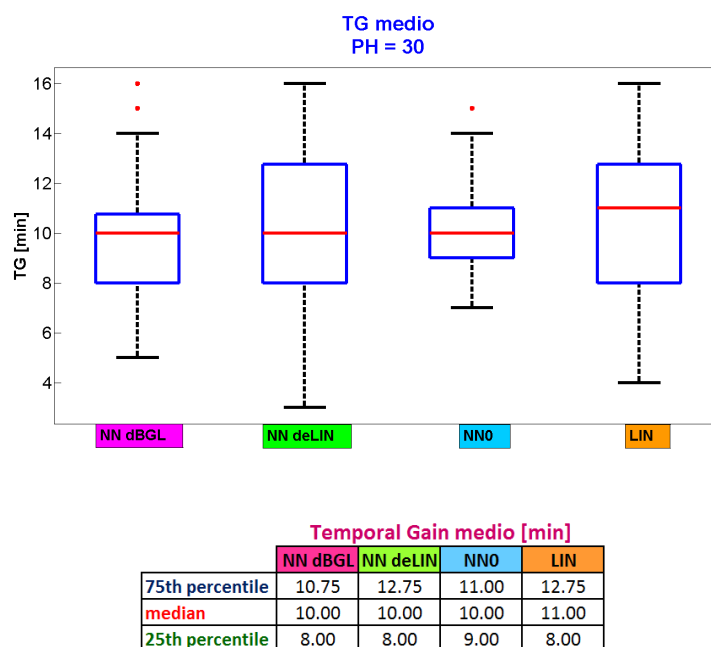


Figura 7.5: Boxplot e tabella riassuntiva con i valori del TG medio, per i 4 algoritmi di predizione considerati.

neurali proposte forniscano una predizione significativamente più regolare rispetto a NN0. Come accadeva con i dati simulati, anche con i dati reali il modello lineare è il migliore in termini di J. Vale la pena sottolineare nuovamente che il forgetting factor μ del modello lineare è ottimizzato proprio minimizzando J. In Figura 7.6 e in Figura 7.7 sono rappresentate la predizione ottenuta con NN0 (pannello superiore) e quella ottenuta con gli altri modelli (pannello inferiore), per due profili glicemici distinti. Le performance di NN0 relative al profilo ritratto in Figura 7.6 sono confrontabili con quelle degli altri modelli. Per serie glicemiche più rumorose, come quella ritratta in Figura 7.7 la predizione di NN0 è caratterizzata da oscillazioni ad altissima frequenza e di ampiezza non trascurabile. Gli altri modelli, invece, mantengono un comportamento accettabile. Ciò giustifica gli altissimi valori assunti da J quando calcolato sulle predizioni fornite da NN0.

RMSE: (Figura 7.4). Il RMSE risulta significativamente più basso per le reti neurali che per il modello lineare. La differenza tra le varie reti neurali, invece, non è significativa. Come già sottolineato in precedenza, i pesi delle reti neurali sono settati minimizzando il RMSE, quindi questo confronto è polarizzato a favore delle reti neurali. Inoltre il modello caratterizzato dal RMSE minore non è necessariamente quello che fornisce la predizione migliore. Dai risultati riportati in Figura 7.4 NN0 risulta essere il miglior predittore in termini di RMSE. Tuttavia, come dimostra la Figura 7.7, questa rete neurale produce spesso profili troppo irregolari.

TG medio: (Figura 7.5). Il TG medio risulta molto basso e leggermente maggiore per il modello lineare. Il *Wilcoxon rank-sum test* indica, comunque, che le differenze tra i valori ottenuti non sono significative. Questo risultato, inoltre, rispecchia più i limiti del metodo con cui viene calcolato il guadagno temporale medio che le reali performance dei 4 modelli. Infatti, il TG del modello lineare risulta più elevato solamente perchè il RMSE non si abbassa mai di molto, nemmeno shiftando il segnale predetto. In Tabella 7.1 sono riportate la media e la standard deviation dei valori minimi assunti dal RMSE per i 4 predittori. Il RMSE minimo relativo al modello lineare è considerevolmente più grande di quello relativo a NN dBGL ed è maggiore anche di quello relativo a NN deLIN.

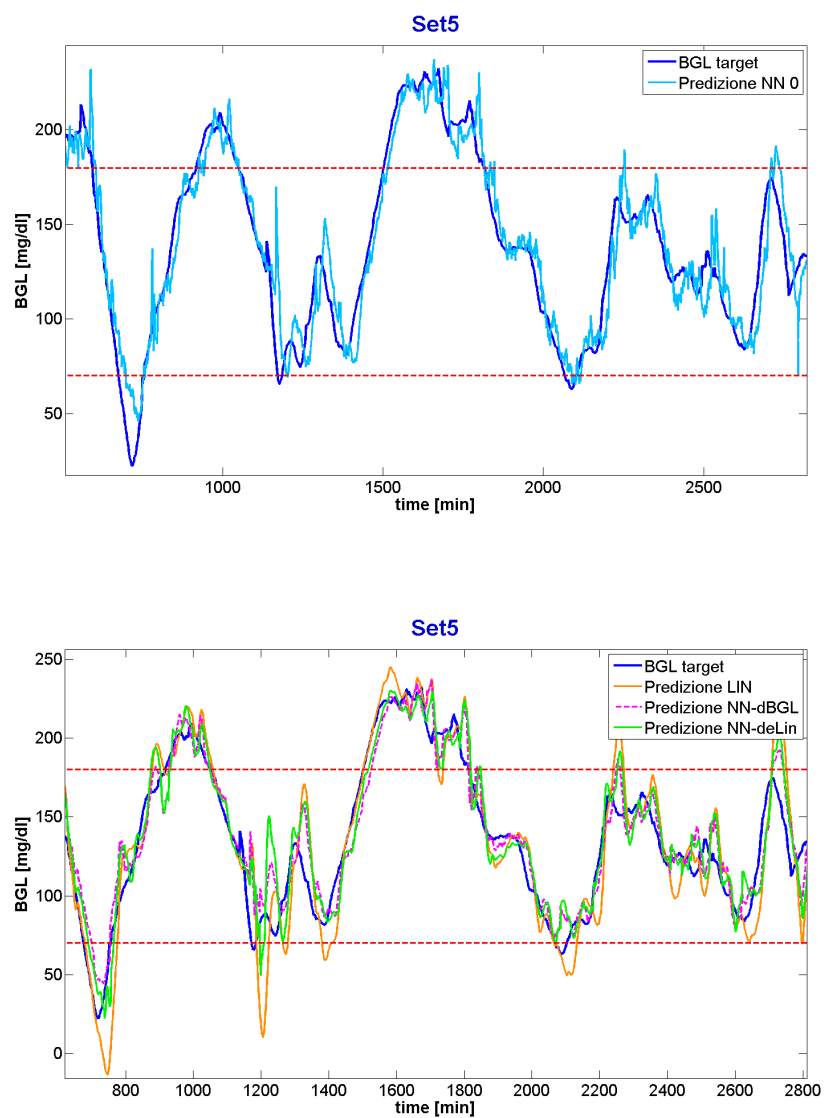


Figura 7.6: Predizione ottenuta su un profilo poco rumoroso.

7.2 Risultati: predizione del profilo glicemico

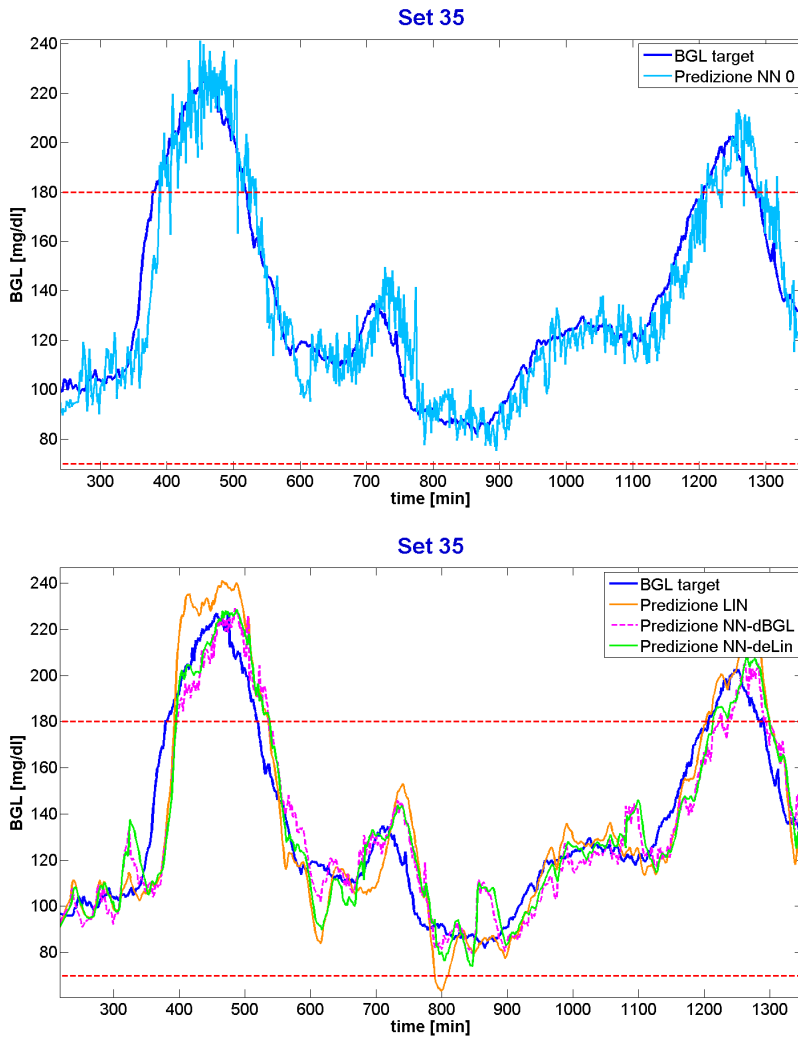


Figura 7.7: Predizione ottenuta su un profilo molto rumoroso.

| MINIMUM RMSE | | | | |
|--------------|---------|----------|------|-------|
| | NN dBGL | NN deLIN | NN 0 | LIN |
| mean | 10.95 | 14.48 | 9.73 | 18.27 |
| sd | 2.00 | 3.20 | 2.73 | 4.84 |

Tabella 7.1: Media e standard deviation del RMSE minimo dei vari predittori. Il modello lineare ha RMSE minimo superiore rispetto alle reti neurali.

7.3 Risultati: attraversamento di soglie critiche

In Figura 7.8 e in Figura 7.10 sono riassunte le statistiche relative al guadagno temporale in corrispondenza all'attraversamento delle soglie ipo e iperglicemiche. Nel predire le ipoglicemie il modello lineare permette di

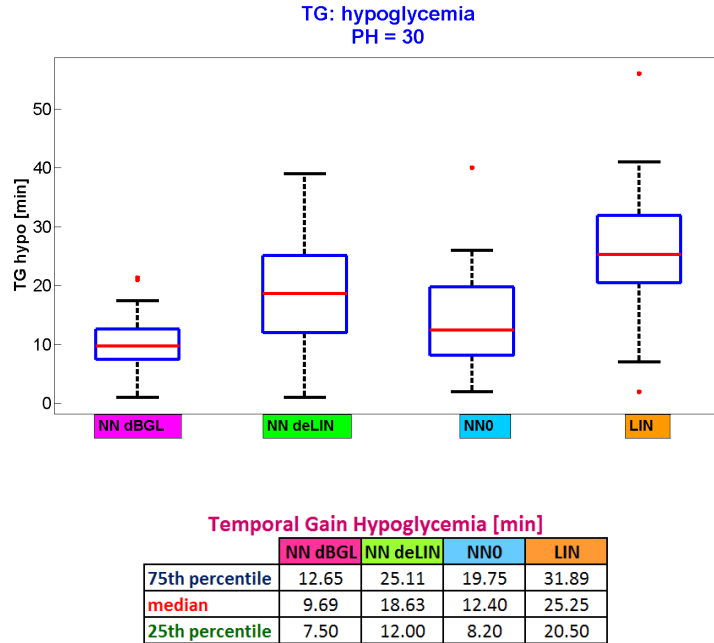


Figura 7.8: Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le ipoglicemie, con i 4 algoritmi di predizione considerati.

ottenere un anticipo superiore in maniera significativa, rispetto a quello fornito dalle reti neurali. In particolare NN dBGL individua le crisi ipoglicemiche mediamente con solo 10 minuti di anticipo rispetto all'istante in cui si verificheranno, mentre il modello lineare permette di ottenere un anticipo medio di 25 minuti. La rete neurale per la correzione dell'errore del predittore lineare ha performance intermedie, con un guadagno temporale medio di circa 18 minuti, che risulta significativamente superiore a quello che caratterizza NN dBGL, ma inferiore in modo significativo a quello del modello lineare.

Questi risultati devono essere considerati tenendo conto delle percentuali di veri e falsi allarmi generati dai vari modelli. Come si può notare nella colonna destra di Figura 7.9 il numero di false ipoglicemie predette dal modello lineare è nettamente superiore al numero di falsi allarmi generati da NN dBGL. Le percentuali di crisi ipoglicemiche correttamente individuate dai due modelli, invece, sono circa uguali, come dimostrano i veri positivi ad esse relativi. Le performance peggiori sono quelle di NN deLIN. Per questa rete neurale, infatti, il numero di crisi ipoglicemiche correttamente predette

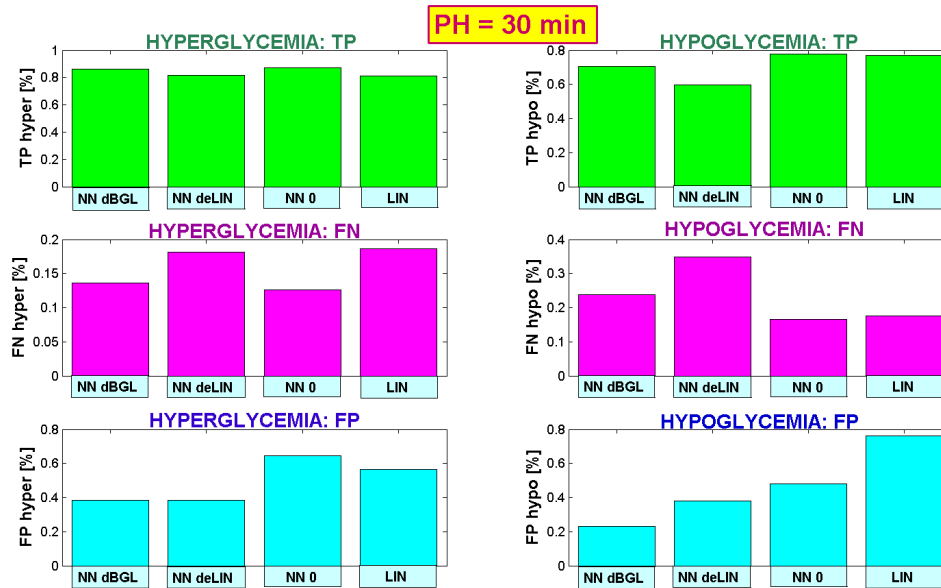


Figura 7.9: Confusion Matrix relativa all'attraversamento delle soglie ipoglicemiche (colonna di destra) e iperglicemiche (colonna di sinistra).

è basso (solamente il 60% di quelle che si verificheranno effettivamente), mentre il numero di crisi erroneamente predette non è trascurabile (circa il 40% delle ipoglicemie predette in verità non si verificheranno).

I risultati ottenuti in corrispondenza dell'attraversamento della soglia iperglicemica sono più incoraggianti. Il predittore con il guadagno temporale più elevato è il modello lineare (in media 24 minuti), mentre NN dBGL dà il TG più basso (in media 17 minuti). La differenza, anche se significativa, risulta comunque molto più limitata di quella osservata per la predizione delle ipoglicemie. La rete NN deLIN presenta anche in questo caso un comportamento intermedio, con un TG medio pari a 20 minuti. Tale guadagno temporale risulta confrontabile con quello fornito da NN0 (che vale mediamente 24 minuti). Le percentuali di iperglicemie correttamente individuate sono molto simili per tutti e 4 i modelli, mentre la percentuale di false iperglicemie predette è più elevata per il modello lineare e per NN0, rispetto alle 2 nuove reti neurali sviluppate.

Da quest'analisi generale NN deLIN risulta caratterizzata da un guadagno temporale superiore rispetto a quello di NN dBGL, in particolare nella predizione delle ipoglicemie, in corrispondenza alle quali la differenza risulta significativa. Per quanto riguarda le iperglicemie, invece, la superiorità del TG ottenuto con NN deLIN, rispetto al TG ottenuto con NN dBGL, non è significativa.

In verità le due reti neurali generano profili molto simili. In corrisponden-

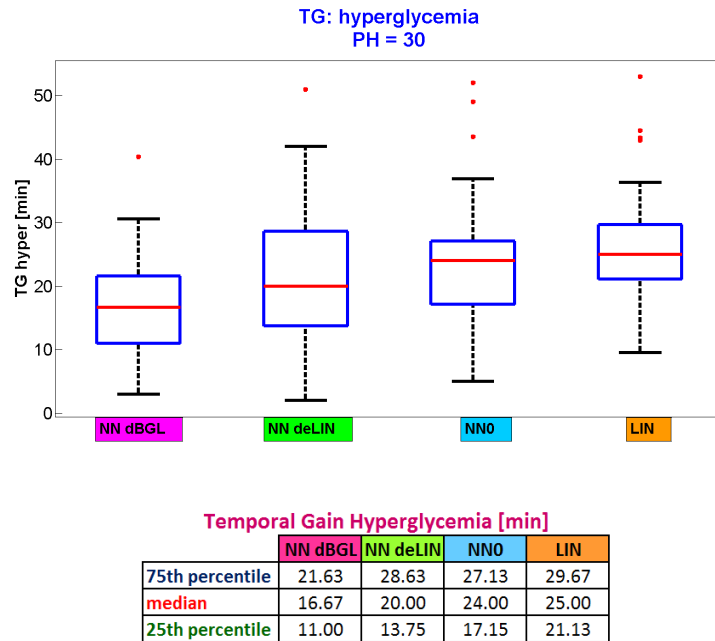


Figura 7.10: Boxplot e tabella riassuntiva con i valori del TG ottenuto predicendo le iperglicemie, con i 4 algoritmi di predizione considerati.

za di attraversamenti difficili da predire, però, NN dBGL individua l'evento con un guadagno temporale minimo (pari a 1 o 2 minuti), NN deLIN, invece, non individua l'evento o lo individua con un ritardo superiore al PH (come dimostra la più alta percentuale di falsi negativi). Di conseguenza, il bassissimo TG di questi eventi va ad abbassare il TG medio di NN dBGL, mentre essi non vengono considerati nel calcolo del TG medio di NN deLIN, che risulta più elevato per questo motivo, e non perchè la predizione ottenuta con NN deLIN sia effettivamente migliore di quella fornita da NN dBGL.

7.4 Risultati: performance dei predittori al variare del PH

All'aumentare di PH diventa sempre più difficile stimare in maniera attendibile il valore glicemico futuro. Ciò è confermato dal peggioramento degli indici che quantificano la bontà della predizione.

In Figura 7.11 notiamo che, all'aumentare del PH, il profilo predetto continua ad essere regolare, come dimostrano i bassi valori di J, mentre da Figura 7.12 deduciamo che il RMSE peggiora linearmente con l'orizzonte di predizione, e segue un andamento molto simile per tutti i predittori.

Il TG medio percentuale peggiora gradualmente, passando da PH di 10 minuti a PH di 30 minuti, ma si assesta successivamente, per tutti i modelli,

7.4 Risultati: performance dei predittori al variare del PH

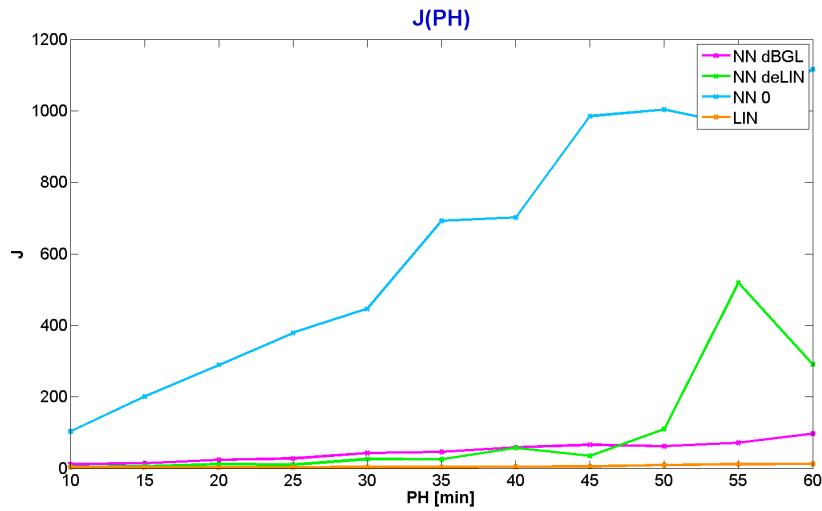


Figura 7.11: Andamento di J al variare del PH .

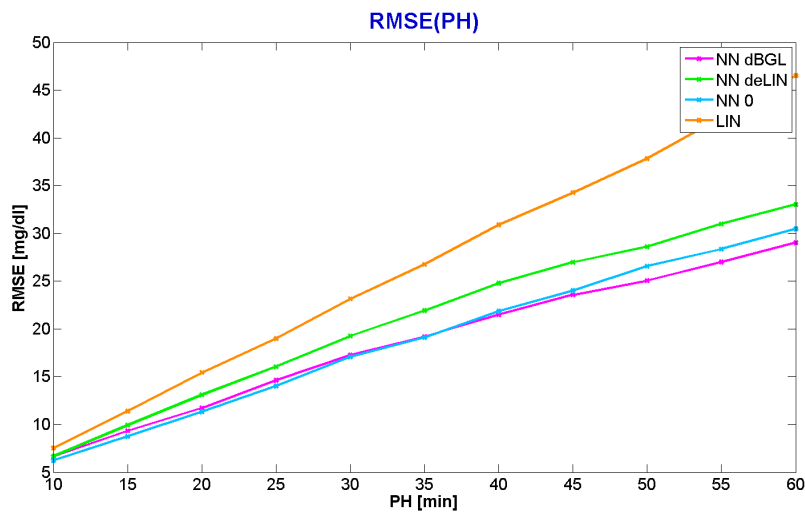
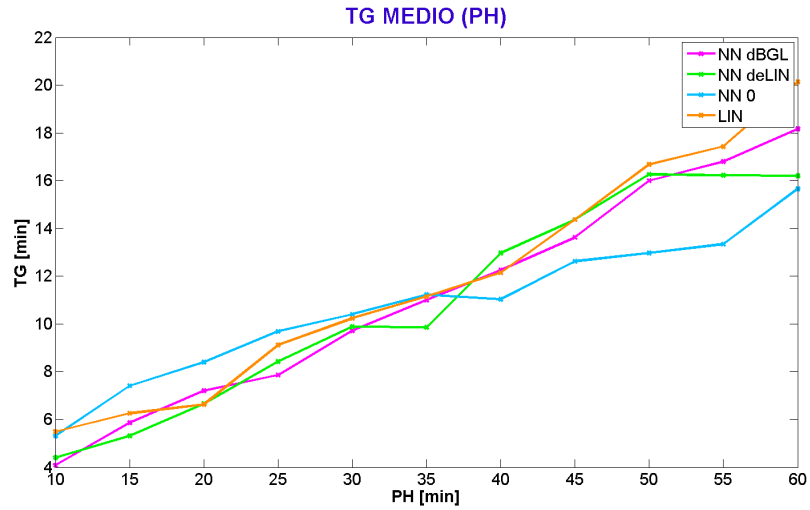
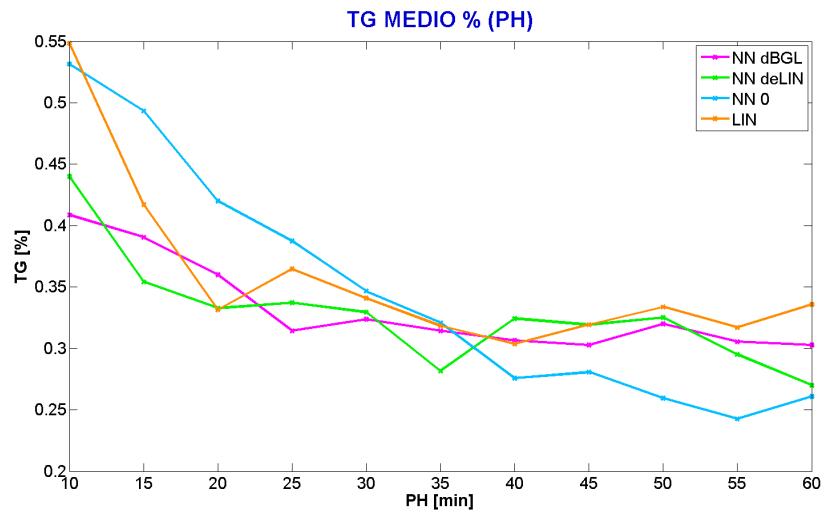


Figura 7.12: Andamento del RMSE al variare del PH .



(a) Guadagno temporale medio effettivo.

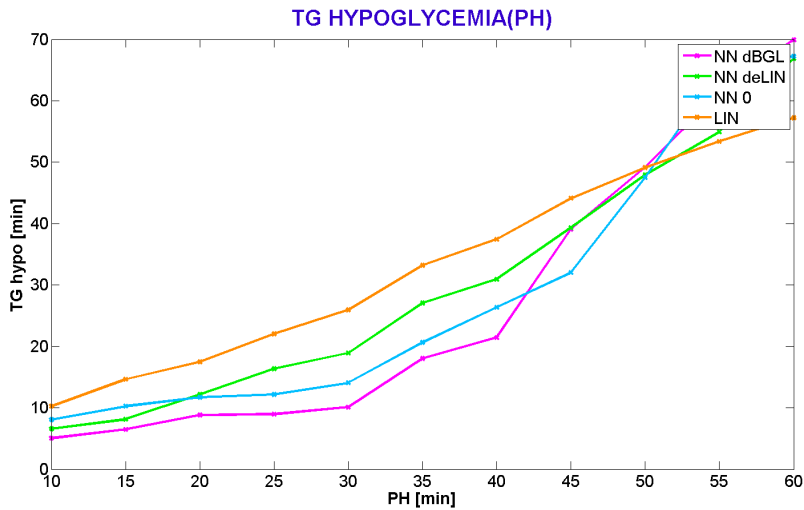


(b) Guadagno temporale medio percentuale.

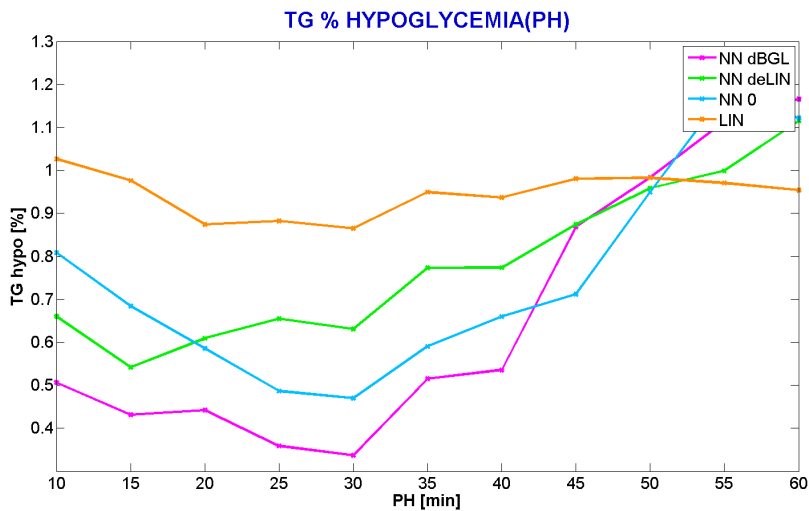
Figura 7.13: Guadagno temporale medio al variare del PH.

ad uno stesso valore asintotico, pari a circa 0.35.

Il guadagno temporale in corrispondenza dell'attraversamento di soglie critiche, al variare del PH, segue lo stesso andamento sia per l'ipo che per l'iperglicemia. Il TG del modello lineare è circa pari al PH, per qualsiasi



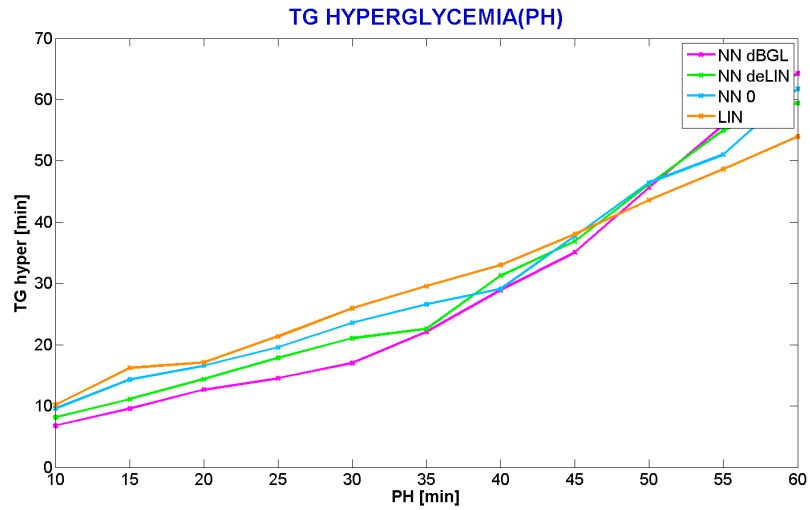
(a) Guadagno temporale effettivo.



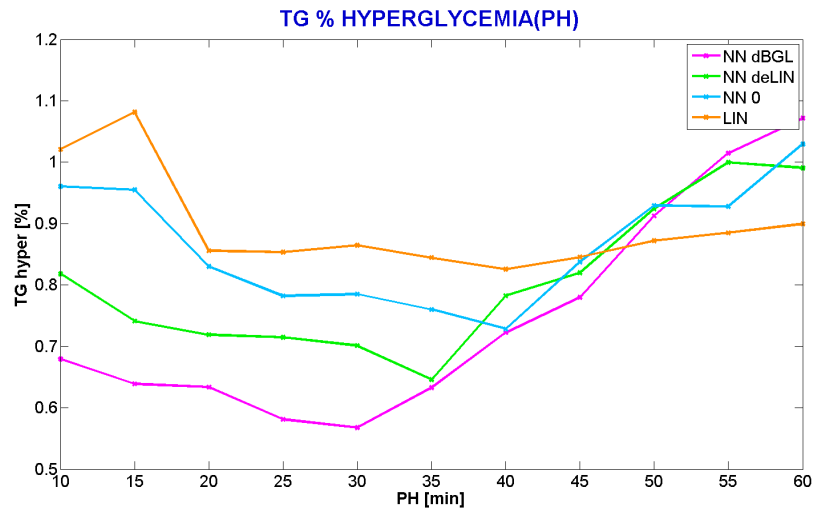
(b) Guadagno temporale percentuale.

Figura 7.14: Guadagno temporale in corrispondenza dell'attraversamento della soglia ipoglicemica, al variare del PH.

orizzonte di predizione. Per quanto riguarda le reti neurali, invece, per PHs inferiori a 30 minuti il TG peggiora gradualmente, all'aumentare del PH stesso, mentre per PHs superiori ai 30 minuti c'è un netto miglioramento



(a) Guadagno temporale effettivo.



(b) Guadagno temporale percentuale.

Figura 7.15: Guadagno temporale in corrispondenza dell'attraversamento della soglia iperglicemica, al variare del PH.

del guadagno temporale.

Questo risultato potrebbe essere fuorviante, se non si tenesse conto delle percentuali di vere e false crisi ipo e iperglicemiche che i vari modelli sono in grado di predire, al variare del PH.

Sia per le ipo che per le iperglicemie, in Figura 7.16, notiamo un crollo della percentuale di veri positivi per PHs superiori ai 30 minuti. Quindi, anche in questo caso, il miglioramento del TG è dovuto al fatto che per PHs inferiori a 30 minuti la maggior parte degli eventi difficili da predire viene individuata, ma con un ritardo tale da abbassare il valore del TG medio, mentre per PHs superiori ai 30 minuti questi eventi non vengono predetti. Di conseguenza essi non vanno ad abbassare il guadagno temporale medio, che risulta superiore poichè è calcolato tenendo conto dei pochi attraversamenti delle soglie critiche predetti correttamente e in maniera molto tempestiva. La percentuale di false crisi ipoglicemiche predette rimane pressochè costante, al variare del PH, per tutte e tre le reti neurali, mentre aumenta notevolmente per il modello lineare. La percentuale di false iperglicemie predette, invece, cresce per tutti i predittori all'aumentare del PH, ma rimane comunque più bassa per le reti neurali che per il modello lineare.

7.5 Conclusioni

I risultati ottenuti sui dati reali, benchè siano peggiori di quelli relativi ai dati simulati, sono comunque incoraggianti. L'uso di reti neurali per predire la glicemia permette di tenere conto di informazioni relative ai pasti e ad altri fattori che influenzano il livello di glucosio ematico. La rate of appearance dei carboidrati migliora la predizione delle iperglicemie e, spesso, se questo ingresso compare durante un trend di discesa della glicemia, evita la falsa predizione di una ipoglicemia. Come si può notare dalle percentuali di falsi positivi (Figura 7.16), le due nuove reti neurali proposte sono più affidabili del modello lineare e di NN0, poichè predicono un numero minore di falsi attraversamenti delle soglie ipo e iperglicemica. Tali attraversamenti, però, sono individuati con un anticipo significativamente inferiore rispetto a quello che si può ottenere col modello lineare.

Come già sottolineato presentando i risultati ottenuti sui dati simulati, non vi è una netta superiorità di un predittore rispetto agli altri. Inoltre, le performance dei vari modelli devono sempre essere valutate in base al fine della predizione. Se ci si concentra esclusivamente sull'attraversamento di soglie critiche il profilo predetto lontano da esse è di scarso interesse e può scostarsi anche di molto dal target, senza per questo peggiorare la qualità finale della predizione. In tal caso il modello lineare permette di ottenere un guadagno temporale maggiore, rispetto alle reti neurali, ed individua la quasi totalità di crisi ipo e iperglicemiche. Esso ha però il difetto di predire anche molte false ipo e iperglicemie. Se, invece, l'obiettivo è avere

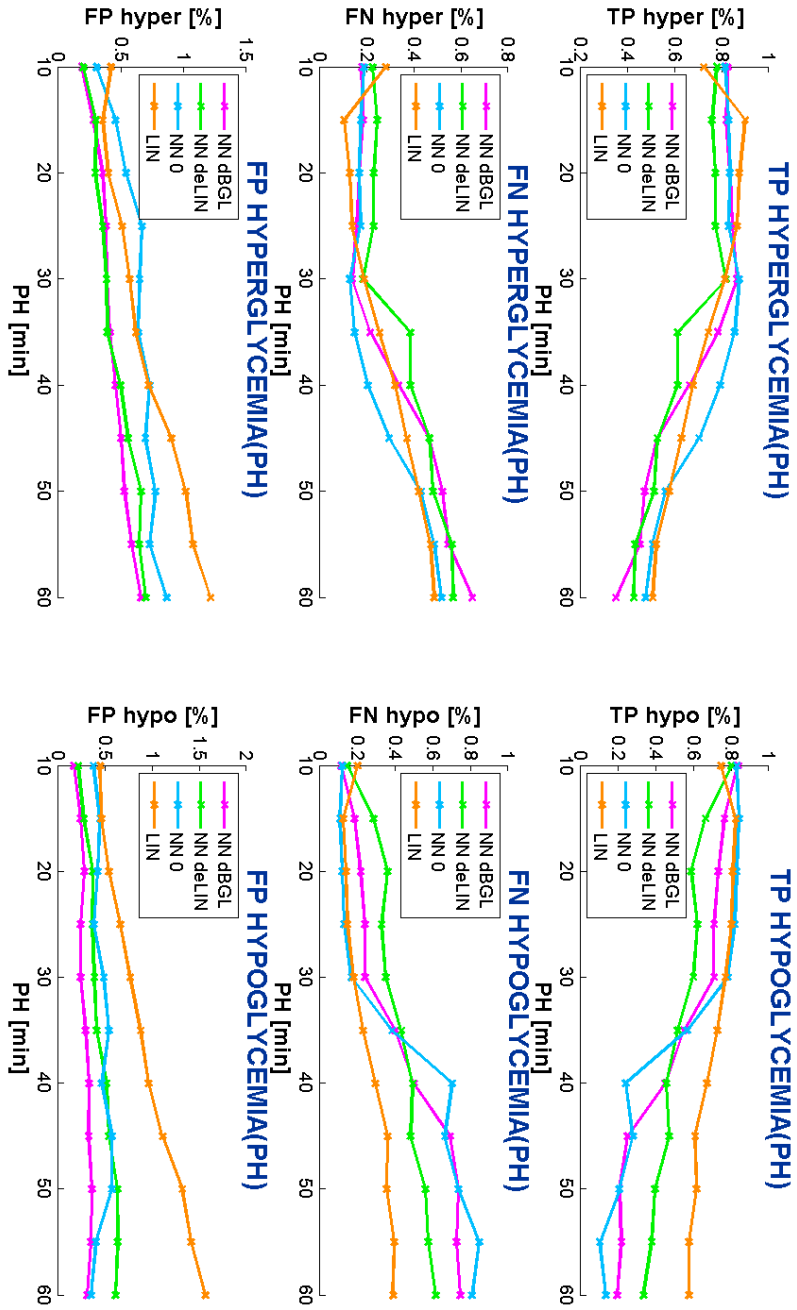


Figura 7.16: Percentuale di vere e false crisi ipo e iperglicemiche predette, al variare del PH.

in anticipo un'informazione abbastanza accurata di quale sarà la glicemia futura, il predittore dovrà fornire uscite veritiere e affidabili per qualsiasi valore glicemico. In questo caso le due nuove reti neurali sono da preferire, poichè sono mediamente più aderenti al target, sia durante le escursioni ipo e iperglicemiche, sia nella fascia euglicemica.

Capitolo 8

Conclusioni e possibili sviluppi futuri

Lo scopo di questo elaborato era sviluppare e testare nuove reti neurali per predire la glicemia futura utilizzando dati di glucosio registrati con sensori per il continuous glucose monitoring, e individuare potenziali pregi e difetti di questa categoria di predittori. Con questa tesi sono state proposte due semplici reti neurali feedforward, una per predire direttamente il livello glicemico futuro, (approccio 1, introdotto nella Sezione 5.1, a pag. 89), l'altra per correggere l'errore di predizione di un modello lineare tempo variante recentemente proposto in letteratura, (approccio 2, introdotto nella Sezione 5.2, a pag. 108). Entrambe le reti neurali, dopo la procedura di training, potrebbero essere potenzialmente incorporate in un sensore CGM, e sarebbero in grado di predire quasi autonomamente la glicemia futura, in quanto l'unico onere per il paziente sarebbe registrare tempestivamente la quantità di carboidrati ingeriti nel corso della giornata. Nella tesi, queste reti neurali sono state sottoposte a training e sono state successivamente testate sia su un dataset simulato, sia su un dataset reale, e le loro performance sono risultate confrontabili con quelle di un modello lineare tempo-variante [39] e con quelle di un'altra rete neurale [34], (indicata in questa tesi con l'abbreviazione NN0), entrambi descritti in letteratura.

8.1 Bilancio sui risultati ottenuti

I risultati relativi alla predizione del profilo glicemico con orizzonte di predizione di 30 minuti sono molto incoraggianti sia per la rete neurale sviluppata nell'approccio 1, sia per quella sviluppata nell'approccio 2. In particolare l'uscita delle reti neurali è mediamente più aderente al target, rispetto all'uscita del modello lineare. In tal modo, in base al valore glicemico predetto, il paziente può modificare in tempo reale la propria terapia,

per evitare crisi ipo o iperglicemiche e per mantenere all'interno del range euglicemico il proprio livello di glucosio nel sangue.

Entrambe le nuove reti, in un'implementazione pratica, richiederebbero al paziente solamente di registrare la quantità di carboidrati assunti durante il pasto. Successivamente la rate of appearance dei carboidrati è simulata sfruttando un generatore che non necessita di informazioni aggiuntive (a parte alcuni parametri fisiologici del paziente, che è sufficiente siano inseriti una volta, e rimangono invariati). Inoltre i dati registrati dal sensore CGM vengono utilizzati senza essere preventivamente filtrati, minimizzando in tal modo il carico computazionale richiesto per effettuare la predizione.

Bisogna sottolineare che anche le reti neurali, come tutti i predittori, forniscono solo un'indicazione del probabile livello glicemico futuro. È perciò necessario che il paziente sia adeguatamente istruito per poter sfruttare questa informazione nel miglior modo possibile e sia consapevole dei possibili errori di predizione, inevitabili con qualsiasi modello.

8.1.1 Risultati su dati simulati

Le performance delle reti neurali su dati simulati sono molto buone, come dimostrano i valori numerici dei vari indici calcolati, riportati in Tabella 8.1. In particolare, i profili predetti dalle 2 nuove reti neurali proposte sono caratterizzati da un guadagno temporale medio significativamente¹ superiore a quello dei profili ottenuti col modello lineare e con NN0. Per quanto riguarda la regolarità del profilo predetto, l'indice J mostra una netta superiorità delle 2 nuove reti rispetto a NN0, mentre esse sono inferiori al modello lineare. Bisogna però sottolineare che il valore del *forgetting factor* μ del modello lineare è scelto proprio minimizzando J, quindi il confronto è polarizzato a favore del predittore lineare stesso. In termini di RMSE, le performance ottenute con le 2 nuove reti proposte sono marcatamente migliore sia di quelle che caratterizzano NN0, sia di quelle fornite dal modello lineare. Infine, il numero di falsi attraversamenti ipo e iperglicemici generati dalla rete sviluppata nell'approccio 1 è nettamente inferiore a quello osservato per NN0 e per il modello lineare.

Nella valutazione dei risultati bisogna tenere conto dei limiti dei dati simulati. I profili, infatti, sono generati attraverso una complessa funzione matematica, in base ad alcuni parametri fisiologici dei pazienti virtuali, alla quantità di carboidrati ingeriti e alle dosi di insulina iniettata. Le reti neurali sono particolarmente indicate per individuare complesse funzioni

¹Per poter affermare che le differenze tra i risultati ottenuti con 2 predittori diversi sono significative è stato eseguito il *Wilcoxon rank-sum test*. Si tratta di un test statistico non parametrico equivalente al *t-test*, ma utilizzabile anche quando i dati non sono distribuiti in maniera gaussiana. Esso testa l'ipotesi nulla che i dati contenuti nei vettori \mathbf{x} e \mathbf{y} siano campioni indipendenti di una identica distribuzione continua, con mediana coincidente, contro l'ipotesi che la mediana dei 2 vettori sia differente. Il livello di significatività adottato è pari al 5%.

Simulated dataset

PH = 30 min

| | | NN dBGL | NN deLIN | NN0 | LIN |
|----------|---------|---------|----------|--------|-------|
| J | [-] | 5.65 | 4.07 | 149.56 | 7.25 |
| RMSE | [mg/dl] | 8.46 | 9.17 | 10.66 | 15.73 |
| TG | [min] | 23.00 | 24.50 | 17.00 | 11.00 |
| TG hypo | [min] | 11.00 | 19.83 | 15.23 | 32.86 |
| TG hyper | [min] | 27.00 | 24.25 | 22.40 | 21.40 |
| TP hypo | [%] | 61% | 52% | 62% | 69% |
| FP hypo | [%] | 23% | 56% | 28% | 120% |
| TP hyper | [%] | 84% | 81% | 95% | 92% |
| FP hyper | [%] | 17% | 26% | 48% | 84% |

Tabella 8.1: Valori mediani degli indici utilizzati per quantificare la bontà della predizione. Le caselle evidenziate in giallo corrispondono agli indici in cui una delle due nuove reti neurali proposte risulta migliore degli altri modelli.

descrivibili in modo analitico, a partire da coppie di ingressi e relative uscite. Inoltre le 2 nuove reti neurali sviluppate hanno, tra gli ingressi, la *rate of appearance* dei carboidrati assunti col pasto, che viene sfruttata anche dal simulatore [9] per generare i profili glicemici virtuali. Di conseguenza esse risultano avvantaggiate rispetto al modello lineare e rispetto a NN0.

8.1.2 Risultati su dati reali

Visti i promettenti risultati ottenuti sul dataset simulato, le 2 reti neurali sono state testate anche su un dataset reale, composto da misure rilevate in 29 soggetti diabetici di tipo 1, e le loro performance sono state nuovamente confrontate con quelle ottenute con NN0 e col modello lineare. I valori numerici mediani degli indici utilizzati per quantificarne le performance sono riportati in Tabella 8.2. Per quanto riguarda il guadagno temporale medio, le performance dei 4 modelli non presentano differenze significative. Il RMSE delle reti neurali, invece, è significativamente inferiore a quello del modello lineare, mentre l'indice J porta a prediligere il modello lineare. Il valore di J che caratterizza le 2 nuove reti neurali è comunque nettamente inferiore a quello ottenuto per NN0, come si può intuire anche osservando i profili predetti, (Figura 7.6 e Figura 7.7, a pag. 138 e a pag. 139, rispettivamente), che risultano molto più regolari delle serie glicemiche ottenute con NN0 e presentano solo sporadiche oscillazioni ad alta frequenza. Per quanto riguarda il guadagno temporale in corrispondenza di eventi ipo e iperglicemici, il modello lineare è superiore alle 2 nuove reti neurali. Vale però la pena sottolineare che le 2 nuove reti neurali producono un numero di falsi attraversamenti delle soglie critiche nettamente inferiore al modello

lineare e la loro predizione risulta perciò più affidabile, anche se fornisce un anticipo inferiore.

Real dataset

PH = 30 min

| | | NN dBGL | NN deLIN | NNO | LIN |
|----------|---------|---------|----------|--------|-------|
| J | [-] | 45.98 | 18.45 | 430.89 | 2.82 |
| RMSE | [mg/dl] | 16.85 | 19.31 | 16.92 | 23.07 |
| TG | [min] | 10.00 | 10.00 | 10.00 | 11.00 |
| TG hypo | [min] | 9.69 | 18.63 | 12.40 | 25.25 |
| TG hyper | [min] | 16.67 | 20.00 | 24.00 | 25.00 |
| TP hypo | [%] | 71% | 60% | 78% | 77% |
| FP hypo | [%] | 23% | 38% | 48% | 76% |
| TP hyper | [%] | 86% | 82% | 87% | 81% |
| FP hyper | [%] | 38% | 38% | 65% | 57% |

Tabella 8.2: Valori mediani degli indici utilizzati per quantificare la bontà della predizione. Le caselle evidenziate in giallo corrispondono agli indici in cui una delle due nuove reti neurali proposte risulta migliore degli altri modelli.

8.2 Possibili sviluppi futuri

Il campo delle reti neurali è molto ampio, ed esse sono state solo marginalmente utilizzate per la predizione della glicemia futura. Gli incoraggianti risultati ottenuti con questa tesi dimostrano che vale la pena approfondirne lo studio. Di seguito verranno elencati alcuni possibili sviluppi futuri.

8.2.1 Nuovi algoritmi per l'allenamento della rete neurale

Come abbiamo avuto modo di sottolineare più volte, il RMSE non è il miglior indice da utilizzare per quantificare la bontà di un predittore. Esso, infatti, non tiene minimamente conto della regolarità del profilo predetto, che è una caratteristica essenziale del segnale glicemico e della maggior parte dei segnali biologici. La funzione obiettivo minimizzata durante l'allenamento della rete è il quadrato del RMSE. Di conseguenza i pesi vengono determinati ottimizzando un indice il cui valore rispecchia solo marginalmente le performance della rete. Sarebbe quindi auspicabile utilizzare una funzione obiettivo diversa in fase di training. L'algoritmo di back propagation richiede che la funzione da minimizzare sia differenziabile e che dipenda in modo esplicito da tutti i pesi della rete. Quando queste condizioni non sono soddisfatte si può ricorrere ad altri algoritmi, come, per esempio, gli algoritmi stocastici.

Le 2 reti neurali descritte nel presente elaborato sono state sviluppate sfruttando il Neural Network Toolbox™ di Matlab® 7.10.0. In questo Toolbox non sono implementati gli algoritmi stocastici per l'allenamento delle reti neurali, inoltre le performance della rete possono essere quantificate solo attraverso funzioni analoghe al RMSE. Per questo motivo nel corso di questa tesi non sono stati indagati ulteriori algoritmi per l'allenamento della rete neurale. Si tratta però di un campo molto interessante e promettente all'interno del quale muoversi, consapevoli che l'individuazione di una funzione obiettivo adeguata ad esprimere le performance della rete neurale è tutt'altro che facile.

La possibilità di utilizzare una nuova funzione obiettivo permetterebbe anche di pesare in maniera diversa gli errori di predizione commessi dalla rete, in base agli obiettivi della predizione stessa. Si potrebbe, per esempio, penalizzare maggiormente gli errori in corrispondenza delle ipo e delle iperglicemie, tener conto del numero di falsi positivi, nonché del guadagno temporale ottenuto dalla rete neurale.

8.2.2 Reti ricorsive

Un ulteriore campo particolarmente interessante è quello delle reti neurali ricorsive. Le reti feedforward sono prive di memoria e, di conseguenza, quando vengono utilizzate per predire serie temporali, diviene cruciale il numero di campioni passati presentati loro in ingresso. Inoltre una rete feedforward risponde sempre allo stesso modo a una stessa combinazione di ingressi. Questo comportamento è poco fisiologico. È noto che a parità di glicemia attuale e a parità di carboidrati ingeriti, insulina iniettata ed esercizio fisico svolto, la glicemia futura potrebbe assumere valori differenti, in base allo stato fisico e psicologico del soggetto. Una rete ricorsiva è in grado di rispondere in modo diverso a uno stesso ingresso, poichè la sua uscita è influenzata anche dallo stato interno del sistema. Questa caratteristica risulta particolarmente attraente.

Nel corso del presente elaborato sono state testate le performance di semplici reti neurali ricorsive allenate con l'algoritmo di back propagation through time. I risultati ottenuti non si sono dimostrati migliori di quelli raggiunti con semplici reti feedforward, mentre il carico computazionale richiesto in fase di training è molto superiore. Bisogna sottolineare che queste reti spesso non arrivano a convergenza. Ciò potrebbe essere dovuto al fatto che l'algoritmo di back propagation è poco indicato per le reti ricorsive, poichè nel calcolare le modifiche da apportare ai pesi non tiene conto dell'effetto dei ritardi interni alla rete. Sarebbe perciò interessante indagare in maniera più approfondita questa tipologia di reti neurali, allenandole con algoritmi più adeguati².

²Nel Neural Network Toolbox™ di Matlab® 7.10.0 è implementato solamente l'algoritmo di back propagation through time.

8.2.3 Classificazione del livello glicemico

Spesso ciò che veramente interessa non è l'esatto valore puntuale della glicemia, ma la zona della scala glicemica a cui esso appartiene. Si potrebbe perciò trasformare il problema di predizione in un problema di classificazione. A tale scopo è necessario suddividere la scala di variazione glicemica in varie classi. Ogni classe contiene un intervallo più o meno ampio di valori glicemici che richiedono al paziente diabetico di mantenere un comportamento simile. In tal modo, pur non conoscendo l'esatto valore della glicemia futura, si può ugualmente intervenire quando la rete neurale predice una classe corrispondente a valori di ipo o di iperglicemia.

Un problema di classificazione viene affrontato in modo molto diverso da un problema di predizione. È perciò necessario ripetere le procedure di ottimizzazione della struttura della rete neurale e scegliere in maniera molto accorta gli ingressi.

Non è detto sia necessario ricorrere a una rete neurale per classificare il livello glicemico futuro, potrebbe essere sufficiente un classificatore statistico. Le reti neurali costituiscono comunque una soluzione molto interessante da testare, data la loro abilità a risolvere problemi di classificazione, come dimostra il loro ampio utilizzo in letteratura.

8.3 Conclusione

Dalla panoramica presentata e dai possibili sviluppi futuri descritti emergono le molte potenzialità delle reti neurali. Tuttavia il loro utilizzo comporta anche dei problemi e delle difficoltà, per esempio la loro natura intrinsecamente stocastica impedisce di controllarne completamente il comportamento e capire a fondo i motivi che hanno portato a una certa configurazione dei pesi invece che a un'altra. Questa loro caratteristica le rende particolarmente affascinanti e, assieme alla loro versatilità e alla loro abilità nel risolvere problemi di svariata natura, ha sicuramente contribuito alla grande diffusione di questi modelli in molti settori dell'economia e della scienza. I risultati ottenuti, inoltre, ne suggeriscono e incoraggiano anche l'utilizzo per la predizione della glicemia futura di pazienti diabetici monitorati con sensori CGM.

Bibliografia

- [1] M.J. Aitkenhead, A.J.S. McDonald, J.J. Dawson, R.P. Couper, G. Smart, M. Billett, D. Hope, and S. Palmer. A novel method for training neural networks for time-series prediction in environmental systems. *Ecological Modelling*, 162:87–95, 2003.
- [2] O. Amir, D. Weinstein, S. Zilberman, M. Less, D. Perl-Treves, H. Primack, A. Weinstein, E. Gabis, B. Fikhte, and A Karasik. Continuous non-invasive glucose monitoring technology based on occlusion spettroscopy. *Journal of Diabetes Science and Technology*, 1(4):463–469, 2007.
- [3] P.J. Angeline, G.M. Saunders, and J.B. Pollack. An evolutionary algorithm that constructs recurrent neural network. *IEEE Transactions on Neural Networks*, 5(1):54–66, 1994.
- [4] G. Baghdadi and A.M. Nasrabadi. Controlling blood glucose levels in diabetics by neural network predictor. In *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, pages 3216–3219, Lyon, France, August 2007. IEEE EMBS.
- [5] D.A. Bender. *Introduction to Nutrition and Metabolism*, chapter 10, pages 310–312. CRC Press, London, 3rd edition, 2002.
- [6] T. Bremer and D.A. Gough. Is blood glucose predictable from previous values? a solicitation for data. *Diabetes*, 48:445–451, 1999.
- [7] W.L. Clarke and B.P. Kovatchev. Continuous glucose sensors: Continuing questions about clinical accuracy. *Journal of Diabetes Science and Technology*, 1(5):669–675, 2007.
- [8] C. Cobelli and R. Bonadonna. *Bioingegneria dei Sistemi Metabolici*. Pàtron Editore, Bologna, 1st edition, 1998.
- [9] 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.

BIBLIOGRAFIA

- [10] A.K. El-Jabali. Neural network modeling and control of type 1 diabetes mellitus. *Bioprocess Biosystem Engineering*, 27:75–79, 2005.
- [11] E.K. Eskaf, O. Badawi, and T. Ritchings. Predicting blood glucose levels in diabetics using feature extraction and artificial neural networks. In *3rd International Conference on Information and Communication Technology: from Theory to Applications*, pages 1–6, Damascus, April 2008. IEEE.
- [12] A. Facchinetti. *On-line Filtering Algorithms for Continuous Glucose Monitoring*. PhD thesis, Università degli Studi di Padova, 2009.
- [13] B. Feldman, R. Brazg, S. Schwartz, and R. Weinstein. A continuous glucose sensor based on wired enzyme™ technology - results from a 3-day trial in patients with type 1 diabetes. *Diabetes Technology and Therapeutics*, 5(5):769–779, 2003.
- [14] B.J. Feldman and G.V. McGarraugh. Method of calibrating an analyte-measurement device, and associated methods, devices and systems. US Patent, 2007.
- [15] P. Felig, J.D. Baxter, A.E. Broadus, and L.A. Frohman. *Endocrinologia e Metabolismo*, chapter 19, pages 1033–1102. Mc Graw-Hill Libri Italia srl, Piazza Emilia 5, 20129 Milano, 2nd edition, Aprile 1991.
- [16] D.A. Finan, III F.J. Doyle, C.C. Palerm, W.C. Bevier, H.C. Zisser, L. Jovanović, and D.E. Seborg. Experimental evaluation of a recursive model identification technique for type 1 diabetes. *Journal of Diabetes Science and Technology*, 3(5):1192–1202, 2009.
- [17] A. Gani, A.V. Gribok, S. Rajaraman, W.K. Ward, and J. Reifman. Predicting subcutaneous glucose concentration in humans: Data-driven glucose modeling. *IEEE Transactions on Biomedical Engineering*, 56(2):246–254, 2009.
- [18] A.C. Guyton and J.E. Hall. *Fisiologia medica*, chapter 78, pages 970–974. Masson S.p.A, Via Paleocapa 7, 20121 Milano, 11th edition, 2006.
- [19] S. Haykin. *Neural Networks: a Comprehensive Foundation*. Macmillan College Publishing Company, 866 Third Avenue, New York, 10022, 1st edition, 1994.
- [20] M. Hudson Beale, M.T. Hagan, and H.B. Demuth. Neural network toolbox™ 7. user’s guide. http://www.mathworks.com/help/pdf_doc/nnet/nnet.pdf.
- [21] E. Hultén and M. Wahde. Improved time series prediction using evolutionary algorithms for the generation of feedback connections in neural networks. In *Computational Finance*, 2004.

- [22] G.A. Jastrebski and D.V. Arnold. Improving evolution strategies through active covariance matrix adaptation. In *Congress on Evolutionary Computation*, Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, July 2006. IEEE.
- [23] S. Kiranyaz, T. Ince, A. Yildirim, and M. Gabbouj. Evolutionary artificial neural networks by multi-dimensional particle swarm optimization. *Neural Networks*, 22:1448–1462, 2009.
- [24] P. Kok. Predicting blood glucose levels of diabetics using artificial neural networks. Master’s thesis, Delft University of Technology, 2004.
- [25] B.P. Kovatchev, S. Anderson, L. Heinemann, and W.L. Clarke. Comparison of the numerical and clinical accuracy of four continuous glucose monitors. *Diabetes Care*, 31(6):1160–1164, 2008.
- [26] B.P. Kovatchev, L.A. Gonder-Frederick, D.J. Cox, and W.L. Clarke. Evaluating the accuracy of continuous glucose monitoring sensors: Continuous error grid analysis illustrates by theasense freestyle navigator data. *Diabetes Care*, 27(8):1922–1928, 2004.
- [27] B.P. Kovatchev, M. Straume, D.J. Cox, and L.S. Farhy. Risk analysis of blood glucose data: A quantitative approach to optimizing the control of insulin dependent diabetes. *Journal of Theoretical Medicine*, 3:1–10, 2000.
- [28] K.K. Kuok, S. Harun, and S.M. Shamsuddin. Particle swarm optimization feedforward neural network for hourly rainfall-runoff modeling in bedup basin, malaysia. *International Journal of Civil and Environmental Engineering*, 9(10):20–39, 2009.
- [29] C.C. Palerm and W. Bequette. Hypoglycemia detection and prediction using continuous glucose monitoring- a study on hypoglycemic clamp data. *Journal of Diabetes Science and Technology*, 1(5):624–629, 2007.
- [30] C.C. Palerm, J.P. Willis, J. Desemone, and B.W. Bequette. Hypoglycemia prediction and detection using optimal estimation. *Diabetes Technology and Therapeutics*, 7(1):3–14, 2005.
- [31] S.M. Pappada, B.D. Cameron, and P.M. Rosman. Development of a neural network for prediction of glucose concentration in type 1 diabetes patients. *Journal of Diabetes Science and Technology*, 2(5):792–801, 2008.
- [32] S. Pereverzyev, S. Sampath, S. McKennoch, and J. Randlov. The design of a predictor of the blood glucose concentration based on regularization networks with adaptively chosen kernels and regularization parameters. November 2009.

- [33] J.L.C.M. Pijenburg. Nonlinear system identification using neural networks. Master's thesis, Eindhoven University of Technology, 1992.
- [34] C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E.J. Gómez, M. Rigla, A. de Leiva, and M.E. Hernando. Artificial neural network algorithm for on-line glucose prediction from continuous glucose monitoring. *Diabetes Technology and Therapeutics*, 12(1):81–88, 2010.
- [35] K. Rebrin and G.M. Steil. Can interstitial glucose assessment replace blood glucose measurements? *Diabetes Technology and Therapeutics*, 2(3):461–472, 2000.
- [36] J. Reifman, S. Rajaraman, A. Gribok, and W.K. Ward. Predictive monitoring for improved management of glucose levels. *Journal of Diabetes Science and Technology*, 1(4):478–486, 2007.
- [37] W. Sandham, D. Nikolettos, D. Hamilton, K. Paterson, A. Japp, and C. MacGregor. Blood glucose prediction for diabetes therapy using a recurrent artificial neural network. <http://www.eurasip.org/Proceedings/Eusipco/Eusipco1998/sessions/VM/WMS-4a/673/Sandham.PDF>.
- [38] W.A. Sandham, E.D. Lehmann, D.J. Hamilton, and M.L. Sandilands. Simulatind and predicting blood glucose levels for improved diabetes healthcare. In *4th Internationa Conference on Advances in Medical, Signal and Information Processing*, pages 121–124, Santa Margherita Ligure, Italy, July 2008. IET.
- [39] 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 Transactions on Biomedical Engineering*, 54(5):931–937, 2007.
- [40] N.S. Thomaidis, Y. Marinakis, M. Marinaki, and G.D. Dounias. Training neural network-based models using trading-related objective functions. http://www.nisis.risk-technologies.com/events/symp2007/papers/BD13_P_Thomaidis.pdf.
- [41] M.J. Tierney, J.A. Tamada, R.O. Potts, R.C. Eastman, K. Pitzer, N.R. Ackerman, and S.J. Fermi. The glucowatch biographer: a frequent, automatic and noninvasive glucose monitor. *Ann Med*, 32(9):632–641, 2000.
- [42] V. Tresp, T. Briegel, and J. Moody. Neural-network models for the blood glucose metabolism of a diabetic. *IEEE Transactions on Neural Networks*, 10(5):1204–1213, 1999.

- [43] E. Trifoglio. Il monitoraggio continuo del glucosio: Sviluppo di una metodologia per il confronto e l'ottimizzazione di algoritmi predittivi. Master's thesis, Università degli Studi di Padova, 2010.
- [44] M. Wahde. *Biologically Inspired Optimization Methods. An Introduction*. WIT Press, Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK, 1st edition, August 2008.
- [45] R. Weiss, Y. Yegorchikov, A. Shusterman, and I. Raz. Noninvasive continuous glucose monitoring using photoacoustic technology-results from the first 62 subjects. *Diabetes Technology and Therapeutics*, 9(1):68–74, 2007.
- [46] J.X Xu and J.S. Lim. A new evolutionary neural network for forecasting net flow of a car sharing system. In *Congress on Evolutionary Computation*, pages 1670–1676. IEEE, 2007.
- [47] S.X. Yang and N. Li. Power demand forecast based on optimised neural network by improved genetic algorithm. In *Proceedings of the 5th International Conference on Machine Learning and Cybernetics*, Dalian, August 2006. IEEE.
- [48] Z. Zainuddin, O. Pauline, and C. Ardil. A neural network approach in predicting the blood glucose level for diabetic patients. *International Journal of Computation Intelligence*, 5(1):72–79, 2009.
- [49] G. Zhang, B.E. Patuwo, and M.Y. Hu. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14:35–62, 1998.
- [50] R.A. Zitar and A.K. El-Jabali. Towards neural network model for insulin/glucose in diabetics-ii. *Informatica*, 29:227–232, 2005.
- [51] <http://www.abbottdiabetescare.com>.
- [52] <http://www.dexcom.com>.
- [53] <http://www.diadvisor.eu>.
- [54] <http://www.epicentro.iss.it>.
- [55] <http://www.medcompare.com/details/33037/CGMS-System-Gold-Continuous-Glucose-Monitoring.html>.
- [56] <http://www.menarini.com>.
- [57] <http://www.minimed.com/products/guardian/>.
- [58] <http://www.progettodiabete.org>.

BIBLIOGRAFIA

- [59] <http://www.solianis.com>.
- [60] <http://superagatoide.altervista.org/images/potenziale%20%27azione.jpg>.
- [61] <http://training.seer.cancer.gov/anatomy/nervous/tissue.html>.
- [62] <http://www.who.int>.