POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Biomedica

Tesi di Laurea Magistrale

Ottimizzazioni di piani di radioterapia tramite radial basis functions e algoritmi genetici: costruzione del sistema con e senza clustering

Relatori
Prof.ssa Gabriella Balestra
Prof.ssa Samanta Rosati

Candidato
Andrea Scotto

Ottobre 2022

Sommario

La radioterapia a fasci esterni è una delle possibili terapie per la cura del cancro. Una parte cruciale in tale trattamento è la creazione di un piano radioterapeutico che irradi la lesione tumorale con la massima dose possibile e che, al contempo, salvaguardi le regioni anatomiche circostanti non affette dalla patologia, dette organi a rischio. Nel caso del cancro della prostata, questi sono il retto e la vescica.

In uno studio precedente è stata dimostrata la validità di piani di radioterapia ottimizzati tramite un algoritmo genetico integrato con una rete neurale a base radiale, allenata per associare un valore di fitness alle soluzioni esplorate dall'algoritmo. In questo lavoro di tesi sono stati adoperati nuovi metodi per l'implementazione di due sistemi di ottimizzazione dei piani di radioterapia. Per il primo è stata allenata una rete neurale per ogni paziente e successivamente è stato implementato un algoritmo genetico per identificare i piani migliori per ciascun di essi. Per il secondo è stata condotta un'analisi sul numero migliore di gruppi in cui suddividere i pazienti sulla base di una caratterizzazione geometrica della prostata, effettuata in uno studio precedente. Successivamente sono stati ripetute le stesse tecniche del caso precedente, considerando ciascun cluster di pazienti in maniera indipendente.

Dopo aver prodotto dei piani con entrambi i metodi si è proceduto con la loro valutazione tramite software e al confronto con quelli standard prodotti dal software stesso, attualmente utilizzato per questo scopo. Dai risultati si evidenzia che i piani generati con entrambi i metodi sono clinicamente validi e presentano performance migliori rispetto a quelli del software.

Indice

Elenco delle tabelle 5

Elenco delle figure 7

1 Introduzione 10
- 1.1 Panoramica . . . 10
- 1.2 La prostata . . . 10
  - 1.2.1 Anatomia della prostata . . . 10
  - 1.2.2 Tumore della prostata . . . 11
    - 1.2.2.1 Diagnosi . . . 11
    - 1.2.2.2 Terapia . . . 12
- 1.3 Radioterapia . . . 12
  - 1.3.1 Radioterapia a fasci esterni . . . 13
  - 1.3.2 Pianificazione del trattamento . . . 14

2 Materiali e metodi 17
- 2.1 Il Dataset . . . 17
- 2.2 RayStation . . . 17
- 2.3 Reti neurali . . . 18
  - 2.3.1 Misure di similarità . . . 19
  - 2.3.2 Rete neurale a base radiale . . . 20
    - 2.3.2.1 Architettura della rete . . . 21
    - 2.3.2.2 Algoritmi di apprendimento . . . 22
- 2.4 Clustering . . . 23
  - 2.4.1 Metodo di clustering . . . 24
  - 2.4.2 Dendrogramma . . . 24
- 2.5 K-fold . . . 24
- 2.6 Algoritmi Genetici . . . 25
  - 2.6.1 Codifica delle variabili . . . 25
  - 2.6.2 Operatori evoluzionari . . . 26
  - 2.6.3 Algoritmo generico . . . 26
- 2.7 Analisi statistica . . . 27
  - 2.7.1 diagramma a scatola e baffi . . . 27
  - 2.7.2 Grafico a barre . . . 27

2.7.3 Grafico a dispersione . . . 27
2.7.4 Silhouette plot . . . 28
2.7.5 Test di Mann-Whitney . . . 29

3 Ottimizzazione dei piani di radioterapia 30
3.1 Radial Basis Function Neural Network . . . 30
3.1.1 Estrazione dei dati tramite RayStation . . . 30
3.1.2 Allenamento della rete . . . 32
3.1.3 Training set e validation set invariati . . . 32
3.1.3.1 Definizione dei set . . . 32
3.1.3.2 Analisi del numero di centri . . . 32
3.1.4 K-fold . . . 35
3.1.4.1 Definizione dei set . . . 35
3.1.4.2 Analisi del numero di centri . . . 35
3.1.5 Validazione del numero di centri . . . 38
3.2 Algoritmo genetico . . . 39
3.2.1 Codifica delle variabili . . . 40
3.2.2 Parametri . . . 40
3.2.3 Risultati . . . 41
3.3 Generazione dei piani di radioterapia . . . 42

4 Ottimizzazione dei piani di radioterapia per cluster di pazienti 43
4.1 Clustering dei pazienti . . . 43
4.2 Radial basis function neural network . . . 47
4.2.1 Analisi del numero di centri . . . 47
4.2.2 Assegnazione dei pazienti del test set al relativo cluster . . . 50
4.2.3 Validazione del numero di centri . . . 51
4.3 Algoritmo genetico . . . 53
4.4 Generazione dei piani di radioterapia . . . 54

5 Analisi dei piani radioterapeutici 55
5.1 Valutazione dei piani radioterapeutici . . . 55
5.2 Confronto dei piani radioterapeutici . . . 55

6 Conclusioni 58

A Valori MSE: allenamento RBFNN 59

B Confronto tra target e output della rete con 15 centri 72

C Valori MSE: allenamento RBFNN con il metodo k-fold 74

D Risultati delle analisi di clustering 87

E Assegnazione dei pazienti ai cluster 99

F Risultati piani di radioterapia ottimizzati 101

G Score percentuali dei piani radioterapeutici 106

Elenco delle tabelle

3.1 Dati in uscita da RayStation per il paziente 40 . . . 31
3.2 Valori minimi, medi e massimi della distribuzione dei valori MSE per le reti con 21, 24 e 30 centri considerando tutte le prove per ogni paziente . . . 36
3.3 Valori minimi, medi e massimi della distribuzione dei valori MSE per le reti con 21, 24 e 30 centri considerando la prova migliore per ogni paziente . . . 37
3.4 Valori MSE per la rete con 24 centri sul validation set dei pazienti del test set . . . 39
3.5 Parametri fissi del GA . . . 40
3.6 Parametri variabili del GA . . . 41
3.7 Analisi della tredicesima combinazione per i pazienti 106, 113 e 117 . . . 42
4.1 Variabili geometriche estratte dai pazienti. . . . 44
4.2 Clustering a due cluster. . . . 45
4.3 Clustering a tre cluster. . . . 45
4.4 Clustering a quattro cluster. . . . 45
4.5 Clustering a cinque cluster. . . . 45
4.6 Numero di reti che hanno un valore MSE inferiore a uno per ogni numero di centri . . . 49
4.8 Valori MSE per la rete con 21 centri sul validation set dei pazienti del cluster 1 del test set . . . 51
4.7 Assegnazione dei pazienti del test set ai due cluster . . . 52
4.9 Valori MSE per la rete con 26 centri sul validation set dei pazienti del cluster 2 del test set . . . 53
4.10 Valori MSE per la rete con 26 centri sul validation set dei pazienti del cluster 2 del test set . . . 54
5.1 Medie e mediane delle distribuzioni degli score associati ai piani dei pazienti del training set . . . 56
5.2 Confronto dei piani per i pazienti del training set . . . 56
5.3 Medie e mediane delle distribuzioni degli score associati ai piani dei pazienti del test set . . . 56
5.4 Confronto dei piani per i pazienti del training set . . . 57
5.5 P-value ottenuti con il test di Mann-Whitney . . . 57
E.1 Assegnazione dei pazienti del training set al relativo cluster di appartenenza 99
E.2 Assegnazione dei pazienti del training set al relativo cluster di appartenenza 100
F.1 Piani ottimizzati per i pazienti del training set . . . 104

F.2 Piani ottimizzati per i pazienti del test set 105
G.1 Punteggi dei piani di trattamento per i pazienti del training set 109
G.2 Punteggi dei piani di trattamento per i pazienti del test set 110

Elenco delle figure

1.1 Definizione dei volumi bersaglio. IV: Volume irradiato, TV: volume trattato [7]. . . . 15
1.2 Esempio di DVH prodotto dal TPS RayStation. . . . 16
2.1 Struttura del neurone [12]. . . . 18
2.2 Struttura a livelli di una rete neurale [13]. . . . 19
2.3 RNN e Feed Forward [14]. . . . 19
2.4 Architettura di una RBFNN [16]. . . . 21
2.5 Esempio di RBF gaussiana [17]. . . . 22
2.6 Rappresentazione grafica della divisione del dataset effettuata con il k-fold [22]. . . . 25
2.7 Esempio di diagramma a scatola e baffi 27
2.8 Esempio di grafico a barre 28
2.9 Esempio di grafico a dispersione 28
2.10 Esempio di silhouette plot [23] 29
3.1 Rappresentazione delle coppie di punti in uno spazio cartesiano 33
3.2 Scatter plot dei valori MSE per 10 pazienti del training set 33
3.3 Scatter plot dei valori MSE per i 140 pazienti del training set 34
3.4 Confronto target e valore di output restituito - paziente 156 34
3.5 Istogramma della frequenza del numero di centri migliore 35
3.6 Suddivisione dei punti in sei gruppi 36
3.7 Istogramma della frequenza del numero di centri migliore 36
3.8 Numero di reti che hanno un valore MSE inferiore a 1 per ogni numero di centri 37
3.9 Boxplot delle distribuzioni dei valori MSE per le reti con 21, 24 e 30 centri considerando tutte le prove per ogni paziente 37
3.10 Boxplot delle distribuzioni dei valori MSE per le reti con 21, 24 e 30 centri considerando la prova migliore per ogni paziente 38
3.11 Esempio di soluzione 40
3.12 Parametro di stabilità 42
4.1 Dendrogramma. . . . 44
4.2 Dendrogrammi. . . . 45
4.3 Coefficienti di silhouette per le 4 tipologie di clustering. . . . 46
4.4 Silhouette plot. . . . 46
4.5 Istogramma della frequenza del numero di centri migliore 48

4.6 Istogramma della frequenza del numero di centri migliore 48
4.7 Boxplot delle distribuzioni dei valori MSE migliori per le reti con il numero di centri migliore per due cluster 50
4.8 Boxplot delle distribuzioni dei valori MSE migliori per le reti con il numero di centri migliore per tre cluster 50
4.9 Criterio 1 54
5.1 Boxplot delle distribuzioni degli score associati ai piani dei pazienti del training set 56
5.2 Boxplot delle distribuzioni degli score associati ai piani dei pazienti del test set 57
A.1 Valori MSE sul validation set dei pazienti 2, 4, 5, 6, 7 e 8 59
A.2 Valori MSE sul validation set dei pazienti 9, 10, 12, 14, 16, 17, 18, 19, 20, 22 e 25 60
A.3 Valori MSE sul validation set dei pazienti 26, 27, 29, 30, 31, 33, 34, 36, 40, 43, 45 e 48 61
A.4 Valori MSE sul validation set dei pazienti 49, 52, 58, 60, 62, 63, 64, 65, 68, 73, 87 e 89 62
A.5 Valori MSE sul validation set dei pazienti 90, 92, 94, 96, 109, 112, 115, 116, 118, 120, 121 e 122 63
A.6 Valori MSE sul validation set dei pazienti 126, 127, 128, 132, 136, 137, 140, 143, 145, 146, 147 e 150 64
A.7 Valori MSE sul validation set dei pazienti 151, 152, 153, 154, 155, 156, 158, 159, 160, 162, 163 e 166 65
A.8 Valori MSE sul validation set dei pazienti 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177 e 179 66
A.9 Valori MSE sul validation set dei pazienti 180, 182, 183, 184, 185, 187, 188, 189, 190, 191, 193 e 194 67
A.10 Valori MSE sul validation set dei pazienti 196, 197, 198, 201, 203, 205, 207, 208, 209, 210, 212 e 213 68
A.11 Valori MSE sul validation set dei pazienti 215, 217, 218, 219, 220, 221, 223, 225, 226, 227, 229 e 230 69
A.12 Valori MSE sul validation set dei pazienti 232, 233, 234, 237, 238, 240, 242, 243, 245, 246, 247 e 248 70
A.13 Valori MSE sul validation set dei pazienti 252 e 254 71
B.1 Confronto target e valore di output restituito - paziente 53 72
B.2 Confronto target e valore di output restituito - paziente 131 72
B.3 Confronto target e valore di output restituito - paziente 6 73
B.4 Confronto target e valore di output restituito - paziente 62 73
B.5 Confronto target e valore di output restituito - paziente 95 73
C.1 Valori MSE sul validation set dei pazienti 2, 4, 5, 6, 7 e 8 74
C.2 Valori MSE sul validation set dei pazienti 9, 10, 12, 14, 16, 17, 18, 19, 20, 22 e 25 75
C.3 Valori MSE sul validation set dei pazienti 26, 27, 29, 30, 31, 33, 34, 36, 40, 43, 45 e 48 76

C.4 Valori MSE sul validation set dei pazienti 49, 52, 58, 60, 62, 63, 64, 65, 68, 73, 87 e 89 . . . 77
C.5 Valori MSE sul validation set dei pazienti 90, 92, 94, 96, 109, 112, 115, 116, 118, 120, 121 e 122 . . . 78
C.6 Valori MSE sul validation set dei pazienti 126, 127, 128, 132, 136, 137, 140, 143, 145, 146, 147 e 150 . . . 79
C.7 Valori MSE sul validation set dei pazienti 151, 152, 153, 154, 155, 156, 158, 159, 160, 162, 163 e 166 . . . 80
C.8 Valori MSE sul validation set dei pazienti 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177 e 179 . . . 81
C.9 Valori MSE sul validation set dei pazienti 180, 182, 183, 184, 185, 187, 188, 189, 190, 191, 193 e 194 . . . 82
C.10 Valori MSE sul validation set dei pazienti 196, 197, 198, 201, 203, 205, 207, 208, 209, 210, 212 e 213 . . . 83
C.11 Valori MSE sul validation set dei pazienti 215, 217, 218, 219, 220, 221, 223, 225, 226, 227, 229 e 230 . . . 84
C.12 Valori MSE sul validation set dei pazienti 232, 233, 234, 237, 238, 240, 242, 243, 245, 246, 247 e 248 . . . 85
C.13 Valori MSE sul validation set dei pazienti 252 e 254 . . . 86
D.1 Variabilità intra-cluster . . . 87
D.2 Distanze inter-cluster . . . 88
D.3 Variabile 2 . . . 88
D.4 Variabile 3 . . . 89
D.5 Variabile 4 . . . 89
D.6 Variabile 5 . . . 90
D.7 Variabile 6 . . . 90
D.8 Variabile 7 . . . 91
D.9 Variabile 8 . . . 91
D.10 Variabile 9 . . . 92
D.11 Variabile 10 . . . 92
D.12 Variabile 11 . . . 93
D.13 Variabile 12 . . . 93
D.14 Variabile 13 . . . 94
D.15 Variabile 14 . . . 94
D.16 Variabile 15 . . . 95
D.17 Variabile 16 . . . 95
D.18 Variabile 17 . . . 96
D.19 Variabile 18 . . . 96
D.20 Variabile 19 . . . 97
D.21 Variabile 20 . . . 97
D.22 Variabile 21 . . . 98

Capitolo 1

Introduzione

1.1 Panoramica

Il carcinoma prostatico è divenuto la patologia tumorale più frequente nella popolazione maschile dei paesi occidentali, rappresentando circa il 19% di tutti i tumori diagnosticati nell'uomo [1]. Gli approcci terapeutici per il tumore della prostata dipendono da diversi fattori, quali l'età del paziente, il suo stato di salute generale e le caratteristiche del cancro.

Un trattamento standard valido per quasi tutti gli stadi di rischio è la radioterapia [2], che consiste nell'erogazione di radiazioni ionizzanti per distruggere le cellule patologiche e arrestarne lo sviluppo. La pianificazione del trattamento è una fase estremamente importante perché la cura deve essere personalizzata per ciascun paziente al fine di somministrare la dose curativa massimizzando il controllo sulle cellule neoplastiche e cautelando, al contempo, le strutture sane adiacenti. Tali strutture sono dette organi a rischio (OAR) e nel caso del tumore della prostata coincidono con la vescica e l'intestino retto.

L'irradiazione degli OAR può causare disturbi urinari e sintomi intestinali, ad esempio dei bruciori durante la minzione o la perdita di sangue e muco dal retto dopo la defecazione. La maggior parte di questi effetti sono acuti e scompaiono gradualmente nel giro di un mese dalla conclusione della terapia, ma, a volte, a distanza di anni dalla conclusione dalla terapia, possono comparire effetti collaterali tardivi che tendono a diventare permanenti [3]. Date le conseguenze di una simile procedura è doveroso ricorrere all'uso di un sistema di pianificazione del trattamento (Treatment Planning System, TPS) automatico per supportare le decisioni dell'oncologo radioterapista.

1.2 La prostata

1.2.1 Anatomia della prostata

La prostata è un organo ghiandolare appartenente all'apparato genitale maschile dei mammiferi. Nell'uomo è collocata al di sotto della vescica, circa 5 cm anteriormente al retto e ha tipicamente una forma piramidale, con base verso l'alto e apice in basso, che ricorda

quella di una castagna. Nei soggetti sani ha un diametro di circa 4 cm ed è formata da una ricca componente fibro-muscolare di forma rotondeggiante.

La sua funzione principale è quella di produrre una secrezione lievemente acida, detta liquido prostatico, che contribuisce per il 30% al volume del liquido seminale. Dal momento che la componente muscolare circonda la prima porzione dell'uretra, la prostata svolge anche un ruolo nella minzione, contribuendo a regolarizzare il flusso dell'urina.

1.2.2 Tumore della prostata

Un tumore si forma quando il normale e ordinato meccanismo di divisione cellulare viene improvvisamente alterato a favore di un processo di moltiplicazione e differenziazione incontrollato da cui si formerà un nodulo che continuerà a crescere in maniera abnorme.

Esistono molte forme di tumore della prostata ma le loro cause non sono ancora certe. La concomitanza dell'avanzare dell'età e della comparsa dell'ipertrofia prostatica fa ipotizzare che la variazione dell'equilibrio ormonale tipica dell'andropausa ricopra un ruolo importante. Oggigiorno, infatti, è noto che la crescita e la sopravvivenza delle cellule tumorali nella prostata è influenzata dal testosterone che circola nel sangue [4]. Tra gli altri fattori di rischio vi sono la storia familiare, il fumo, l'inquinamento ambientale e una dieta ricca di latticini e povera di frutta e verdura.

L'incidenza di questa patologia è in continua crescita ma alla sua base, più che la presenza di fattori di rischio, vi sono due considerazioni. La prima è l'aumento dell'età media della popolazione. Infatti, sebbene il tumore si presenti di rado prima dei quarantacinque anni, le probabilità di ammalarsi aumentano dopo i cinquant'anni e circa due tumori su tre vengono scoperti in uomini ultrasessantacinquenni. La seconda è la maggiore probabilità di diagnosi di tale patologia attraverso tecniche di screening precoce, che contribuiscono alla diminuzione del tasso di mortalità.

1.2.2.1 Diagnosi

La diagnosi del tumore della prostata può essere effettuata attraverso diversi esami. Il primo controllo è l'esplorazione rettale digitale (ERD) effettuata dallo specialista urologo in pochi minuti, ma è un'indagine con componente soggettiva, a bassa specificità e sensibilità. Poiché non consente di palpare tumori troppo piccoli ha un ruolo limitato e la sua negatività non esclude la presenza di una neoplasia in fase iniziale. L'ecografia transrettale e la flussometria permettono di rilevare rispettivamente le dimensioni della prostata e l'eventuale pressione sull'uretra e la velocità del getto urinario, da cui si può dedurre una rilevante ostruzione del flusso.

Negli ultimi anni l'unica indagine per la diagnosi precoce del tumore della prostata è il test dell'antigene prostatico specifico (PSA). Il PSA è una proteina prodotta dalla prostata, presente nel liquido seminale e in condizioni fisiologiche in minima quantità anche nel sangue, per cui per effettuare questa tipologia d'esame è sufficiente un prelievo sanguigno. I valori considerati nella norma sono inferiori a 4 ng/ml, tuttavia, anche in questo caso, i risultati non devono essere interpretati in maniera assoluta in quanto il valore può aumentare con l'età o in presenza di altre condizioni.

Se dopo i primi accertamenti si sospetta la presenza di un tumore si ricorre alla biopsia, cioè al prelievo di piccoli frammenti di ghiandola da sottoporre ad analisi di laboratorio in grado di diagnosticare e confermare la lesione.

La gravità della patologia viene definita attraverso due parametri: il grading e lo stadio [5]. Col termine grading si fa riferimento al grado di differenziazione delle cellule, legato all'aggressività del tumore e indica la rapidità con cui la massa tende a crescere e a diffondersi ad altri organi. Esso viene stabilito in seguito alla biopsia da parte dell'anatomopatologo ed è misurato tramite il grado Gleason. Quest'ultimo varia da 1 a 5 ma per definire il punteggio si sommano il primo e il secondo grado più comuni assegnati ai campioni prelevati, per cui può assumere valori compresi tra 2 e 10:

- Da 2 a 4: tumore in crescita ben differenziato e con scarsa tendenza a diffondersi ad altri organi;
- Da 5 a 6: tumore moderatamente differenziato con un grado di aggressività intermedio;
- Da 7 a 10: tumore aggressivo scarsamente differenziato e con maggiore probabilità di diffondersi ad altri organi.

La stadiazione consente di definire la localizzazione e le dimensioni del tumore attraverso il sistema TNM, in cui T si riferisce all'estensione della neoplasia, N allo stato dei linfonodi (0 se non sono stati intaccati, 1 alternativamente) e M alla presenza di metastasi (0 se assenti, 1 se presenti). La correlazione del punteggio di Gleason, del sistema TMN e del valore di PSA consente di classificare la malattia in una delle tre classi di rischio: basso, intermedio e alto.

1.2.2.2 Terapia

La terapia più adatta per ciascun paziente dipende da un'analisi delle sue caratteristiche, ad esempio l'età e l'aspettativa di vita, e dal grado di rischio. Per pazienti che presentano un basso rischio si procede con una sorveglianza attiva, effettuando controlli frequenti in modo da seguire l'evoluzione della malattia e individuare eventuali cambiamenti. Nei pazienti in cui il tumore è in stadio intermedio o metastatico può essere indicata una terapia ormonale, nota come terapia di deprivazione androgenica (ADT), che ha lo scopo di ridurre il livello di testosterone, ormone che stimola la crescita delle cellule tumorali. Nei casi più gravi la scelta ricade spesso sull'asportazione tramite intervento chirurgico, sulla radioterapia o su un approccio combinato delle due.

1.3 Radioterapia

La radioterapia è una branca della medicina che utilizza le radiazioni ionizzanti per provocare la necrosi cellulare. Le radiazioni ionizzanti sono onde elettromagnetiche dotate di un'energia talmente alta da danneggiare il DNA delle cellule tramite attacco fisico diretto oppure tramite attacco indiretto, attraverso la creazione di particelle radioattive, dette

radicali liberi, che si legano alle molecole danneggiandone le strutture. Conseguentemente, le cellule tumorali che subiscono danni irreversibili al DNA perdono la capacità di riprodursi e muoiono.

La quantità di energia assorbita da un mezzo per unità di massa in seguito all'esposizione è detta dose assorbita e l'unità di misura nel Sistema internazionale è il gray (Gy). Tale grandezza è estensiva, cioè non è direttamente legata al danno biologico causato ad un tessuto in quanto una stessa quantità di energia assorbita da due radiazioni diverse produce effetti biologici differenti sull'organismo. Per questo motivo è più conveniente usare la dose uniforme equivalente (EUD: Equivalent Uniform Dose), definita come la dose che, se distribuita uniformemente sul volume target, porta allo stesso numero di cellule morte.

La radioterapia può essere classificata in due modi:

- In base allo specifico obiettivo con cui viene prescritta al paziente, che può variare a seconda delle sue condizioni cliniche e dal tipo di tumore:
  - Radioterapia curativa: ha lo scopo di eliminare radicalmente il tumore;
  - Radioterapia preoperatoria: ha lo scopo di ridurre le dimensioni del tumore per rendere più facile l'operazione di asportazione chirurgica e di ridurre il rischio che alcune cellule malate possano diffondersi durante l'intervento;
  - Radioterapia postoperatoria: ha lo scopo di aumentare la probabilità che ogni cellula cancerosa residua dopo l'asportazione chirurgica venga eliminata;
  - Radioterapia palliativa: ha lo scopo di allentare la crescita del tumore e alleviarne i sintomi nei casi in cui la malattia non possa essere curata;
- In base alla tecnica di somministrazione:
  - Radioterapia a fasci esterni: il sito tumorale è irradiato da raggi prodotti da una sorgente posizionata all'esterno del corpo;
  - Radioterapia interna o brachiterapia: le radiazioni vengono emesse da una sorgente radioattiva inserita all'interno di minuscolo sferette, dette semi, che sono impiantate nel paziente all'interno del tessuto tumorale o in sua prossimità. In questo tipo di somministrazione le sorgenti sono confinate, cioè non possono entrare nel circolo sanguigno e non vengono metabolizzate.

Gli effetti collaterali sono dovuti principalmente alla possibilità che le radiazioni colpiscano gli OAR, sebbene in misura minore rispetto al sito tumorale, provocando fenomeni infiammatori nella zona trattata. Questi possono essere a breve termine se compaiono durante il trattamento o a lungo termine se compaiono dopo qualche tempo. I primi, nella maggior parte dei casi, sono di modesta entità e svaniscono nel corso di qualche settimana mentre i secondi possono persistere per tutta la vita.

1.3.1 Radioterapia a fasci esterni

La radioterapia a fasci esterni è caratterizzata dall'uso di raggi X ad alta energia prodotti da una sorgente esterna, detta acceleratore lineare. Tale apparecchiatura è costituita

da un lettino, su cui il paziente giace durante il trattamento ed eventualmente viene immobilizzato per garantire il corretto posizionamento, e da una testata rotante, detta gantry, che eroga le radiazioni intorno al paziente.

La radioterapia non è dolorosa e un singolo trattamento può durare da una decina di minuti fino a un'ora nel caso di irradiazione corporea totale, sebbene l'erogazione vera e propria duri solo pochi minuti.

Esistono diverse tecniche di radioterapia a fasci esterni, le più utilizzate sono:

- Radioterapia conformazionale tridimensionale (3D-CRT): sfrutta un collimatore multilamellare per focalizzare il fascio di radiazioni concentrando dosi più elevate sul tumore ed esponendo le cellule sane circostanti a dosaggi più bassi.

- Radioterapia a modulazione di intensità (IMRT): è un'evoluzione della tecnica precedente in quanto utilizza un collimatore con delle lamelle mobili per irradiare tumori con forma complessa e irregolare.

- Volumetric Modulated Arc Therapy (VMAT): favorisce la massima focalizzazione delle radiazioni sui tessuti tumorali, che vengono colpiti da tutte le angolazioni possibili grazie alla continua rotazione della testata attorno al paziente.

- Radioterapia adattiva: a ogni seduta viene elaborato un nuovo piano di radioterapia per tenere conto dell'anatomia del paziente in quella specifica giornata.

- Radioterapia 4D (4D-RT): tiene conto del movimento degli organi dovuti alla respirazione e alla peristalsi intestinale durante la seduta.

- Radioterapia stereotassica (SBRT): permette di erogare un'alta dose su piccoli tumori garantendo un'elevata accuratezza.

- Adroterapia: al posto dei raggi X usa gli adroni, cioè protoni e ioni carbonio. Dal momento che tali particelle sono più energetiche attraversano i tessuti in superficie più velocemente, senza danneggiarli, rilasciando la loro energia nella massa tumorale.

1.3.2 Pianificazione del trattamento

Come descritto precedentemente esistono diversi trattamenti di radioterapia. Non ne esiste in assoluto uno migliore o più efficace rispetto agli altri, ma per ciascun paziente viene prescritto quello più adatto in base alle sue caratteristiche. Oltre a ciò, durante la pianificazione del trattamento deve essere stabilita la dose di radiazioni complessiva necessaria a distruggere il tumore, in quante frazioni deve essere somministrata e con quale frequenza [6].

Il piano di cura di un paziente viene redatto in occasione della prima visita medica da un'équipe di specialisti. Le figure coinvolte sono:

- Medico radioterapista oncologo: conosce con competenza le radiazioni ionizzanti e sceglie il trattamento e le tecniche da utilizzare.

- Fisico medico: ha la responsabilità del funzionamento e dei controlli di sicurezza delle apparecchiature e collabora con il radioterapista nel calcolo della distribuzione della dose.
- Tecnico di radioterapia: gestisce le apparecchiature utili e segue il paziente durante le varie sessioni. In particolare, predispone il corretto posizionamento del paziente all'inizio della seduta.
- Infermiere: segue i pazienti durante le visite e le sedute di terapia.

Alla visita segue una procedura di simulazione e centraggio per definire con la massima precisione la zona da irradiare (target), proteggendo dalle radiazioni gli OAR, che nel caso di tumore alla prostata sono il retto e la vescica. Il paziente viene, quindi, sottoposto a una tomografia computerizzata (TC) o a una risonanza magnetica nucleare (RMN) per acquisire le immagini necessarie e definire il suo posizionamento sul lettino. In seguito, la zona da irradiare viene delimitata attraverso dei tatuaggi permanenti sulla cute, in modo che non si cancellino nel tempo che intercorre tra le varie sedute. I volumi contornati sono:

- Gross tumor volume (GTV): volume di neoplasia palpabile o visibile;
- Clinical target volume (CTV): volume che comprende il GTV e l'estensione microscopica che deve essere trattata;
- Planning target volume (PTV): volume che comprende il CTV e i margini di sicurezza intorno ad esso affinché si abbia la certezza che la dose sia erogata al CTV.

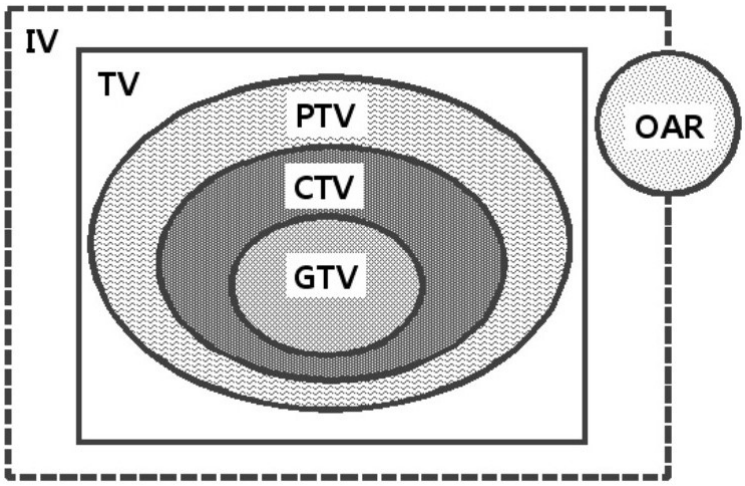


Figura 1.1: Definizione dei volumi bersaglio. IV: Volume irradiato, TV: volume trattato [7].

In questo ambito risultano fondamentali i Treatment Planning Systems (TPS), cioè software di pianificazione automatica dei piani di radioterapia. Dopo aver importato

i dati e le immagini del paziente per l'identificazione del tumore, il sistema calcola la distribuzione di dose che ci si aspetta nei tessuti del paziente, considerando anche la penetrazione dei raggi sulla base dei tipi di tessuto che la radiazione incontrerà lungo il suo cammino. Lo scopo è quello di ottenere una distribuzione ottimale secondo quanto richiesto per il volume target e la quantità di dose tollerata dai tessuti sani circostanti.

Il genere la pianificazione è effettuata tramite il metodo detto Inverse Planning. In questo caso l'addetto alla pianificazione definisce i limiti di dose per il volume target, gli OAR e alcuni parametri per l'irradiazione. In seguito, un algoritmo di ottimizzazione determina automaticamente i parametri del piano che soddisfano al meglio le richieste attraverso la minimizzazione di una funzione obiettivo.

In alcuni centri, invece, viene utilizzato un metodo più tradizionale, detto Forward Planning, che si basa su un approccio di tipo "trial and error". Con questa modalità deve essere manualmente specificato e modificato l'insieme dei parametri del piano finché non si raggiunge un piano soddisfacente. I risultati dipendono dall'esperienza, dall'impegno e dalle decisioni di colui che si occupa della pianificazione. Inoltre, è meno efficiente dal punto di vista del tempo richiesto per la pianificazione.

La valutazione del piano elaborato avviene tramite l'analisi dell'istogramma dose-volume (DVH), che permette di osservare la quantità di dose ricevuta da ciascun organo irradiato. Sull'asse delle ascisse è riportato il valore di dose cumulativa specifica, sull'asse delle ordinate è riportato il valore percentuale di volume che ha assorbito una certa dose. Dal momento che nel piano ideale gli OAR assorbono dosi limitate, le curve di tali organi dovrebbero trovarsi sulla sinistra del grafico. Viceversa, gli organi target devono assorbire una dose maggiore, per cui le loro curve dovrebbero trovarsi sulla destra.

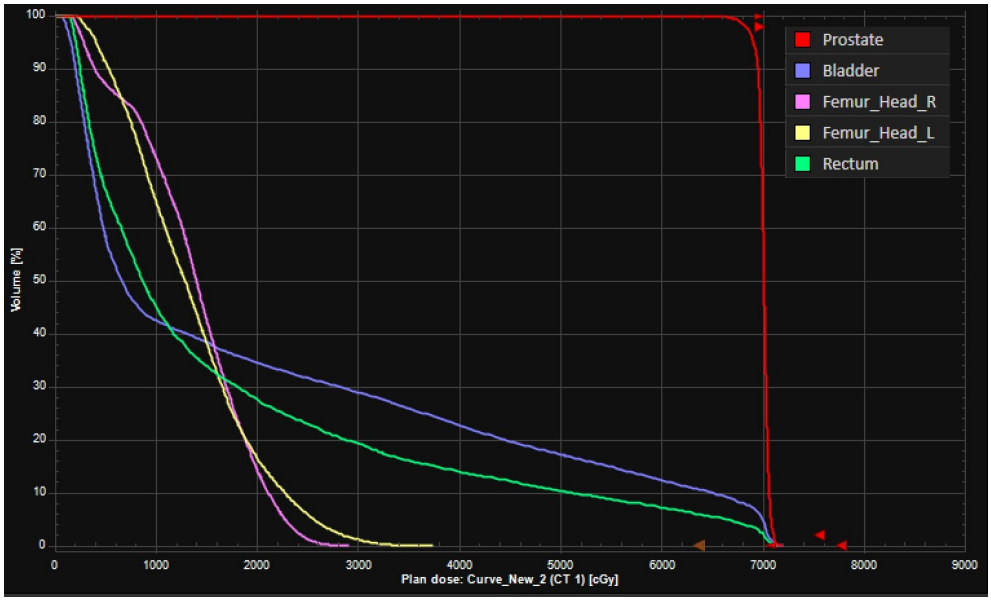


Figura 1.2: Esempio di DVH prodotto dal TPS RayStation.

Capitolo 2

Materiali e metodi

I metodi di seguito descritti sono stati sviluppati con il linguaggio di programmazione Matlab (© 1984-2022 MathWorks, Inc) e con Excel (© 1985-2022 Microsoft Corporation).

2.1 Il Dataset

Il dataset utilizzato nel presente lavoro di tesi è composto dai dati di 175 pazienti affetti da tumore della prostata e sono stati forniti dal dipartimento di Oncologia dell'ospedale Molinette di Torino, in seguito ad un processo di anonimizzazione, al fine di garantire la privacy dei proprietari.

Nel corso del seguente lavoro di tesi il dataset è suddiviso in due sottoinsiemi:

- Training set: formato da 140 pazienti, usato per permettere alla rete neurale di imparare;
- Test set: formato dai restanti 35 pazienti, per fare delle valutazioni sulle reti allenate.

2.2 RayStation

RayStation è un TPS sviluppato da RaySearch Laboratories [8] e distribuito in Italia da Tecnologie Avanzate [9]. Il software permette di combinare le funzionalità del machine learning e del deep learning per elaborare in modo efficiente piani di trattamento VMAT, IMRT o 3D-CRT specifici per ogni paziente. Tra le varie funzionalità è implementata l'ottimizzazione multicriterio (MCO), che permette di generare un insieme di piani ottimali che delimitano lo spazio delle soluzioni, cioè che rispettano i vincoli imposti dall'utente. I medici, poi, possono navigare interattivamente tra questi piani per selezionare quello con il miglior compromesso clinico per assicurare l'adeguata copertura del tumore ed evitare l'eccessiva irradiazione delle strutture sane.

2.3 Reti neurali

Le reti neurali artificiali (ANN: Artificial Neural Network) [11] sono modelli matematici che tentano di simulare il comportamento del cervello umano. Il cervello è un calcolatore prodigioso, in grado di interpretare le informazioni fornite dai sensi con una velocità elevata e di imparare ad elaborarle autonomamente. Già dalla seconda metà dello scorso secolo gli studiosi tentano di riprodurre in maniera artificiale tale comportamento progettando delle reti che simulassero le caratteristiche di base dei neuroni e le loro interconnessioni.

I neuroni sono costituiti da tre elementi: dendriti, sinapsi e assoni. Un neurone riceve i segnali prodotti dalle altre cellule attraverso i dendriti e si propagano sotto forma di impulsi elettrici attraverso l'assone, che termina in tante ramificazioni. Nella parte terminali di ogni diramazione è presente una struttura, detta sinapsi, che converte l'attività proveniente dall'assone in impulsi eccitatori o inibitori per i neuroni a esso interconnessi.

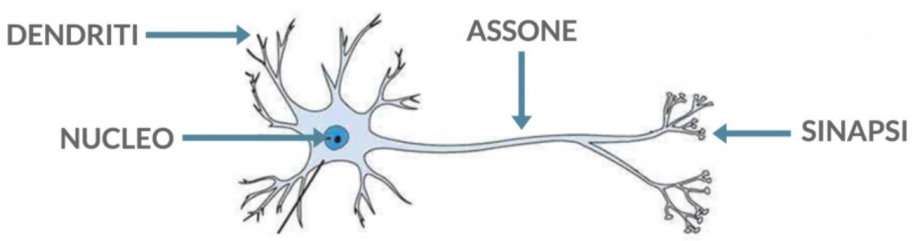


Figura 2.1: Struttura del neurone [12].

L'apprendimento avviene quando le sinapsi trasmettono i segnali da un neurone all'altro. Da qui deriva l'analogia con le reti neurali: sono composte da unità interconnesse, che simulano i neuroni, e da pesi modificabili associati a ogni connessione, che traslano la funzione delle sinapsi. Ciascuna unità moltiplica il segnale in ingresso per il peso della connessione e poi somma tutti i risultati, considerando eventualmente un valore detto scostamento. A questo risultato viene applicata una funzione di attivazione che trasforma il valore in un segnale in uscita, che viene poi trasmesso alle altre unità. L'addestramento avviene variando i valori associati ai pesi e agli scostamenti, in questo modo ogni neurone si specializza a riconoscere determinati stimoli e ad apprenderli.

Una generica rete neurale è composta da uno strato di input e da uno strato di output, mentre il numero di strati interni, detti strati nascosti, dipendono dal tipo di rete implementata. Quando l'informazione procede solo nella direzione che va dal layer di input al layer di output senza cicli, la rete prende il nome di rete feed-forward; se, invece, contiene dei cicli, si parlerebbe di reti neurali ricorrenti (RNN: Recurrent Neural Network), che si avvicinano maggiormente al comportamento umano ma sono meno utilizzate in termini di applicazioni e computazione.

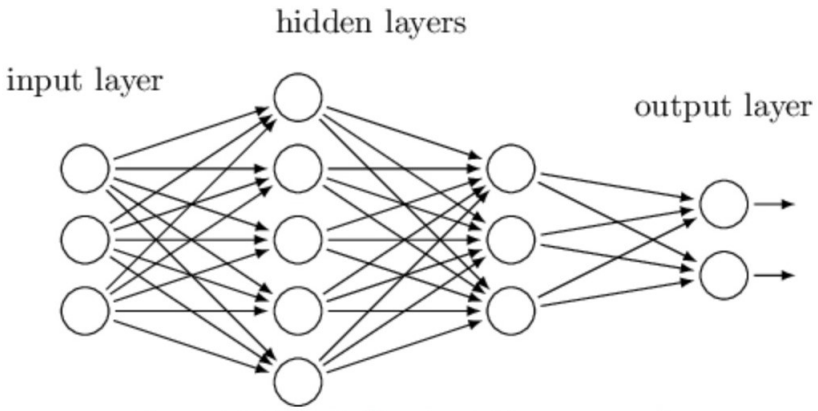


Figura 2.2: Struttura a livelli di una rete neurale [13].

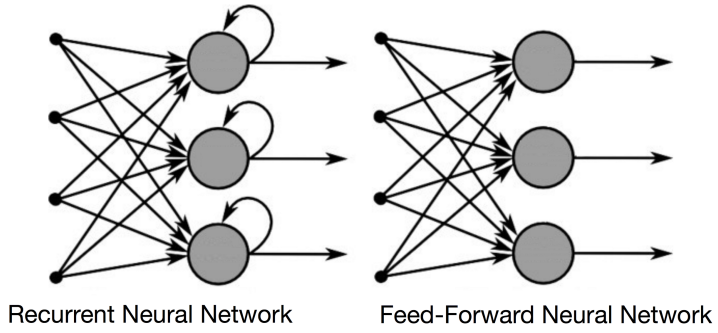


Figura 2.3: RNN e Feed Forward [14].

2.3.1 Misure di similarità

Alla base dell'allenamento delle reti neurali vi sono le misure di similarità. Il modo più comune per stimare una misura di similarità è quello che fa riferimento alla misura di una distanza. Le metriche più utilizzate per trattare dataset numerici sono:

- Distanza euclidea:

$$d(x, y) = \sqrt{\sum_{j=1}^{d}(x_j - y_j)^2} \tag{2.1}$$

- Distanza di Minkowsky:

$$d(x, y) = (\sum_{j=1}^{d}(x_j - y_j)^p)^{\frac{1}{p}} \tag{2.2}$$

- Distanza di Manhattan (o norma L_1):

$$d(x,y) = \sqrt{\sum_{j=1}^{d} |x_j - y_j|} \tag{2.3}$$

- Distanza massima (o norma L_∞):

$$d(x,y) = \max_{1 \leq j \leq d} |(x_j - y_j)| \tag{2.4}$$

- Distanza di Mahalanobis:

$$d(x,y) = \sqrt{(x-y)\sum (x-y)^T} \tag{2.5}$$

Le misure di similarità possono essere influenzate dai diversi range che le diverse variabili numeriche possono assumere, per cui è opportuno normalizzare i dati. Le principali tipologie di normalizzazione sono:

- Normalizzazione min-max: i valori saranno compresi nell'intervallo [0,1].

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{2.6}$$

  in cui x è il vettore di valori assunti dalla variabile mentre $min(x)$ e $max(x)$ sono rispettivamente il minimo e il massimo valore che essa assume.

- Normalizzazione media: i valori vengono riscalati nell'intervallo [-1,1].

$$x' = \frac{x - \mu}{max(x) - min(x)} \tag{2.7}$$

  in cui μ è la media dei valori della variabile.

- Standardizzazione:

$$x' = \frac{x - \mu}{\sigma} \tag{2.8}$$

  dove σ è la sua deviazione standard.

2.3.2 Rete neurale a base radiale

Una rete neurale a base radiale (RBFNN: Radial Basis Function Neural Network) è una ANN di tipo feed forward, per migliorare l'accuratezza e ridurre i tempi di allenamento grazie alla teoria dell'interpolazione. Questa tipologia di rete è comunemente usata per risolvere problemi di approssimazione di curve in uno spazio ad alta dimensione. Tale funzionamento è giustificato dal teorema di Cover [15]: "Un complesso problema di classificazione dei modelli, proiettato in uno spazio ad alta dimensione in modo non lineare, è più probabile che sia separabile linearmente che in uno spazio a bassa dimensione, a condizione che lo spazio non sia densamente popolato.

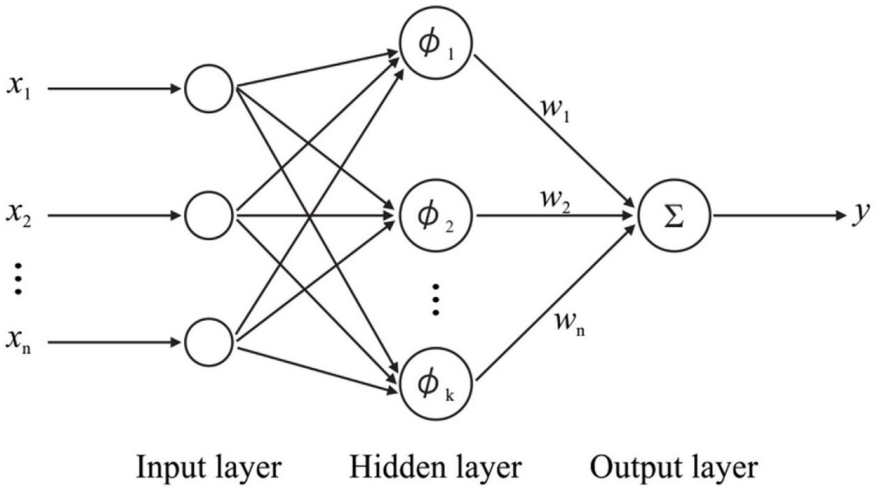


Figura 2.4: Architettura di una RBFNN [16].

2.3.2.1 Architettura della rete

Una rete neurale a base radiale è caratterizzata da un singolo layer nascosto. I collegamenti tra i neuroni di input e i neuroni del layer nascosto sono rappresentati da archi non pesati, caratterizzati da una trasformazione non lineare mentre i collegamenti tra i neuroni del layer nascosto e i neuroni di output sono archi pesati (w) che implementano una trasformazione lineare.

Ciascun neurone del layer nascosto, a differenza di altre reti neurali, non implementa una funzione di attivazione bensì una funzione a base radiale, ossia una funzione a valori reali il cui valore dipende unicamente dalla distanza fra il suo argomento e un punto fisso del dominio, detto centro (μ), che rappresenta il suo unico massimo. Alcune RBF sono caratterizzate anche dalla larghezza (σ) del campo ricettivo nello spazio di input.

Nel corso degli anni sono state proposte diverse RBF:

- Lineare:
$$\Phi(\|z_p - \mu_j\|_2) = \|z_p - \mu_j\|_2 \tag{2.9}$$

- Cubica:
$$\Phi(\|z_p - \mu_j\|_2) = \|z_p - \mu_j\|_2^3 \tag{2.10}$$

- Thin-plate-splane:
$$\Phi(\|z_p - \mu_j\|_2) = \|z_p - \mu_j\|_2^3 ln\|z_p - \mu_j\|_2 \tag{2.11}$$

- Multiquadratica:
$$\Phi(\|z_p - \mu_j\|_2, \sigma_j) = \sqrt{\|z_p - \mu_j\|_2^2 + \sigma_j^2} \tag{2.12}$$

- Multiquadratica inversa:

$$\Phi(\|z_p - \mu_j\|_2, \sigma_j) = \frac{1}{\sqrt{\|z_p - \mu_j\|_2^2 + \sigma_j^2}} \tag{2.13}$$

- Gaussiana:

$$\Phi(\|z_p - \mu_j\|_2, \sigma_j) = e^{-}\|z_p - \mu_j\|_2^2 / sigma_j^2 \tag{2.14}$$

- Logistica:

$$\Phi(\|z_p - \mu_j\|_2, \sigma_j) = \frac{1}{1 + e^{-}\|z_p - \mu_j\|_2^2 / 2\sigma_j^2 - \theta_j} \tag{2.15}$$

in cui $\| \bullet \|$ rappresenta la norma euclidea e θ_j rappresenta un valore di bias.

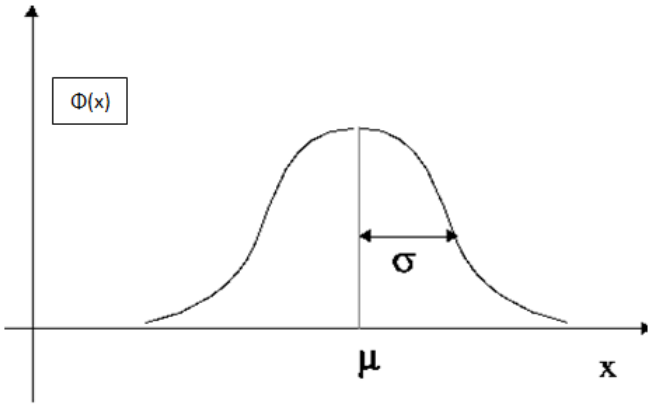


Figura 2.5: Esempio di RBF gaussiana [17].

2.3.2.2 Algoritmi di apprendimento

Per allenare le RBFNN sono stati sviluppati diversi algoritmi che si differenziano principalmente per il numero di parametri che vengono allenati. I principali algoritmi sono il metodo dei centri fissi, il metodo del gradiente discendente e il metodo a due fasi.

Il metodo dei centri fissi è stato proposto da Broomhead e Lowe [18]. Il primo passo consiste nella scelta dei centri, che coincidono con i neuroni nel layer nascosto, tra gli elementi del training set. Successivamente si calcola la larghezza del campo ricettivo di ogni RBF come in formula 2.16:

$$\sigma_j = \sigma = \frac{d_{max}}{\sqrt{j}}, \qquad j = 1, ..., J \tag{2.16}$$

in cui d_{max} è la massima distanza euclidea tra i centri e J è il numero di centri scelto.

Si calcolano i pesi della rete tramite la formula 2.17:

$$\mathbf{w}_k = (\phi^T \phi)^{-1} \phi^T t_k \tag{2.17}$$

in cui $k = 1, ..., K$ è il numero che individua iterativamente l'input della rete, $\mathbf{t}_k$ è il vettore degli output target, $\phi \in R^{KxJ}$ è la matrice con le mappature non lineari RBF prodotte dal layer nascosto.

Infine, si calcola l'output per ogni elemento di input come in formula 2.18:

$$o_{k,p} = \sum_{j=1}^{J} w_{kj} yj, p \tag{2.18}$$

$$y_{j,p}(\mathbf{z}_p) = \phi(\|\mathbf{z}_p - \mu_j\|_2, \sigma_j) = e^{-\frac{\|\mathbf{z}_p - \mu_j\|_2^2}{2\sigma_j^2}} \tag{2.19}$$

in cui $\mathbf{z}_p$ è il vettore di input e μ_j rappresenta il centro della RBF.

L'allenamento con il gradiente discendente è stato proposto da Moody e Darken [19]. e da Poggio e Girosi [20]. In questo algoritmo, oltre a scegliere i centri e le larghezze è necessario inizializzare i pesi della rete con piccoli valori casuali. Successivamente si modificano iterativamente i parametri sulla base dell'errore tra l'output desiderato e l'output della rete, finché non si raggiunge una condizione di stop scelta dall'utente.

$$w(t+1) = w(t) - \eta_w \frac{\partial E}{\partial w} \tag{2.20}$$

$$\mu(t+1) = \mu(t) - \eta_c \frac{\partial E}{\partial \mu} \tag{2.21}$$

$$\sigma(t+1) = \sigma(t) - \eta_\sigma \frac{\partial \sigma}{\partial w} \tag{2.22}$$

Il terzo metodo prevede due fasi. La prima è caratterizzata da un apprendimento non supervisionato con il metodo del k-means, un algoritmo di clustering partizionale, per scegliere i centri. Nella seconda fase si effettua un allenamento supervisionato dei pesi seguendo il metodo del gradiente discendente.

2.4 Clustering

Quando si parla di clustering di dati si fa riferimento a un insieme di tecniche che hanno in comune l'obiettivo di identificare, tra gli elementi di un dataset, dei gruppi naturali, detti cluster, che siano omogenei al loro interno e diversi tra di loro. La prima caratteristica si misura con la variabilità intra-cluster, definita come la media delle distanze tra tutte le coppie di elementi all'interno del cluster. La seconda si misura tramite la distanza inter-cluster, che si ottiene calcolando le distanze tra tutte le coppie di prototipi dei gruppi.

La maggior parte degli algoritmi di clustering possono essere suddivise in due tipologie: clustering partizionale e clustering gerarchico [10]. Nella prima i dati vengono divisi in un numero specificato di cluster attraverso la minimizzazione di un certo criterio. Nella seconda i dati sono divisi secondo una gerarchia di partizioni annidate, caratterizzata da una misura di similarità e da una regola di classificazione.

2.4.1 Metodo di clustering

Il metodo di clustering è ciò che permette di assegnare un elemento a un certo cluster. Tra i metodi più comuni per raggruppare gli elementi vi sono:

- Legame singolo (single link): riunisce i cluster che presentano la minima distanza definita come la minima tra tutte le distanze calcolabili a due a due tra gli elementi dei gruppi. Questo metodo predilige la differenza tra i cluster.

- Legame completo (complete link): riunisce i cluster che presentano la più piccola distanza definita come la massima tra tutte le distanze calcolabili a due a due tra gli elementi dei gruppi. Questo metodo dà maggiore importanza all'omogeneità degli elementi all'interno del cluster.

- Centroide: riunisce i cluster che hanno la minore distanza tra i relativi centroidi. Lo svantaggio di questo metodo è che un cluster molto piccolo viene sopraffatto quando è raggruppato con un cluster più grande.

2.4.2 Dendrogramma

Il dendrogramma è un grafo per la visualizzazione del processo di clustering gerarchico. Sull'asse orizzontale sono distribuiti gli elementi del dataset, mentre l'asse verticale è orientato e riporta la distanza alla quale i cluster sono uniti.

Il processo di costruzione del dendrogramma comincia considerando ogni elemento del dataset come un cluster. Successivamente, vengono calcolate le distanze tra tutti gli elementi e vengono riuniti in un cluster i due elementi che rispettano il metodo di clustering scelto. Si procede iterativamente finché non si giunge all'apice, cioè quella condizione in cui tutti gli elementi sono raggruppati in un singolo cluster. Infine, si effettua il taglio del dendrogramma ad un valore scelto dall'utente per visualizzare il numero di raggruppamenti desiderato.

2.5 K-fold

Il K-fold, anche detto convalida incrociata, è un metodo di ricampionamento dei dati per stimare l'abilità di un modello predittivo di generalizzare [21]. Al tempo stesso questo metodo permette di evitare l'overfitting, cioè quel fenomeno che si verifica quando il modello si adatta a caratteristiche specifiche del training set ma che non hanno riscontro su altri dati e si caratterizza da un'incapacità nel generalizzare.

Nel k-fold, il set di dati viene partizionato in k gruppi disgiunti approssimativamente delle stesse dimensioni. In seguito, il modello viene allenato usando un training set formato da k-1 gruppi e applicato sul validation set, rappresentato dal restante gruppo, per misurarne le performance. Questo procedimento è ripetuto finché tutti i gruppi vengono utilizzati come validation set.

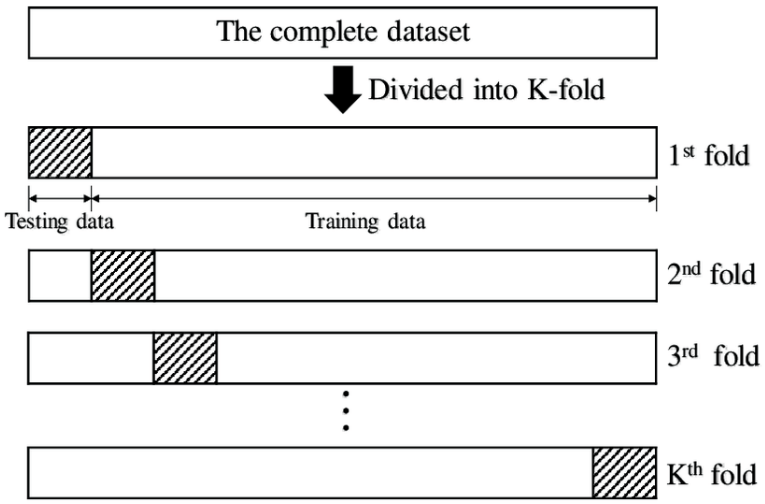


Figura 2.6: Rappresentazione grafica della divisione del dataset effettuata con il k-fold [22].

2.6 Algoritmi Genetici

Gli algoritmi genetici (GA: Genetic Algorithm) fanno parte della computazione evolutiva (EC: Evolutionary Computation), cioè un insieme di metodi euristici di ricerca dell'ottimo che prendono spunto dalla teoria evolutiva di Charles Darwin. L'idea di base, infatti, è quella di partire da una popolazione iniziale di soluzioni di un problema e farle evolvere, usando degli operatori ispirati alle mutazioni genetiche e alla selezione naturale, con un processo iterativo che simula l'evoluzione delle soluzioni verso quella migliore. Dal momento che si tratta di un algoritmo stocastico non è garantito che si ottenga sempre la soluzione migliore ma, in genere, si trovano soluzioni sufficientemente buone.

La bontà delle soluzioni è quantificata attraverso la funzione di fitness. Tipicamente si tratta di una funzione matematica ma può essere un modello di simulazione, una funzione logica o altro. Di conseguenza, l'obiettivo del GA è quello di ricercare la soluzione che minimizza la fitness.

2.6.1 Codifica delle variabili

La scelta della codifica appropriata delle variabili è un passo fondamentale per l'utilizzo di un GA, in quanto una specifica codifica potrebbe ridurre i costi computazionali, ad esempio garantendo automaticamente l'ammissibilità di tutte le soluzioni. Nei GA le variabili sono rappresentate sotto forma di sequenze di bit con una dimensione fissata, che possono assumere valori binari. All'interno di una soluzione possono essere presenti più variabili, non necessariamente con la stessa codifica.

I principali metodi di codifica delle variabili sono:

- Variabili SI/NO: ogni bit della soluzione rappresenta una variabile e il suo valore identifica la presenza o meno della relativa variabile nella soluzione finale;
- Variabili intere: le variabili possono assumere un certo range di valori e vengono codificate con la conversione in binario del loro valore decimale;
- Variabili codificate: le variabili assumono valori non numerici, per cui a ciascuna di esse viene assegnato un codice numerico.

2.6.2 Operatori evoluzionari

Esistono due tipologie di operatori evoluzionari: quelli di riproduzione e quelli di selezione.

Gli operatori di riproduzione si dividono in due tipologie:

- Operatori di crossover: tramite la sua applicazione due soluzioni si scambiano uno o più bit per produrre nuove soluzioni. A essi è associata la probabilità di crossover (tipicamente Pc = [0,9 1]), che permette di applicare l'operatore solo quando un numero estratto casualmente è superiore a tale valore.
- Operatori di mutazione: in seguito alla sua applicazione uno o più bit all'interno della soluzione cambiano il loro valore, da 0 a 1 e viceversa. A essi è associata la probabilità di mutazione (tipicamente Pm = [0,05 0,2]), che funziona come la probabilità di crossover.

Gli operatori di selezione vengono applicati durante il processo di selezione delle soluzioni a cui applicare gli operatori di riproduzione a partire dalla popolazione iniziale e durante il processo di generazione della nuova popolazione al termine di ogni iterazione. A questi operatori è associata la selective pressure, un parametro legato al tempo che occorre per giungere alla costruzione di una popolazione uniforme formata da soluzioni uguali.

2.6.3 Algoritmo generico

L'algoritmo genetico si basa sull'iterazione dei seguenti step:

1. Generazione, solitamente casuale, della popolazione iniziale di soluzioni;
2. Attribuzione del valore della funzione fitness a ogni soluzione della popolazione per valutarne la bontà;
3. Selezione delle soluzioni su cui applicare gli operatori di riproduzione;
4. Generazione della nuova popolazione a partire dalle nuove soluzioni.

Il ciclo iterativo termina al raggiungimento della condizione di stop, ad esempio quando è stato raggiunto il numero massimo di iterazioni o quando non si verificano cambiamenti nelle soluzioni per un certo numero di iterazioni consecutive.

2.7 Analisi statistica

2.7.1 diagramma a scatola e baffi

Il diagramma a scatola e baffi (o boxplot) è uno strumento della statistica descrittiva proposto da Tukey, utile per riportare graficamente una distribuzione di dati. Esso consiste di un rettangolo, detto scatola, delimitato dal primo e dal terzo quartile della distribuzione e diviso in due parti dalla mediana. Dalla scatola si diramano due segmenti, detti baffi, che si estendono fino ai dati più estremi ma che non siano più distanti di 1,5 volte la differenza tra i due quartili precedenti. I valori al di fuori dei baffi sono rappresentati sotto forma di asterischi e sono chiamati outliers.

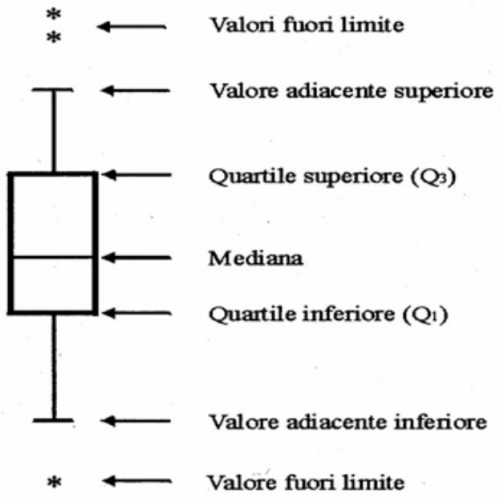


Figura 2.7: Esempio di diagramma a scatola e baffi

2.7.2 Grafico a barre

L'istogramma (o bar graph) è una rappresentazione grafica che permette di confrontare diverse categorie di dati discreti. È composto da rettangoli separati, le cui basi sono allineate su un asse non orientato, che rappresentano le diverse variabili e la cui altezza è proporzionale alla frequenza assoluta dei dati.

2.7.3 Grafico a dispersione

Il grafico a dispersione (o scatter plot) è un metodo della statistica descrittiva che permette di visualizzare la relazione tra due variabili quantitative. I dati sono visualizzati in un piano cartesiano, in cui in ciascuno dei due assi sono riportati i valori di una delle due variabili. Ogni elemento è rappresentato da un punto e la relativa posizione è determinata dai valori che assume.

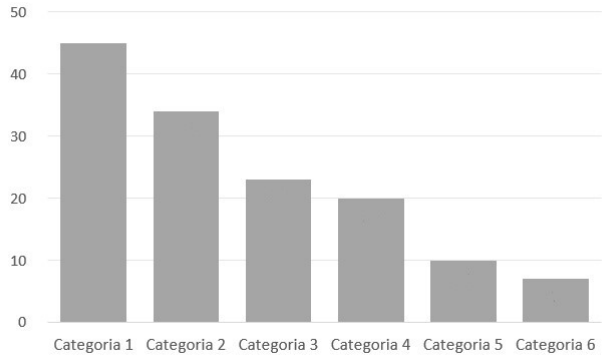


Figura 2.8: Esempio di grafico a barre

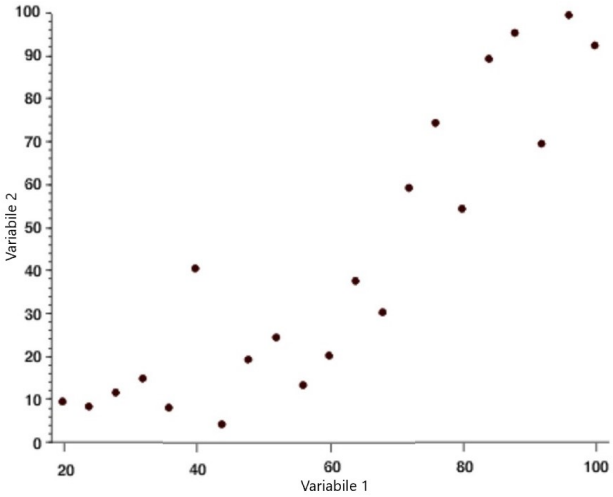


Figura 2.9: Esempio di grafico a dispersione

2.7.4 Silhouette plot

Il silhouette plot si colloca nell'ambito del clustering dei dati e permette di rappresentare graficamente la qualità del partizionamento dei vari elementi. A ogni elemento viene associato un coefficiente di silhouette, valore che può variare tra -1 e 1. In particolare un valore positivo indica che l'elemento è molto simile agli elementi del cluster in cui è stato incluso e viceversa.

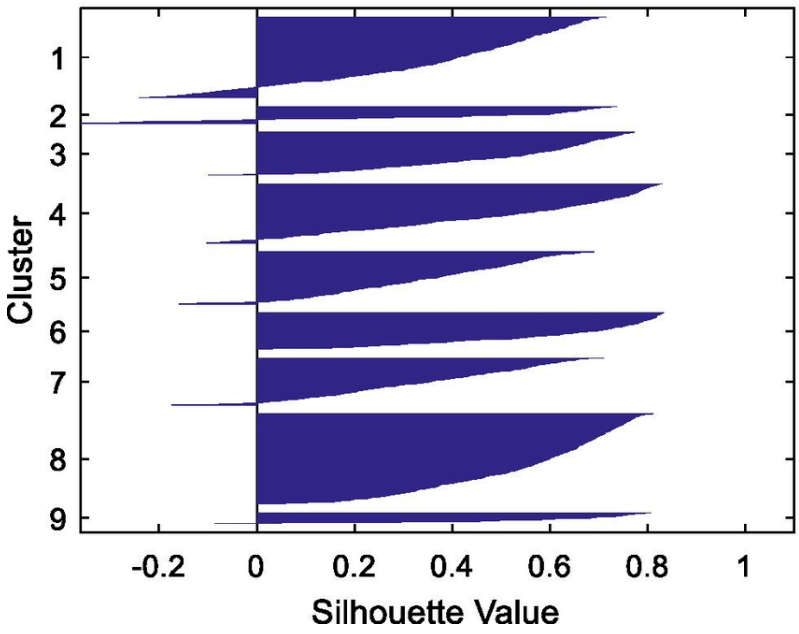


Figura 2.10: Esempio di silhouette plot [23]

2.7.5 Test di Mann-Whitney

Il test di Mann-Whitney, anche noto come test U di Mann-Whitney è un test non parametrico utilizzato per verificare se due campioni statistici indipendenti provengono dalla stessa popolazione. Alla base del test vi è la formulazione dell'ipotesi nulla, secondo cui non c'è differenza tra i due campioni in relazione al parametro in esame e l'eventuale differenza deriva da fluttuazioni casuali dovute all'estrazione del campione. Contemporaneamente si formula l'ipotesi alternativa, complementare alla precedente, e il parametro α, detto livello di significatività, che rappresenta la probabilità di respingere l'ipotesi nulla quando è vera. Si calcola, quindi, una misura di evidenza contro l'ipotesi nulla, detto p-value. Se tale valore è maggiore di α si può concludere che non ci sono evidenze statistiche per rifiutare l'ipotesi nulla e viceversa se è minore di α.

Capitolo 3

Ottimizzazione dei piani di radioterapia

3.1 Radial Basis Function Neural Network

Lo scopo della rete neurale a base radiale è quello di apprendere la relazione tra la coppia di dose fornita agli OAR e il relativo valore di fitness, che ne valuta la bontà. In questo modo la rete allenata sarà in grado di associare a ogni nuovo piano di input il relativo valore di fitness. Per questo motivo la rete è stata integrata all'interno della funzione di fitness dell'algoritmo genetico, così da eliminare i tempi di computazione introdotti dall'elaborazione dei piani da parte di RayStation.

3.1.1 Estrazione dei dati tramite RayStation

I dati con cui sono state allenate le reti sono stati generati tramite il TPS RayStation. Per ogni paziente sono stati creati 36 piani di ratioterapia, ciascuno formato da una delle possibili combinazioni di valori MaxEUD da somministrare alla vescica e al retto considerando la dose da 500 cGy e quelle comprese tra 1000 e 5000 cGy a passi di 1000 cGy. I piani sono stati realizzati vagliando la prescrizione clinica di 2,7 Gy suddivisi in 26 frazioni e utilizzando la tecnica VMAT e l'acceleratore lineare LIN 4.

Per ciascuno di questi piani il software ottimizza la dose al PTV e calcola un valore di fitness, in modo che il piano migliore corrisponda al valore minore. La formula utilizzata dallo script di RayStation per restituire il valore della funzione di fitness è la 3.1 [24].

$$FF = \sqrt{\frac{2(CI)^2 + 4(\frac{rAD}{PD})^2 + 2(\frac{bAD}{PD})^2 + (\frac{lfAD}{PD}^2 + (\frac{rfAD}{PD})^2}{10}} \tag{3.1}$$

in cui CI è l'indice di conformità (copertura del PTV), PD è la dose prescritta al PTV e rAD, bAD, $lfAD$ e $rfAD$ sono rispettivamente le dosi fornite al retto, alla vescica e alle teste femorali sinistra e destra.

Nella tabella 3.1 è riportato un esempio dei valori in uscita dal software.

Valore fitness	Dose vescica	Dose retto
3.166640	4000	500
3.194783	3000	500
3.238791	1000	1000
3.276877	5000	500
3.383622	1000	500
3.410646	5000	1000
3.427088	4000	1000
3.467184	3000	1000
3.474875	2000	1000
3.585049	500	500
3.674493	2000	500
3.886925	500	1000
4.420693	1000	2000
4.517180	2000	2000
4.614820	3000	2000
4.614992	4000	2000
4.677751	5000	2000
4.771912	500	2000
5.624539	500	3000
5.889451	5000	3000
5.911628	3000	3000
5.946976	4000	3000
5.961825	2000	3000
5.970642	1000	3000
7.367275	500	4000
7.373207	4000	4000
7.406356	2000	4000
7.426974	5000	4000
7.458891	3000	4000
7.641830	2000	5000
7.790947	1000	4000
7.927229	5000	5000
7.927276	4000	5000
7.976940	3000	5000
8.708664	500	5000
8.712591	1000	5000

Tabella 3.1: Dati in uscita da RayStation per il paziente 40

3.1.2 Allenamento della rete

Il training della RBFNN è individuale per ogni singolo paziente ed è stato effettuato con il metodo dei centri fissi. Le principali difficoltà di questo metodo risiedono nella scelta del numero adeguato di centri e nella loro selezione, in modo che siano distribuiti equamente su tutto lo spazio di input. Un approccio in tale direzione si basa sull'uso di tecniche di clusterizzazione, in modo da trovare un insieme di centri che riflettano in maniera accurata la distribuzione dei dati. Nel presente lavoro di tesi è stato utilizzato il metodo del dendrogramma sui valori di fitness ottenuto per ogni coppia di dose.

Per l'allenamento della rete, i valori di dose al retto e alla vescica costituiscono gli input della rete, mentre il target è rappresentato dal valore di fitness prodotto da RayStation.

3.1.3 Training set e validation set invariati

3.1.3.1 Definizione dei set

In una prima fase, le coppie di dose per ogni paziente sono state suddivise in due gruppi fissi:

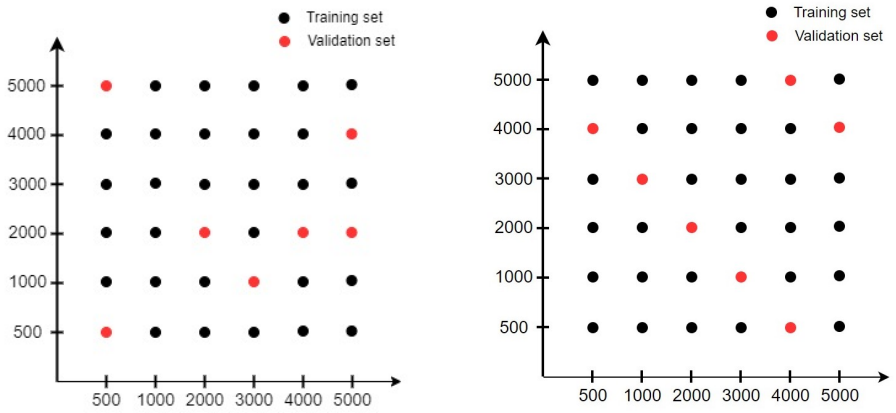
- Training set: contiene l'80% dei dati, cioè 29 coppie, utili per l'allenamento della rete;
- Validation set: contiene il restante 20% dei dati, cioè 7 coppie, per valutare le prestazioni della rete allenata.

Nei precedenti lavori di tesi [25][26][27], la suddivisione è stata effettuata ordinando le coppie sulla base della fitness e inserendo un punto ogni cinque nel validation set. Rappresentando i dati sotto forma di punti in un piano cartesiano [fig. 13] in cui gli assi contengono rispettivamente le dosi al retto e alla vescica, si evince che tale divisione non consente di avere una distribuzione equa dei punti nello spazio. Al fine di avere dei campioni equispaziati uguali per tutti i pazienti, le coppie sono state ordinate sulla base dei valori di dose e di selezionare sette punti così come effettuato precedentemente.

3.1.3.2 Analisi del numero di centri

Per capire l'influenza che il numero di centri selezionato ha sulle prestazioni della rete si è proceduto nel seguente modo per ciascun paziente:

1. È stato realizzato il dendrogramma sulle coppie del training set dei punti;
2. Il dendrogramma è stato tagliato più volte in modo da avere un totale di 28 clusterizzazioni diverse: da 2 fino a 29 cluster;
3. Per ciascuna clusterizzazione sono stati calcolati i centri dei cluster come mediane degli elementi al loro interno, per cui il numero di centri utilizzato corrisponde al numero di cluster individuati;
4. Per ogni clusterizzazione è stata allenata una RBFNN con i punti appartenenti al training set e valutata in termini di errore quadratico medio (MSE: Mean Squared Error) tra la predizione e il target dei punti del validation set.



(a) Suddivisione secondo il metodo precedente - Paziente 96

(b) Suddivisione secondo il nuovo metodo

Figura 3.1: Rappresentazione delle coppie di punti in uno spazio cartesiano

I risultati ottenuti da una prima analisi su 10 pazienti selezionati casualmente dal training set (pazienti: 19, 25, 73, 163, 180, 218, 234, 236, 237 e 253) sono stati riportati in un grafico a dispersione (fig. 3.2). Da questi si evince che le reti allenate con un numero di centri particolarmente basso rispetto al numero di elementi del training set dei punti hanno i risultati peggiori. Estendendo l'analisi a tutti i 140 pazienti del training set si

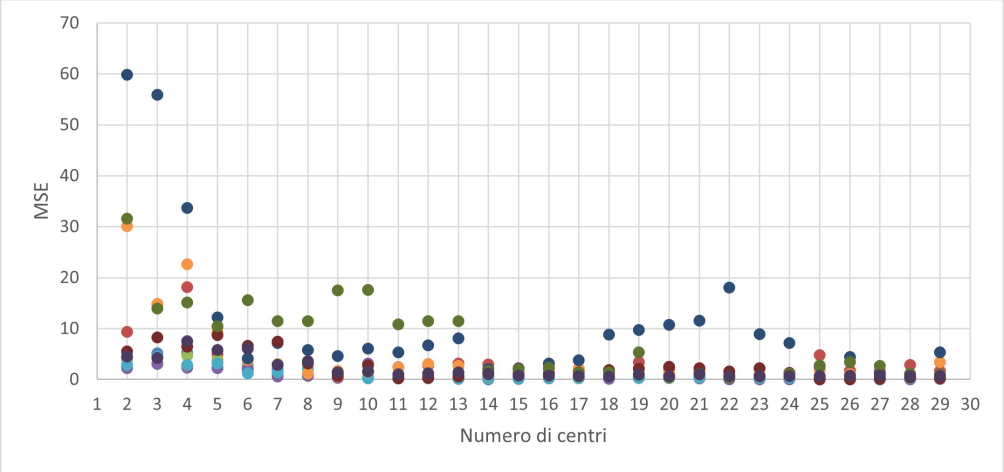


Figura 3.2: Scatter plot dei valori MSE per 10 pazienti del training set

è proceduto come nel caso precedente, ma sono state omesse le clusterizzazioni da 2 a 5 cluster.

I valori MSE ottenuti per ciascun paziente sono riportati in appendice A. Anche in questo caso i risultati sono stati riportati in un grafico a dispersione (fig. 3.3), da cui si osserva che la combinazione migliore è quella che prevede l'allenamento con 15 centri.

Per confermare l'effettiva validità delle reti allenate con 15 centri sono stati confrontati i

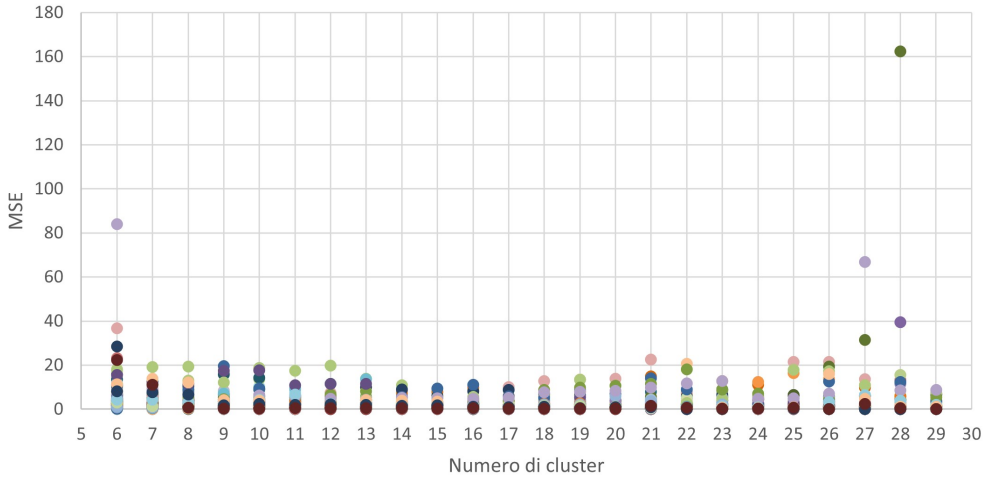
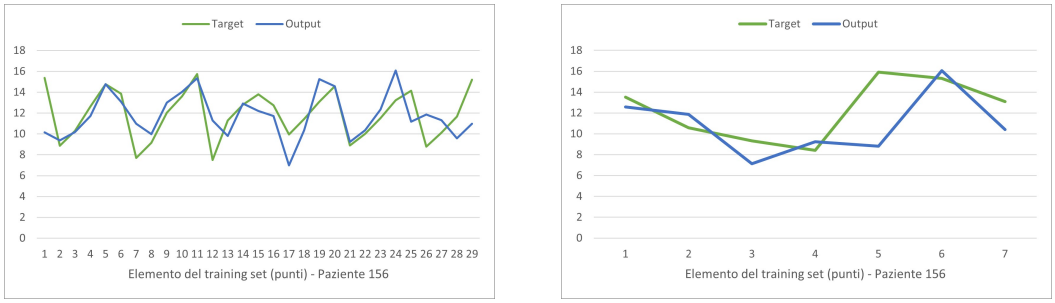


Figura 3.3: Scatter plot dei valori MSE per i 140 pazienti del training set

valori target e gli output restituiti su tutte le coppie di dose per sei pazienti:

- I due pazienti con MSE sui punti del validation set minore: pazienti 53 (MSE: 0,03) e 131 (MSE: 0,04);
- I due pazienti con MSE sui punti del validation set intermedio: pazienti 6 (MSE: 0,71) e 62 (MSE: 0,72);
- I due pazienti con MSE sui punti del validation set maggiore: pazienti 95 (MSE: 7,93) e 156 (MSE: 9,46).



(a) Training set (b) Validation set

Figura 3.4: Confronto target e valore di output restituito - paziente 156

I risultati, di cui la figura 3.4 è un esempio, sono stati allegati nell'appendice B. Da essi emerge che alcune coppie dei pazienti con MSE più alto presentano errori particolarmente

elevati, per cui si è proseguito con l'analisi per cercare una rete che funzionasse meglio. A tal fine sono state selezionate la rete con le prestazioni migliori per ogni paziente del training set ed è stato creato un istogramma delle frequenze per valutare, tra queste, il numero di centri più ricorrente.

Dal grafico in figura 3.5 si evince che la rete più ricorrente tra le migliori sia quella con 29 centri, cioè quella per cui a ogni coppia corrisponde un centro, con un totale di 25 ricorrenze.

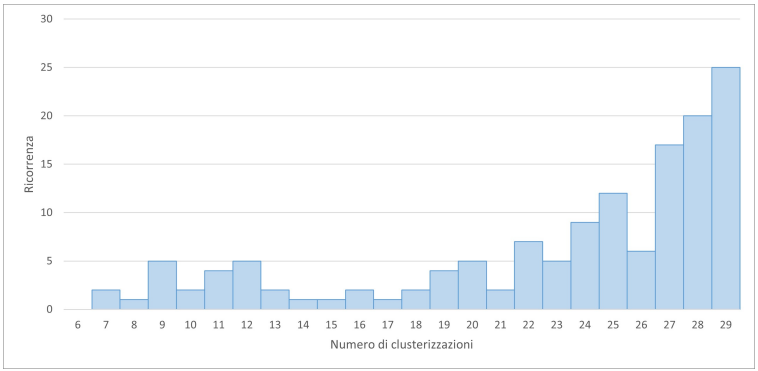


Figura 3.5: Istogramma della frequenza del numero di centri migliore

3.1.4 K-fold

3.1.4.1 Definizione dei set

Per rendere più robusta l'analisi della rete migliore è stato effettuato un nuovo studio sulla scelta del numero migliore di centri con il metodo del k-fold. In questo caso le coppie di dose per ogni paziente sono state suddivise in sei gruppi secondo lo schema riportato in figura 3.6. Dopo di che sono stati usati iterativamente:

- L'unione di cinque sottogruppi come training set: 30 punti;
- Il restante sottogruppo come validation set: 6 punti.

3.1.4.2 Analisi del numero di centri

Per ciascuna delle 6 combinazioni sono state allenate 29 RBFNN, per un totale di 174 reti per paziente, procedendo come nel caso precedente ma con tutte le clusterizzazioni da 2 a 30 cluster.

È stato analizzato il numero di centri della rete migliore per ogni paziente e i risultati sono stati riportati nell'istogramma delle frequenze in figura 3.7. In questo caso la rete migliore più ricorrente risulta essere quella con 30 centri.

Successivamente, per ogni numero di centri si è visto quanti pazienti presentassero almeno una rete con valore MSE inferiore a uno. Dai risultati, riportati in figura 3.8, si

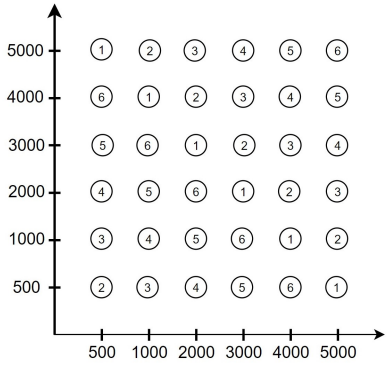


Figura 3.6: Suddivisione dei punti in sei gruppi

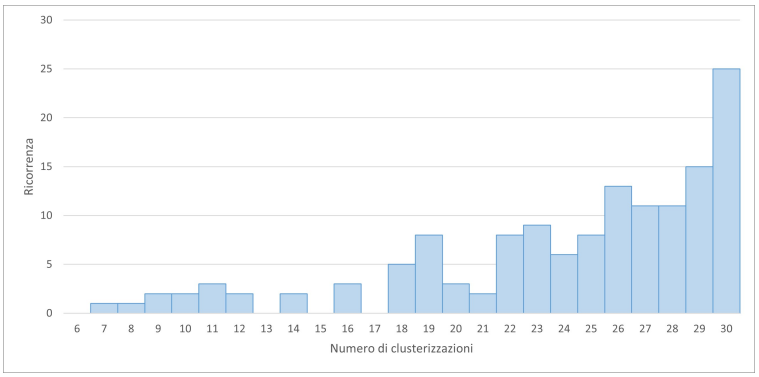


Figura 3.7: Istogramma della frequenza del numero di centri migliore

osserva che le reti migliori sono risultate quelle con 21 e con 24 centri, entrambe con 127 reti su un totale di 174 reti.

Infine, sono state confrontati, tramite due grafici boxplot, i valori MSE delle reti con 21, 24 e 30 centri considerando sia la distribuzione ottenuta con l'unione dei valori ottenuti dalle sei prove per ogni paziente (fig. 3.9) sia quella ottenuta considerando solamente la prova migliore per ciascun paziente (fig. 3.10). I valori minimi, medi e massimi per le tre reti sono riportati, rispettivamente per i due casi, nelle tabelle 3.2 e 3.3.

	21 centri	24 centri	30 centri
Minimo	0.002	0.001	0.001
Media	1.770	1.672	1.089
Massimo	25.841	45.053	25.924

Tabella 3.2: Valori minimi, medi e massimi della distribuzione dei valori MSE per le reti con 21, 24 e 30 centri considerando tutte le prove per ogni paziente

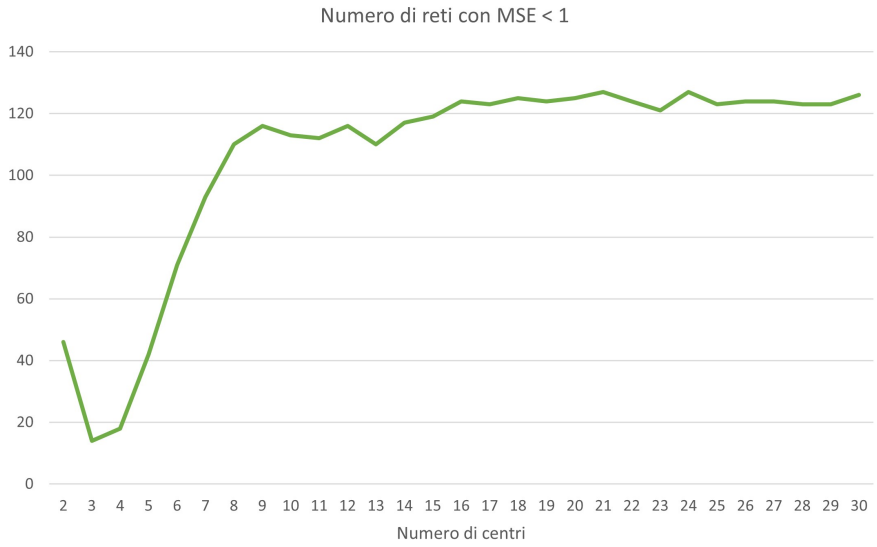


Figura 3.8: Numero di reti che hanno un valore MSE inferiore a 1 per ogni numero di centri

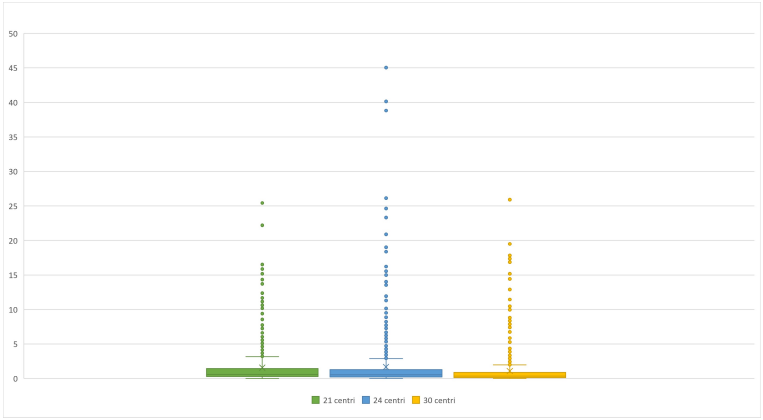


Figura 3.9: Boxplot delle distribuzioni dei valori MSE per le reti con 21, 24 e 30 centri considerando tutte le prove per ogni paziente

	21 centri	24 centri	30 centri
Minimo	0,002	0,001	0,001
Media	0,466	0,404	0,424
Massimo	5,619	3,347	5,997

Tabella 3.3: Valori minimi, medi e massimi della distribuzione dei valori MSE per le reti con 21, 24 e 30 centri considerando la prova migliore per ogni paziente

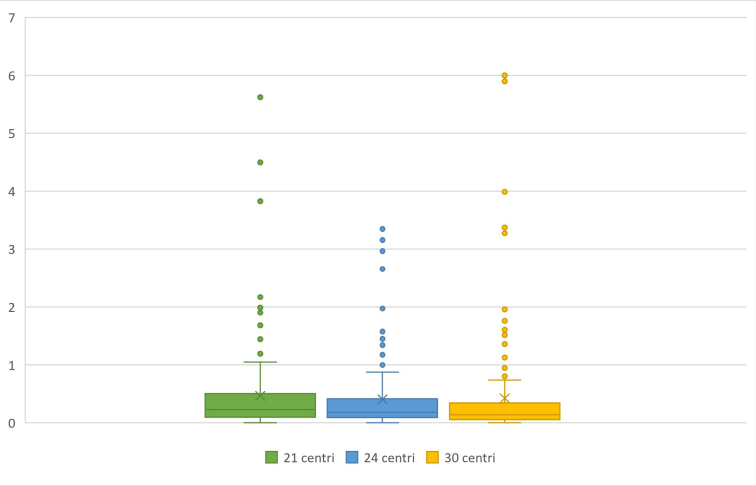


Figura 3.10: Boxplot delle distribuzioni dei valori MSE per le reti con 21, 24 e 30 centri considerando la prova migliore per ogni paziente

Dai risultati si evince che, per quanto riguarda la distribuzione che comprende tutte le prove, la reti con 21 e 30 centri presentano un valore MSE massimo nettamente inferiore rispetto a quello della rete con 24 centri e tra le due quella con 30 centri ha il valore medio inferiore. Tuttavia, nella distribuzione che comprende solo i valori delle prove migliori, cioè quelle che verranno utilizzate per ogni paziente, la rete con 24 centri è quella che presenta i valori più bassi per tutti e tre i parametri. Per questo motivo la scelta è ricaduta su quest'ultima.

3.1.5 Validazione del numero di centri

Per validare la scelta del numero di centri, per ognuno dei trentacinque pazienti del test set sono state allenate, con il metodo del k-fold, le sei reti con 24 centri. I risultati in termini di MSE sono riportati nella tabella (3.4).

I risultati mostrano che esiste almeno una combinazione di validation set e training set dei punti che riporta, per ciascun paziente del test set, un valore MSE sul validation set inferiore a 1.

Paziente	Val. set 1	Val. set 2	Val. set 3	Val. set 4	Val. set 5	Val. set 6
3	0,719	0,239	0,099	0,108	0,924	0,829
13	0,287	0,395	0,189	0,323	0,614	2,028
23	1,553	0,877	0,256	0,149	0,494	0,762
24	1,061	0,909	0,809	1,586	0,059	0,774
39	0,307	0,389	0,174	0,017	0,018	0,282
41	0,691	0,539	0,093	0,126	0,335	1,010

47	0,838	0,738	0,388	0,506	0,453	1,587
53	0,404	0,215	0,167	0,051	0,058	0,143
75	0,485	0,486	0,176	0,136	0,696	0,644
88	0,694	0,704	0,288	0,024	10,328	0,735
93	0,297	0,048	0,084	0,067	0,114	0,183
124	1,847	2,952	0,648	0,903	2,901	5,646
125	0,312	0,191	0,128	0,033	0,081	0,202
138	1,141	0,714	0,083	0,282	0,464	0,964
161	0,332	0,126	0,091	0,009	0,124	0,303
164	0,198	0,065	0,051	0,056	0,099	0,118
165	3,143	3,078	3,907	3,214	0,961	4,803
178	1,941	0,287	0,667	0,250	0,117	0,361
181	1,620	0,715	0,801	0,968	0,856	3,080
192	0,286	0,161	0,174	0,050	0,162	0,100
195	0,623	1,612	1,588	0,267	0,195	0,626
199	0,479	0,139	0,073	0,094	0,170	0,283
200	0,358	0,692	0,256	0,235	1,913	1,140
202	0,064	0,396	0,238	0,295	0,104	0,117
211	1,111	1,881	1,530	0,445	5,373	2,287
214	1,145	0,148	0,060	0,277	0,141	0,302
222	0,461	0,380	0,060	0,209	0,560	0,222
224	0,944	0,235	0,711	0,589	0,471	1,096
231	0,526	0,226	0,053	0,447	0,859	0,886
235	0,840	0,406	0,176	0,047	0,139	0,546
236	0,623	0,796	0,216	0,301	1,215	1,557
244	0,241	0,118	0,196	0,126	0,164	0,595
249	1,497	0,515	0,226	0,934	1,205	1,490
251	0,162	0,090	0,046	0,089	0,082	0,668
253	0,303	0,410	0,499	0,229	0,810	1,305

Tabella 3.4: Valori MSE per la rete con 24 centri sul validation set dei pazienti del test set

3.2 Algoritmo genetico

L'algoritmo genetico è stato implementato per ottimizzare il valore di dose da fornire al retto e alla vescica. Le possibili combinazioni di dose vengono valutate tramite una funzione di fitness che, per ogni paziente, implementa la RBFNN allenata con la combinazione di gruppi, derivanti dal k-fold, che ha riportati il minore valore MSE di predizione sul validation set. La soluzione migliore è quella con il valore di fitness minore.

3.2.1 Codifica delle variabili

Le soluzioni sono state rappresentate con la codifica codified variables. Ogni soluzione è composta da 16 bit divisi in due gruppi da 8 bit (fig. 3.11): il primo codifica la dose MaxEUD destinata alla vescica, il secondo codifica la dose MaxEUD da fornire al retto. Ogni variabile può assumere 255 valori, per cui l'algoritmo esplora in totale 65535 soluzioni, che combinano valori di dose che variano da 500 a 5000 cGy a passi di 20 cGy.

Per decodificare una variabile è necessario convertirne il valore binario in decimale, sommare 25 e moltiplicare per 20.

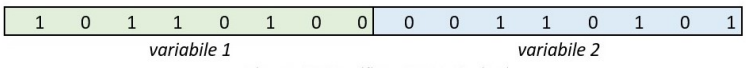


Figura 3.11: Esempio di soluzione

3.2.2 Parametri

I parametri del GA sono stati mantenuti invariati rispetto a quelli utilizzati negli studi precedenti, riportati nella tabella 3.5.

Per quanto riguarda gli operatori di riproduzione si è cercato di ottimizzarne i valori, valutando ventiquattro combinazioni, riportate nella tabella 3.6, ottenute facendo variare i parametri nel seguente modo:

- PC: da 0.8 a 0.95 a passi di 0.05;
- PM: da 0.1 a 0.2 a passi di 0.02.

Parametri	Valori
Numero di individui	100
Numero di iterazioni	200
Numero di genitori	80% della popolazione
Selezione genitori	Roulette wheel
Numero di geni	16
Selezione nuova generazione	Random selection
Mutazione	Random mutation
Crossover	Due punti di taglio
Numero di ripetizioni	50

Tabella 3.5: Parametri fissi del GA

Numero combinazione	PC	PM
1	0.8	0.1
2	0.8	0.12
3	0.8	0.14
4	0.8	0.16
5	0.8	0.18
6	0.8	0.2
7	0.85	0.1
8	0.85	0.12
9	0.85	0.14
10	0.85	0.16
11	0.85	0.18
12	0.85	0.2
13	0.9	0.1
14	0.9	0.12
15	0.9	0.14
16	0.9	0.16
17	0.9	0.18
18	0.9	0.2
19	0.95	0.1
20	0.95	0.12
21	0.95	0.14
22	0.95	0.16
23	0.95	0.18
24	0.95	0.2

Tabella 3.6: Parametri variabili del GA

3.2.3 Risultati

Il tuning degli operatori di riproduzione è stato eseguito su tutti i pazienti del training set. Dal momento che l'algoritmo genetico è di tipo stocastico, per ciascuna combinazione sono state eseguite 50 ripetizioni e per ognuna di esse è stata salvata la soluzione migliore per valutare la stabilità della coppia di parametri. Il criterio di stabilità utilizzato è il numero di pazienti per cui la combinazione ha portato al raggiungimento della soluzione migliore per ogni specifico paziente. I valori assunti da questo criterio sono riportati nel grafico in figura 3.12. Dalla sua analisi la combinazione più stabile risulta essere la numero 13, con un totale di 137 pazienti su un massimo di 140.

Successivamente sono stati osservati i risultati per i pazienti 106, 113 e 117, cioè quelli per cui la tredicesima combinazione non ha portato alla soluzione migliore. In particolare sono stati valutati gli scostamenti tra la fitness, la dose al retto e la dose alla vescica

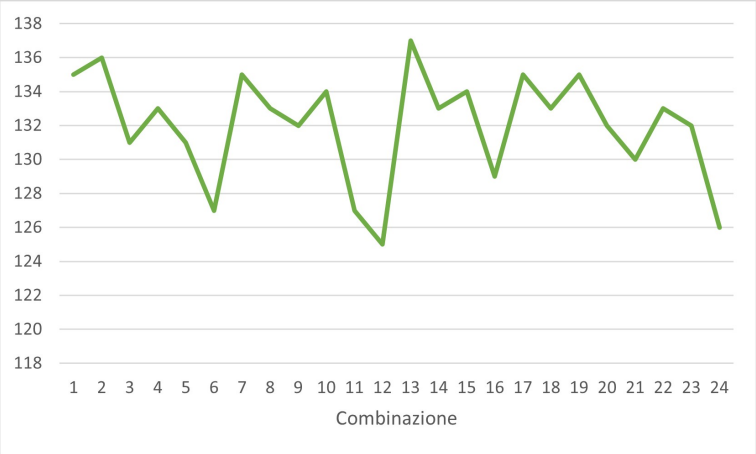


Figura 3.12: Parametro di stabilità

migliori in assoluto, ottenute con le altre combinazioni, e quelle raggiunta con la combinazione scelta. I risultati sono riportati nella tabella 3.7.

In tutti e tre i casi le variazioni sul valore della fitness sono molto basse, mentre per quanto riguarda i valori di dose si osserva una sola variazione di 20 cGy per ogni paziente, per cui si è proceduto con tale combinazione.

Paziente	fitness migliore	fitness comb. 13	Dose retto migliore	Dose retto comb. 13	Dose vescica migliore	Dose vescica comb. 13
106	6,0820	6,0827	1400	1380	1100	1100
113	10,2203	10,2218	1160	1140	1380	1380
117	3,6160	3,6161	4200	4200	1160	1180

Tabella 3.7: Analisi della tredicesima combinazione per i pazienti 106, 113 e 117

3.3 Generazione dei piani di radioterapia

Per ogni paziente del dataset sono stati prodotti dei piani di radioterapia utilizzando un algoritmo genetico con pc = 0.9 e pm = 0.1, come definiti dalla tredicesima combinazione, e che implementa una RBFNN allenata con 24 centri come funzione di fitness. I piani ottimizzati sono riportati in appendice F.

Capitolo 4

Ottimizzazione dei piani di radioterapia per cluster di pazienti

Terminata l'implementazione del primo metodo di ottimizzazione dei piani di radioterapia si è proceduto con l'implementazione del secondo metodo per valutare se è possibile ottenere dei piani radioterapeutici migliori suddividendo i pazienti in cluster omogenei sulla base delle caratteristiche geometriche della loro prostata.

4.1 Clustering dei pazienti

A partire dalle variabili geometriche estratte in un lavoro di tesi precedente [26] e riassunte nella tabella 4.1, è stata effettuata una normalizzazione dei valori con il metodo del min-max scaling al fine di poter confrontare le varie grandezze tra di loro. Successivamente è stato utilizzato un algoritmo gerarchico caratterizzato dalla distanza euclidea, come metrica, e dal legame completo, come metodo di clustering. Il ritultato è rappresentato dal dendrogramma in figura 4.1.

Variabile	Parametro geometrico
1	ID paziente
2	Lunghezza (b) del bounding box
3	Altezza (h) del bounding box
4	Profondità(s) del bounding box
5	Superficie del bounding box
6	Volume del bounding box
7	Convex volume
8	Diametro equivalente
9 - 11	Lunghezza assi principali

12	Superficie
13	Volume
14	Rapporto superficie/volume
15	PCAR
16 - 18	Valore minimo, massimo e medio delle basi del bounding box 2D
19 - 21	Valore minimo, massimo e medio delle altezze del bounding box 2D

Tabella 4.1: Variabili geometriche estratte dai pazienti.

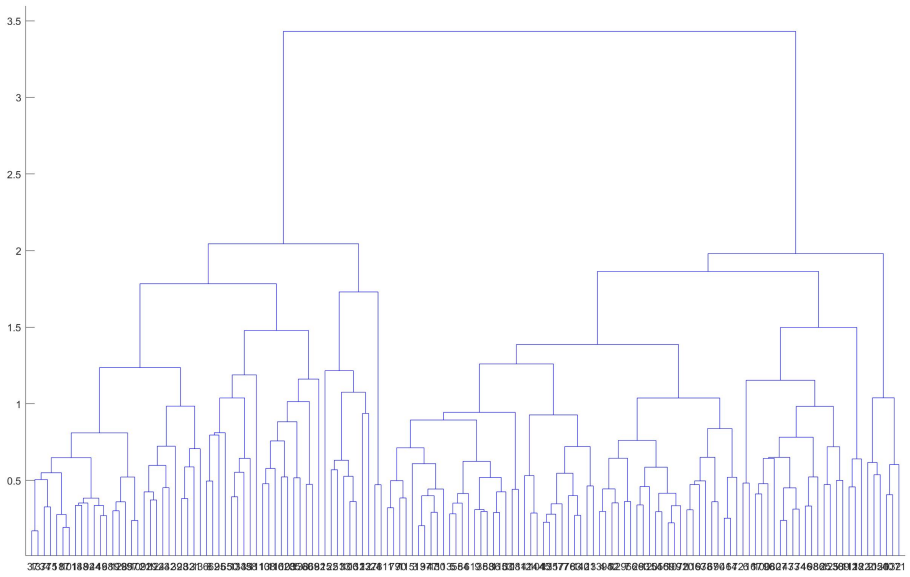
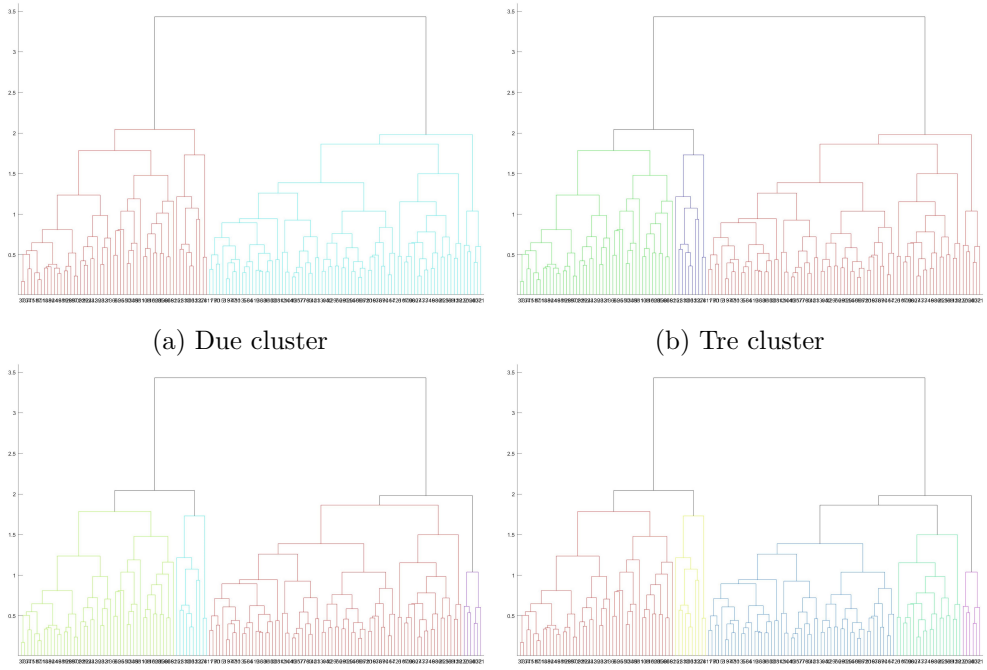


Figura 4.1: Dendrogramma.

In particolare, sono state analizzate le divisioni in due, tre, quattro e cinque cluster 4.2. Per ciascun tipo di clustering sono stati valutati:

- Le numerosità dei cluster, riportate nelle tabelle 4.2, 4.3, 4.4, 4.5;
- Le variabilità intra-cluster come media delle distanze di tutte le coppie tra di loro, le distanze inter-cluster tra i centroidi, calcolati come la mediana degli elementi all'interno del relativo cluster e i boxplot di ciascun parametro, riportati in appendice F;
- I coefficienti di silhouette, riportati in figura 4.3.



(a) Due cluster

(b) Tre cluster

(c) Quattro cluster

(d) Cinque cluster

Figura 4.2: Dendrogrammi.

Cluster	Numerosità
1	83
2	57

Tabella 4.2: Clustering a due cluster.

Cluster	Numerosità
1	83
2	10
3	47

Tabella 4.3: Clustering a tre cluster.

Cluster	Numerosità
1	6
2	77
3	10
4	47

Tabella 4.4: Clustering a quattro cluster.

Cluster	Numerosità
1	6
2	57
3	20
4	10
5	47

Tabella 4.5: Clustering a cinque cluster.

Dall'analisi della silhouette si evince che la divisione migliore è quella a due cluster, sebbene il coefficiente di quella a tre cluster sia molto simile. Si è deciso, quindi, di approfondire l'indagine e, per entrambi i clustering, è stato creato un silhouette plot (fig. 4.4). Dai due grafici si osserva che, nel caso con due cluster, gli elementi assumono valori

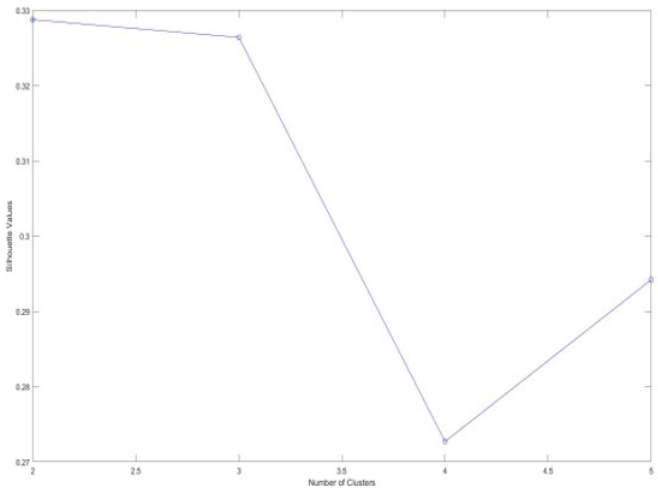
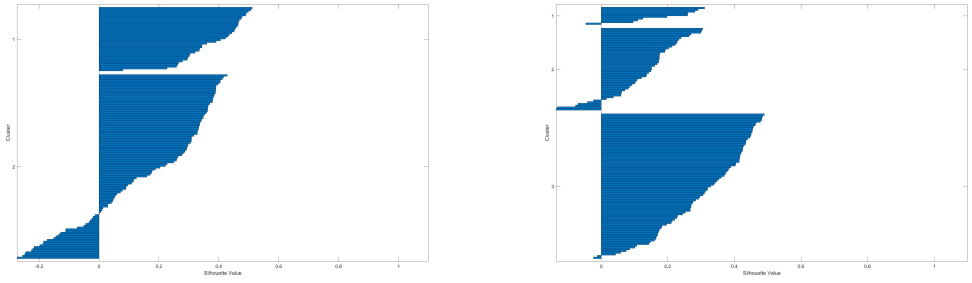


Figura 4.3: Coefficienti di silhouette per le 4 tipologie di clustering.

mediamente maggiori, per cui si può dedurre che sono stati raggruppati correttamente, tuttavia in uno dei due cluster è presente un numero maggiore di elementi con valore negativo rispetto al caso con tre cluster. Per questo motivo si è deciso di proseguire lo studio con entrambe le tipologie di clustering e di valutare successivamente quella con i risultati migliori.



(a) Due cluster

(b) Tre cluster

Figura 4.4: Silhouette plot.

4.2 Radial basis function neural network

4.2.1 Analisi del numero di centri

Per la scelta del numero migliore di centri per l'allenamento delle RBFNN, sono stati prima suddivisi i pazienti tra i vari cluster di appartenenza e successivamente sono state condotte le stesse analisi condotte per il primo metodo di ottimizzazione dei piani. Quindi, a partire dagli stessi dati e per ogni cluster, sono stati individuati i centri più ricorrenti tra le reti migliori di ciascun paziente e i centri per cui il maggior numero di pazienti presentasse almeno una rete, tra le sei ottenute con il metodo del k-fold, con valore MSE inferiore a uno.

I risultati della prima selezione sono riportati sotto forma di istogramma delle frequenze in figura 4.5 e figura 4.6. I numeri di centri più ricorrenti tra le reti migliori sono:

- Due cluster:
  - Cluster 1: 30 centri;
  - Cluster 2: 30 centri.
- Tre cluster:
  - Cluster 1: 30 centri;
  - Cluster 2: 22, 29 e 30 centri;
  - Cluster 3: 30 centri.

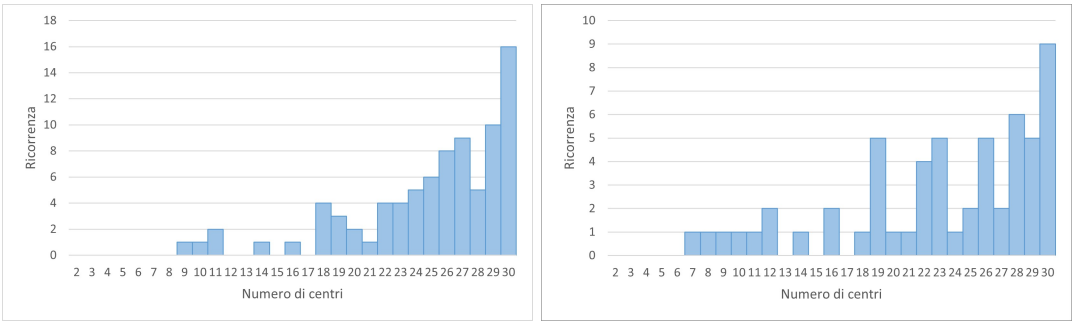
Per quanto riguarda la seconda selezione, i risultati sono riportati nelle tabelle 4.6. I centri selezionati in questo caso sono:

- Due cluster:
  - Cluster 1: 21 centri;
  - Cluster 2: 26 centri.
- Tre cluster:
  - Cluster 1: 21 centri;
  - Cluster 2: 6,7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 e 30 centri;
  - Cluster 3: 26 centri.

Infine, sono state valutate le distribuzioni del valore MSE migliore per ogni paziente ottenute con le reti individuate precedentemente per ciascun clustering. I risultati sono stati riportati nelle figure 4.7 e 4.8.

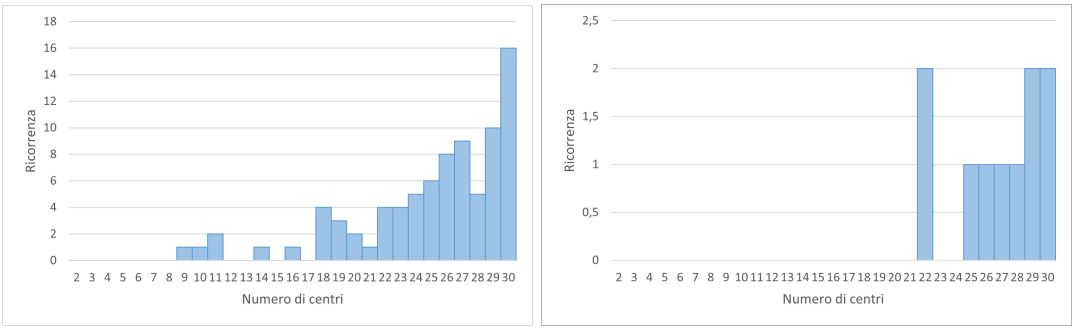
Tenendo conto di tutti i risultati ottenuti, si è deciso di proseguire con la divisione dei pazienti in due cluster, l'associazione di ogni paziente al relativo cluster è riportata in appendice E. In particolare sono state scelte la rete con 21 centri per il primo cluster e la rete con 26 centri per il secondo cluster, in quanto le distribuzioni dei loro valori MSE

presentano valori minimi confrontabili con quelli delle reti con 30 centri e valori medi e massimi inferiori in entrambi i casi.

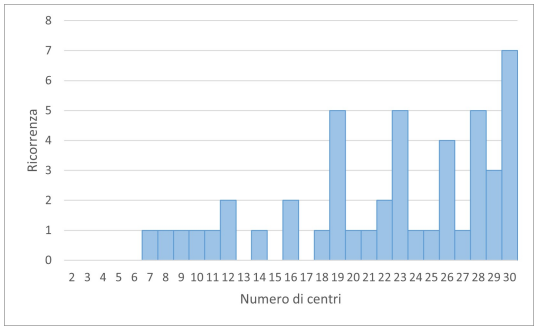


(a) Cluster 1 (b) Cluster 2

Figura 4.5: Istogramma della frequenza del numero di centri migliore



(a) Cluster 1 (b) Cluster 2



(c) Cluster 3

Figura 4.6: Istogramma della frequenza del numero di centri migliore

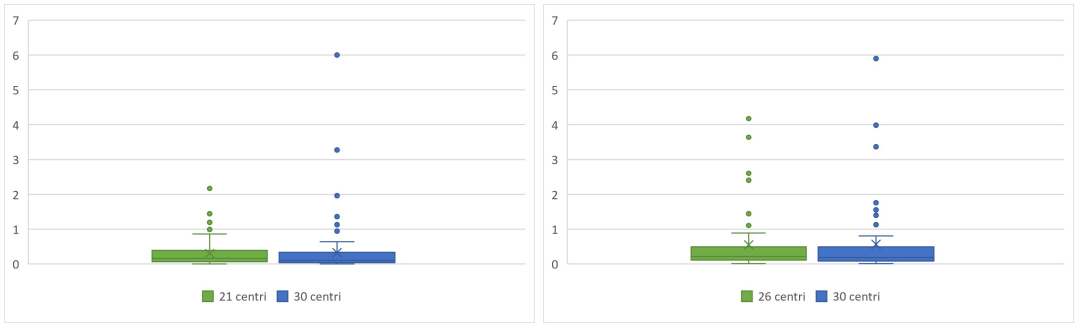
Numero di centri	Numerosità cluster 1	Numerosità cluster 2
2	35	11
3	8	6
4	13	5
5	30	12
6	44	27
7	56	37
8	67	43
9	70	46
10	70	43
11	69	43
12	68	48
13	67	43
14	71	46
15	72	47
16	78	46
17	76	47
18	79	46
19	77	47
20	78	47
21	80	47
22	78	46
23	76	45
24	79	48
25	75	48
26	75	49
27	76	48
28	76	47
29	75	48
30	78	48

(a) Due cluster

Numerosità cluster 1	Numerosità cluster 2	Numerosità cluster 3
35	1	10
8	35	6
13	8	5
30	14	11
44	37	20
56	52	29
67	64	35
70	75	38
70	78	35
69	78	35
68	77	40
67	75	36
71	75	38
72	79	39
78	80	38
76	86	39
79	84	38
77	87	39
78	85	39
80	86	39
78	88	38
76	86	37
79	84	40
75	87	40
75	83	41
76	83	40
76	84	39
75	84	40
78	83	40

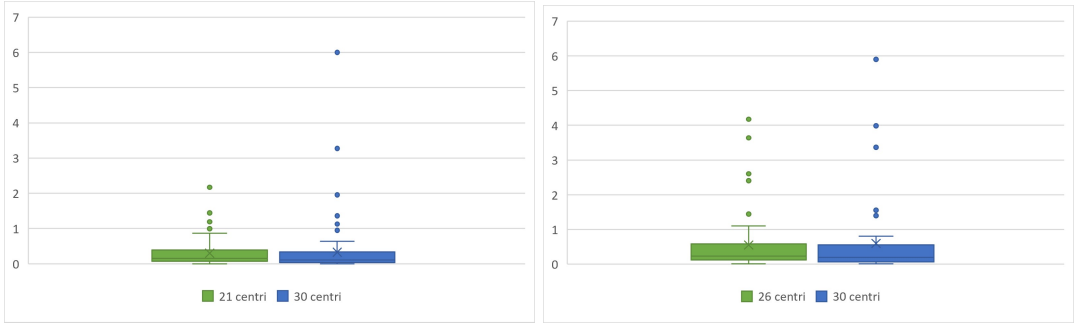
(b) Tre cluster

Tabella 4.6: Numero di reti che hanno un valore MSE inferiore a uno per ogni numero di centri

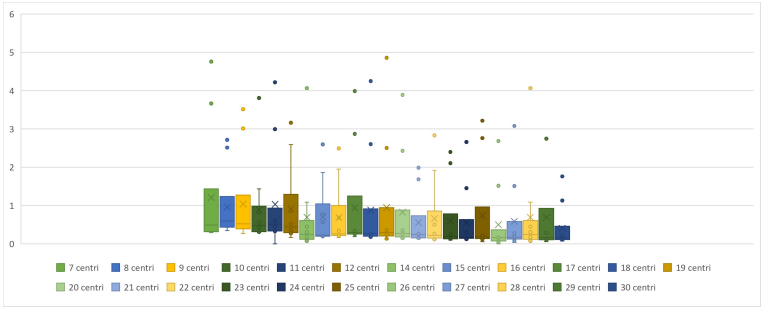


(a) Cluster 1 (b) Cluster 2

Figura 4.7: Boxplot delle distribuzioni dei valori MSE migliori per le reti con il numero di centri migliore per due cluster



(a) Cluster 1 (b) Cluster 3



(c) Cluster 2

Figura 4.8: Boxplot delle distribuzioni dei valori MSE migliori per le reti con il numero di centri migliore per tre cluster

4.2.2 Assegnazione dei pazienti del test set al relativo cluster

Prima di procedere con la validazione dei numeri di centri scelti è stato necessario assegnare ciascuno dei trentacinque pazienti del test set a uno dei due cluster. Tale operazione è stata effettuata in due modi diversi:

- Distanza minore dal centroide: il paziente viene assegnato al cluster per cui la sua distanza dal centroide è minore;

- Metodo complete: il paziente viene assegnato al cluster per cui la distanza dall'elemento più lontano è la minore.

I risultati sono riportati nella tabella 4.7, da cui si vede che i risultati coincidono per entrambi i metodi, per cui l'assegnazione dei pazienti ai due cluster è stabile.

4.2.3 Validazione del numero di centri

Così come nel caso senza clustering, per validare la scelta del numero di centri, sono state allenate, per ciascun pazienti del test set, sei reti con il metodo del k-fold e il relativo numero di centri. I valori MSE ottenuti sono stati riportati nelle tabelle 4.8 e 4.9. Da queste si vede che per il cluster 1 esiste almeno una combinazione per ogni paziente per cui il valore MSE è inferiore a uno, mentre per i pazienti 165 e 211 del cluster 2 non esiste una combinazione tale per cui questa condizione si verifichi, nonostante ciò i loro valori più bassi sono prossimi a uno, per cui le scelte dei centri sono state confermate.

Paziente	Validation set 1	Validation set 2	Validation set 3	Validation set 4	Validation set 5	Validation set 6
3	0,925	0,224	0,864	0,120	0,829	1,387
23	1,357	0,557	0,304	0,053	1,366	1,194
24	2,595	1,782	1,766	1,186	0,842	1,169
47	2,610	0,546	0,517	1,226	0,694	1,840
53	0,417	0,251	0,199	0,080	0,156	0,114
93	0,400	0,098	0,045	0,086	0,104	0,114
125	0,433	0,277	0,115	0,264	0,486	0,400
161	0,323	0,160	0,120	0,087	0,556	0,352
164	0,150	0,334	0,137	0,516	0,345	0,062
178	0,976	0,571	1,139	0,165	0,534	1,267
192	0,471	0,164	0,119	0,049	0,142	0,600
195	0,853	3,110	1,335	0,521	1,470	1,599
199	0,267	0,145	0,563	0,179	0,071	0,320
200	0,468	1,470	0,235	0,352	1,429	0,761
202	0,228	0,592	0,203	0,369	0,110	0,345
222	0,680	0,247	0,056	0,051	0,748	1,083
235	1,077	0,498	0,182	0,145	0,120	1,080
251	0,345	0,142	0,125	0,126	0,221	0,524

Tabella 4.8: Valori MSE per la rete con 21 centri sul validation set dei pazienti del cluster 1 del test set

Paziente	Distanza cluster 1	Distanza cluster 2	Cluster assegnato
3	1,01	1,83	1
13	1,04	0,92	2
23	0,77	1,13	1
24	0,51	1,28	1
39	1,06	0,94	2
41	1,79	1,07	2
47	0,50	0,78	1
53	0,86	0,88	1
75	2,28	1,43	2
88	1,43	0,87	2
93	0,76	1,51	1
124	1,26	1,03	2
125	0,55	1,34	1
138	1,44	0,92	2
161	0,67	0,94	1
164	0,74	1,60	1
165	1,96	1,32	2
178	0,86	1,46	1
181	1,20	0,90	2
192	1,46	2,31	1
195	0,89	0,92	1
199	0,88	1,62	1
200	0,89	1,57	1
202	0,72	1,20	1
211	1,46	0,88	2
214	1,55	0,99	2
222	0,95	1,07	1
224	0,79	0,74	2
231	2,48	1,67	2
235	0,85	0,99	1
236	0,93	0,72	2
244	1,45	0,91	2
249	1,63	1,24	2
251	0,73	1,07	1
253	1,01	0,76	2

(a) Metodo della distanza minore dal centroide

Distanza cluster 1	Distanza cluster 2	Cluster assegnato
1,64	2,99	1
2,03	1,97	2
1,71	2,20	1
1,38	2,50	1
2,13	1,94	2
2,99	1,93	2
1,63	1,99	1
1,86	1,90	1
3,47	2,28	2
2,63	1,55	2
1,49	2,68	1
2,18	2,02	2
1,36	2,55	1
2,58	1,82	2
1,70	2,10	1
1,60	2,81	1
3,12	1,80	2
1,61	2,62	1
2,22	1,82	2
2,03	3,46	1
1,87	2,02	1
1,75	2,83	1
1,52	2,74	1
1,65	2,41	1
2,53	1,82	2
2,62	1,63	2
1,86	2,20	1
1,90	1,80	2
3,67	2,21	2
1,81	2,05	1
2,01	1,77	2
2,57	1,66	2
2,69	1,76	2
1,61	2,21	1
2,06	1,66	2

(b) Metodo complete

Tabella 4.7: Assegnazione dei pazienti del test set ai due cluster

Paziente	Validation set 1	Validation set 2	Validation set 3	Validation set 4	Validation set 5	Validation set 6
13	0,087	0,155	0,195	0,220	0,601	2,326
39	0,330	0,100	0,122	0,028	0,044	0,344
41	0,976	0,569	0,152	0,055	0,473	1,109
75	0,386	0,287	0,214	0,298	0,855	0,353
88	0,689	0,310	0,157	0,143	1,004	0,372
124	1,571	0,919	0,600	0,295	0,072	0,627
138	0,927	0,820	0,155	0,200	0,577	1,232
165	1,506	2,510	2,649	1,296	1,143	2,194
181	2,141	0,666	0,454	0,272	1,111	1,022
211	2,111	1,041	1,651	4,382	1,608	3,197
214	0,955	0,177	0,106	0,298	0,262	0,628
224	0,830	0,400	0,501	0,578	0,469	0,340
231	0,593	0,184	0,113	0,152	1,338	0,490
236	1,068	0,470	0,198	0,364	0,062	0,157
244	0,207	0,125	0,056	0,181	0,175	0,567
249	1,074	0,405	0,529	0,276	0,407	3,918
253	0,629	0,295	0,511	0,444	5,141	1,346

Tabella 4.9: Valori MSE per la rete con 26 centri sul validation set dei pazienti del cluster 2 del test set

4.3 Algoritmo genetico

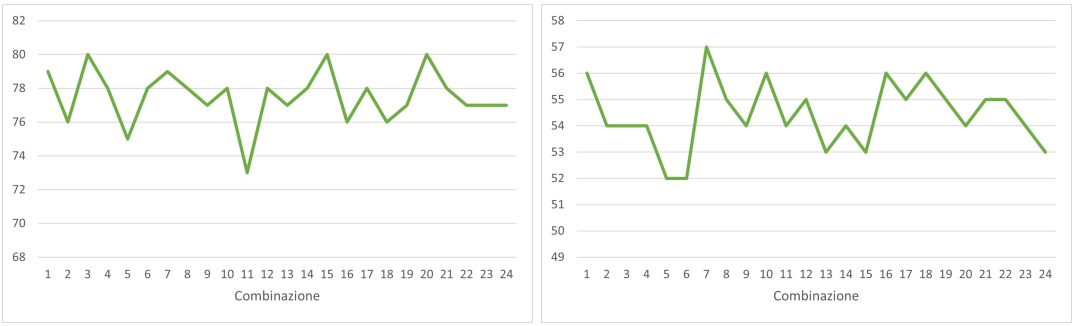
Anche in questo caso sono stati avviati due GA, uno per ogni cluster, con gli stessi parametri fissi adoperati nel capitolo antecedente. Per quanto riguarda il tuning dei parametri variabili sono state utilizzate le stesse modalità e le stesse combinazioni analizzate precedentemente su tutti i pazienti del training set.

I valori ottenuti con il parametro di stabilità sono riassunti nei grafici in figura 4.9.

Per il cluster 2 la combinazione migliore è la settima, in quanto tutti i 57 pazienti hanno raggiunto la combinazione migliore. Per quanto riguarda il cluster 1 sono state individuate tre combinazioni migliori: la numero 3, la numero 15 e la numero 20, con un totale di 80 pazienti su un massimo di 83, per cui è stato condotto uno studio più approfondito sulla stabilità delle tre combinazioni in due modi:

- Numero di soluzioni con fitness minima ottenuta per tutti i pazienti;
- Numero di volte in cui ciascuna combinazione dà il valore più elevato tra le tre scelte.

Con il primo metodo sono stati ottenuti un punteggio pari a 618 per la combinazione 3, pari a 670 per la combinazione 15 e pari a 729 per la combinazione 30. Con il secondo metodo sono stati ottenuti un punteggio pari a 33 per le combinazioni 3 e 15 e un punteggio



(a) Cluster 1 (b) Cluster 2

Figura 4.9: Criterio 1

pari a 41 per la combinazione 20. Per cui è stato confermato che la ventesima risulta essere la più stabile.

Infine, sono stati valutati gli scostamenti per i tre pazienti per cui la combinazione non ha raggiunto il valore di fitness migliore e i risultati sono stati riportati nella tabella 4.10. In tutti e tre i casi le variazioni sul valore della fitness sono molto basse e le variazioni di dose sono limitate a un massimo di 20 cGy per distretto anatomico, per cui si è proceduto con tale combinazione.

Paziente	Fitness migliore	Fitness comb. 20	Dose retto migliore	Dose retto comb. 20	Dose vescica migliore	Dose vescica comb. 20
6	8,157	8,157	1320	1320	2600	2620
189	3,784	3,785	1060	1060	3020	3000
212	3,292	3,292	660	640	3200	3220

Tabella 4.10: Valori MSE per la rete con 26 centri sul validation set dei pazienti del cluster 2 del test set

4.4 Generazione dei piani di radioterapia

In conclusione, per i pazienti del cluster 1 sono stati prodotti dei piani di radioterapia utilizzando un algoritmo genetico con pc = 0.95 e pm = 0.1, come definiti dalla ventesima combinazione, e che implementa una RBFNN allenata con 21 centri come funzione di fitness. Per i pazienti appartenenti al cluster 2 sono stati prodotti dei piani di radioterapia utilizzando un algoritmo genetico con pc = 0,85 e pm = 0,1, come definiti dalla settima combinazione, e che implementa una RBFNN allenata con 26 centri come funzione di fitness. I piani ottimizzati sono riportati in appendice F.

Capitolo 5

Analisi dei piani radioterapeutici

Per dimostrare la validità dei piani prodotti in questo lavoro, questi sono stati confrontati con i piani radioterapeutici di riferimento prodotti tramite il TPS di RayStation.

5.1 Valutazione dei piani radioterapeutici

Per poter comparare i diversi piani per ogni paziente sono stati valutati con un algoritmo implementato da RayStation, che associa un punteggio a ognuno di essi sulla base del raggiungimento dei Clinical Goals, cioè i limiti inferiori e superiori delle dosi fornite al target e agli OAR. I punteggi ottenuti sono riportati sotto forma percentuale in appendice G.

5.2 Confronto dei piani radioterapeutici

Nel boxplot in figura 5.1 sono rappresentate le distribuzioni degli score attribuiti a ciascuno dei tre piani per ogni paziente del training set. Nella tabella 5.1 sono state confrontate le medie e le mediane delle tre distribuzioni. Inoltre sono stati confrontati i piani prodotti in questo lavoro con quelli prodotti tramite il GPS in termini di risultati migliori per ogni singolo paziente. Da un'analisi dei risultati si osserva che entrambi i metodi hanno prodotto un numero di piani migliori maggiore rispetto a quelli prodotti dal GPS e tra i due quello che prevede il clustering dei pazienti funziona meglio.

Figura 5.1: Boxplot delle distribuzioni degli score associati ai piani dei pazienti del training set

	Senza cluster	Con cluster	GPS
Media	95%	95%	92%
Mediana	100%	100%	98%

Tabella 5.1: Medie e mediane delle distribuzioni degli score associati ai piani dei pazienti del training set

	Migliore del GPS	Uguale al GPS	Peggiore del GPS
Senza cluster	53	35	51
Con cluster	57	41	41

Tabella 5.2: Confronto dei piani per i pazienti del training set

Tali osservazioni sono state confermate dalle stesse analisi condotte sui pazienti del test set, i cui risultati sono riportati in figura 5.2 e nelle tabelle 5.3 e 5.4. In questo caso, però, i due metodi possono essere considerati equivalenti tra di loro.

	Senza cluster	Con cluster	GPS
Media	97%	97%	95%
Mediana	98%	98%	99%

Tabella 5.3: Medie e mediane delle distribuzioni degli score associati ai piani dei pazienti del test set

Figura 5.2: Boxplot delle distribuzioni degli score associati ai piani dei pazienti del test set

	Migliore del GPS	Uguale al GPS	Peggiore del GPS
Senza cluster	13	6	15
Con cluster	10	9	15

Tabella 5.4: Confronto dei piani per i pazienti del training set

Infine è stata condotta un'analisi statistica tramite il test di Mann-Whitney accoppiato a due code per confrontare le tre combinazioni di distribuzioni. I p-value ottenuti sia nel confronto tra i pazienti del training set che del test set sono riportati nella tabella 5.5. In tutti i casi è stato ottenuto un valore maggiore del livello di significatività del test (pari a 0,05), per cui le distribuzioni sono statisticamente equivalenti tra di loro e i metodi costruiti in questo lavoro di tesi sono clinicamente validi.

	Senza cluster vs GPS	Con cluster vs GPS	Senza cluster vs Con cluster
Training	0,686	0,125	0,302
Validation	0,926	0,74	0,857

Tabella 5.5: P-value ottenuti con il test di Mann-Whitney

Capitolo 6

Conclusioni

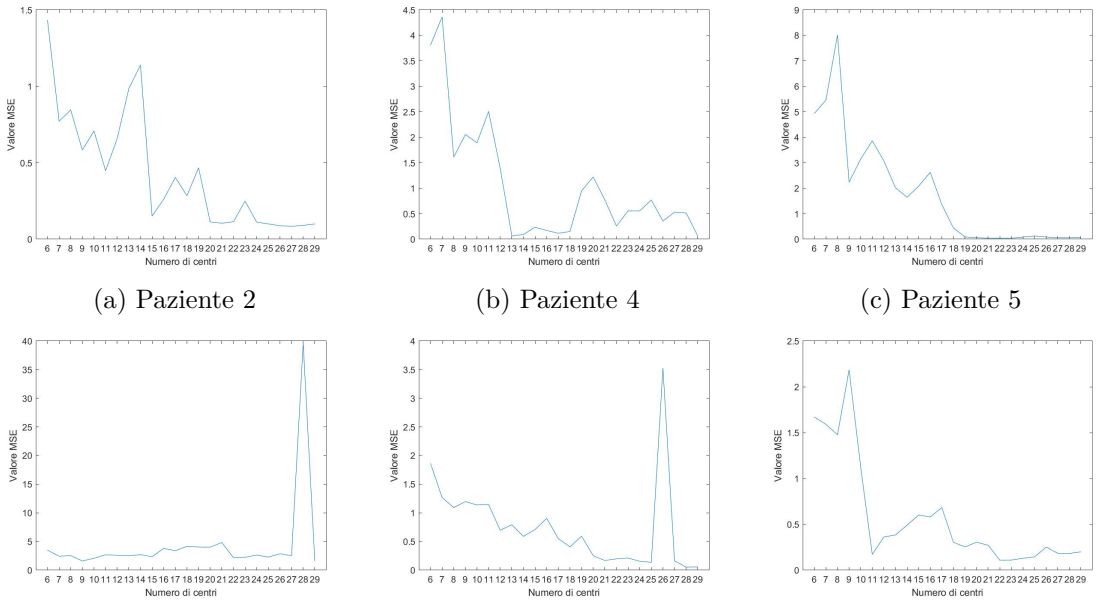
Lo scopo del presente lavoro di tesi consiste nell'implementazione di due metodi di ottimizzazione della coppia di dosi da fornire alla vescica e al retto, per pazienti affetti dal cancro alla prostata, utilizzando approcci differenti rispetto a quelli adoperati negli studi precedenti con il fine di ottenere risultati migliori.

I due metodi avanzati differiscono nella suddivisione dei pazienti del dataset: nel primo caso non è stata condotta; nel secondo caso ha previsto il clustering dei pazienti in due sottogruppi sulla base della caratterizzazione geometrica della prostata. Per ciascuno metodo è stato ottimizzato, tramite tecniche di clustering, il numero di centri della rete RBFNN, successivamente allenata con il metodo dei centri fissi e implementata come funzione di fitness dell'algoritmo genetico. L'algoritmo genetico è stato, a sua volta, utilizzato per individuare il piano migliore per ogni paziente in seguito a un processo di ottimizzazione della combinazione dei parametri variabili effettuato con un nuovo criterio di stabilità. Inoltre, in tutti e due i casi è stato condotta una validazione dei risultati su sottoinsiemi di pazienti esclusi dalla fase di training.

I piani radioterapeutici generati con le due diverse modalità sono stati valutati con un algoritmo sviluppato da RayStation, ottenendo punteggi con media pari al 95%, superiore rispetto a quelli dei piani prodotti dal TPS e in generale migliori per il secondo metodo. Infine, da un punto di vista statistico, le distribuzioni degli score dei tre metodi risultano equivalenti, per cui è stata dimostrata la validità clinica dei due sistemi costruiti in questo lavoro.

Appendice A

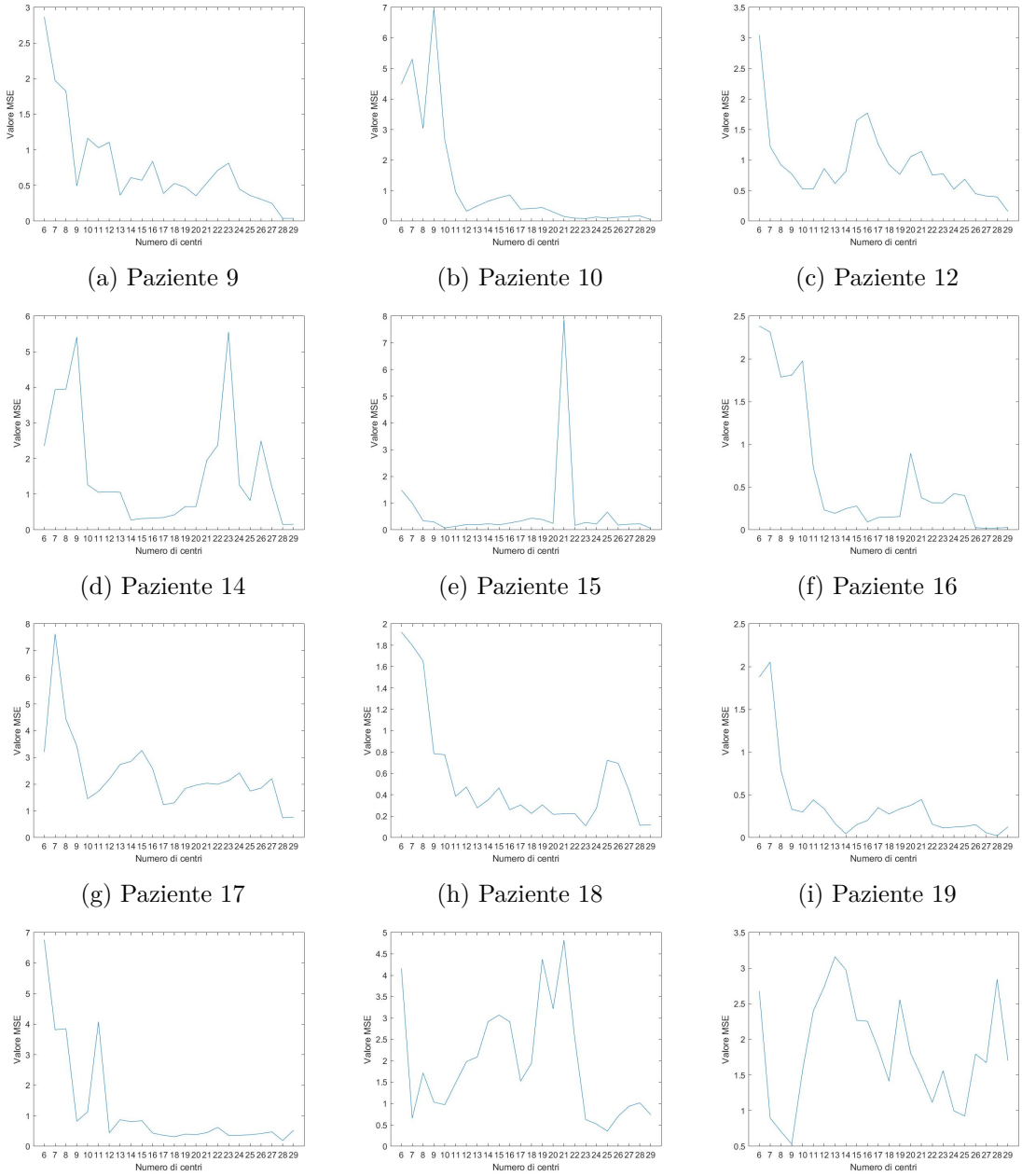
Valori MSE: allenamento RBFNN



(a) Paziente 2 (b) Paziente 4 (c) Paziente 5

(d) Paziente 6 (e) Paziente 7 (f) Paziente 8

Figura A.1: Valori MSE sul validation set dei pazienti 2, 4, 5, 6, 7 e 8



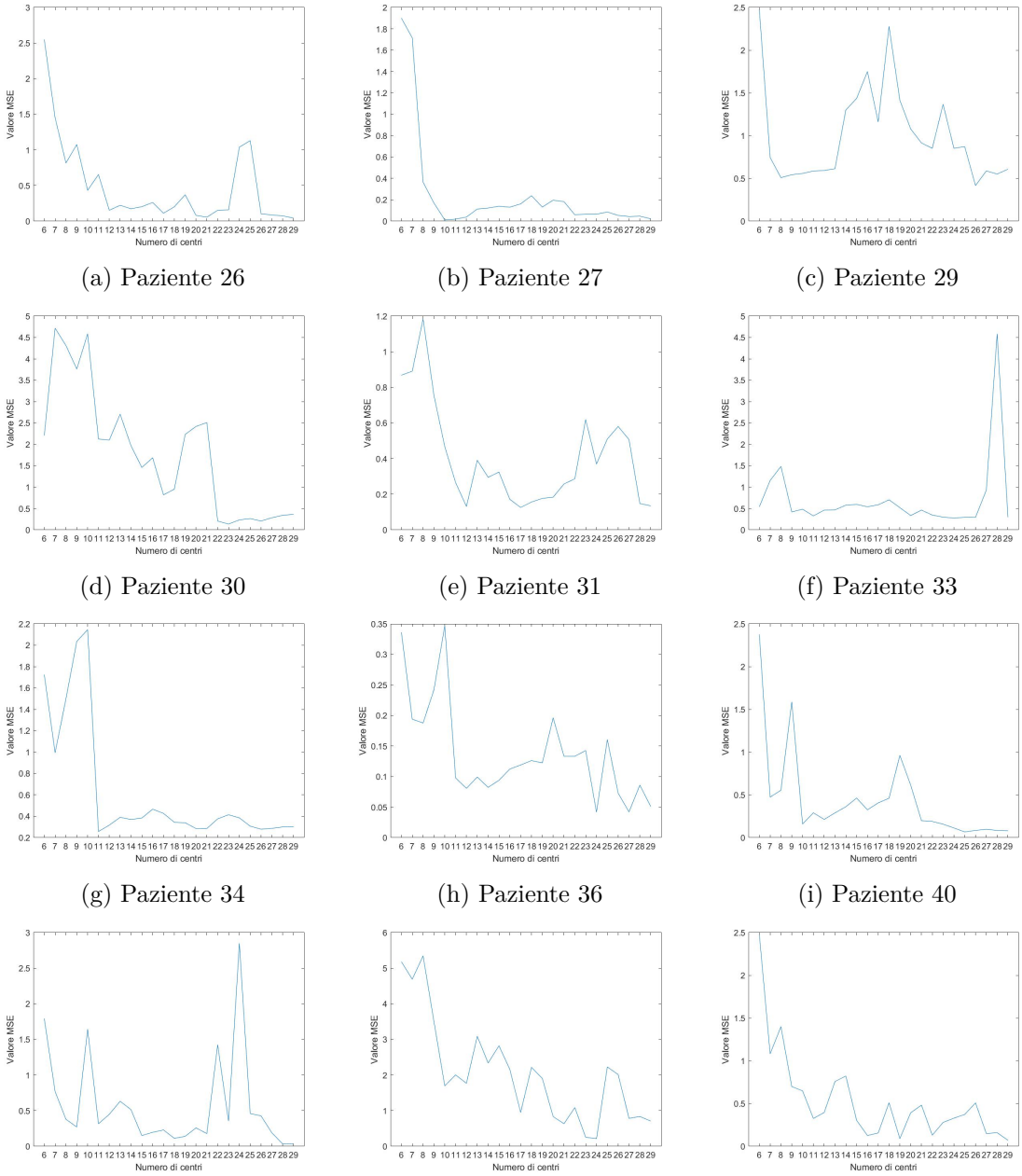
(a) Paziente 9 (b) Paziente 10 (c) Paziente 12

(d) Paziente 14 (e) Paziente 15 (f) Paziente 16

(g) Paziente 17 (h) Paziente 18 (i) Paziente 19

(j) Paziente 20 (k) Paziente 22 (l) Paziente 25

Figura A.2: Valori MSE sul validation set dei pazienti 9, 10, 12, 14, 16, 17, 18, 19, 20, 22 e 25



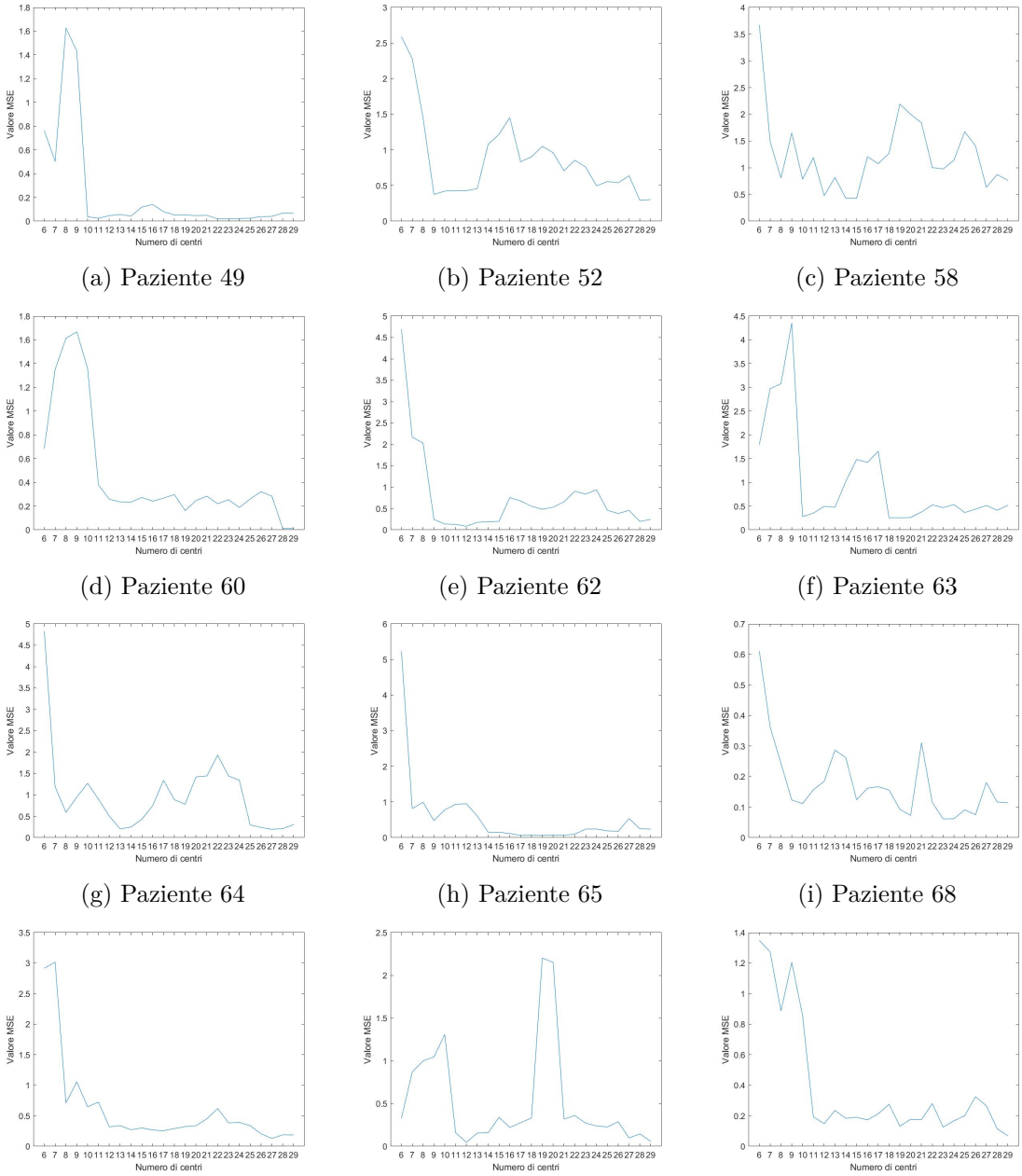
(a) Paziente 26 (b) Paziente 27 (c) Paziente 29

(d) Paziente 30 (e) Paziente 31 (f) Paziente 33

(g) Paziente 34 (h) Paziente 36 (i) Paziente 40

(j) Paziente 43 (k) Paziente 45 (l) Paziente 48

Figura A.3: Valori MSE sul validation set dei pazienti 26, 27, 29, 30, 31, 33, 34, 36, 40, 43, 45 e 48



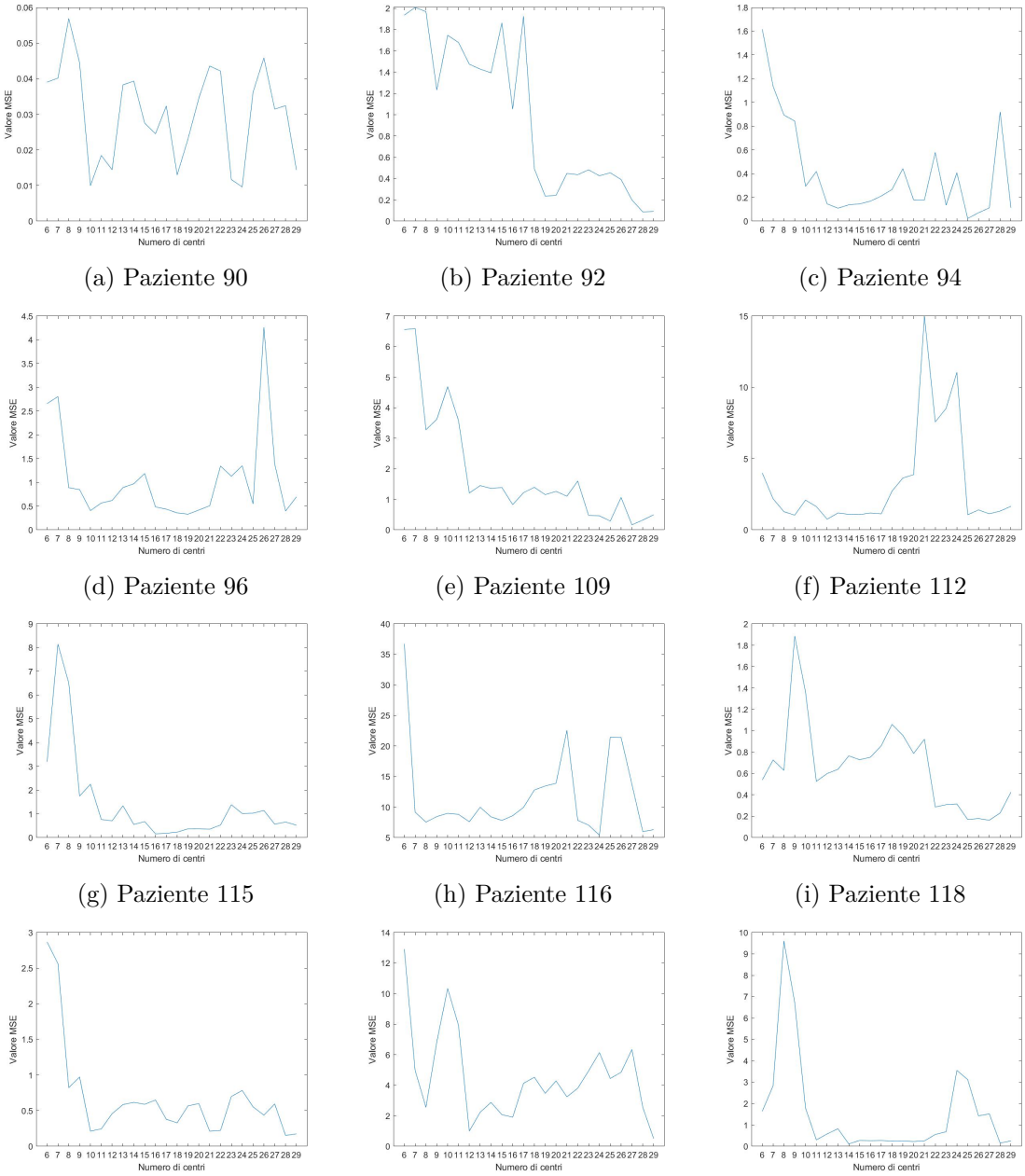
(a) Paziente 49 (b) Paziente 52 (c) Paziente 58

(d) Paziente 60 (e) Paziente 62 (f) Paziente 63

(g) Paziente 64 (h) Paziente 65 (i) Paziente 68

(j) Paziente 73 (k) Paziente 87 (l) Paziente 89

Figura A.4: Valori MSE sul validation set dei pazienti 49, 52, 58, 60, 62, 63, 64, 65, 68, 73, 87 e 89



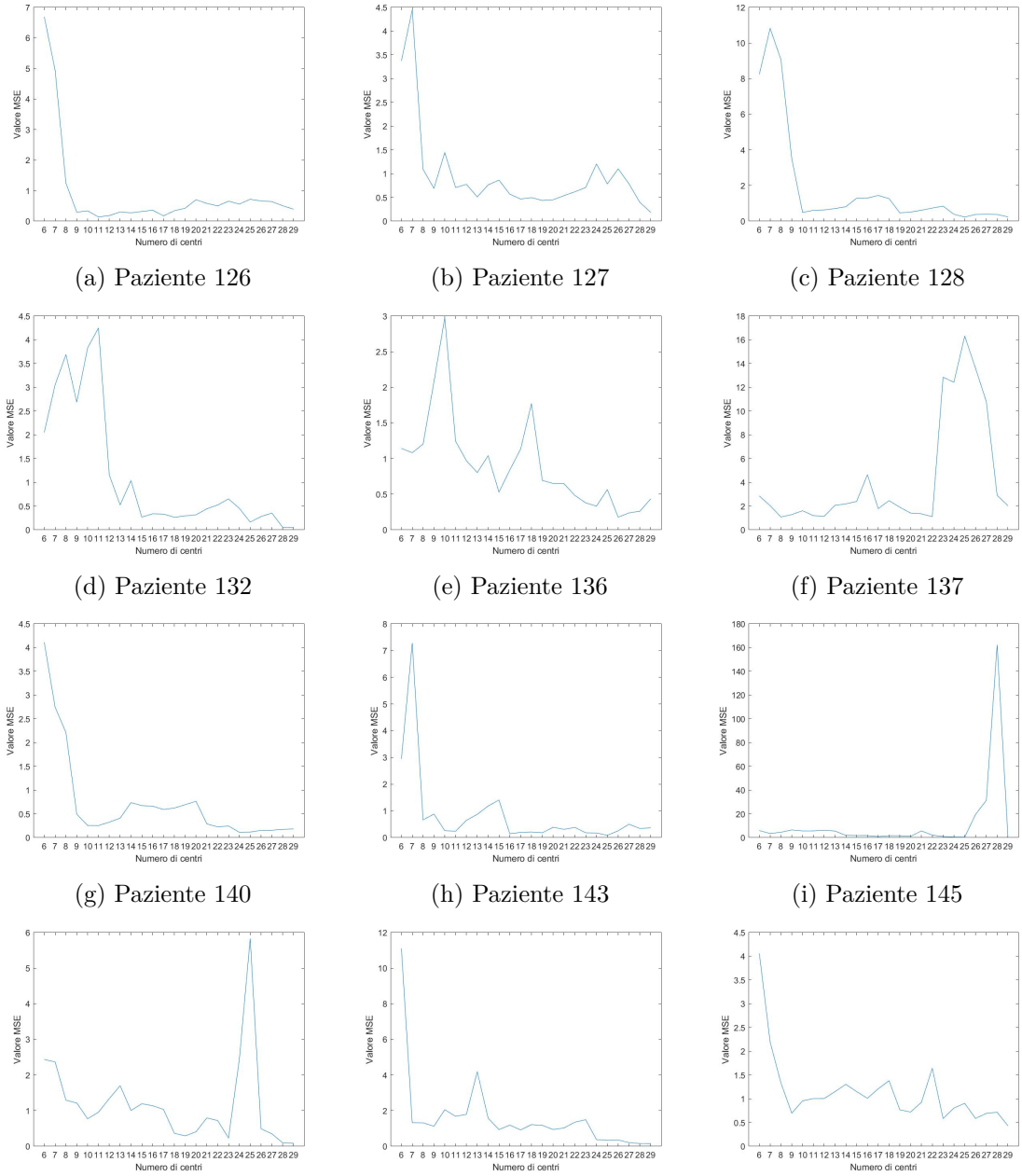
(a) Paziente 90 (b) Paziente 92 (c) Paziente 94

(d) Paziente 96 (e) Paziente 109 (f) Paziente 112

(g) Paziente 115 (h) Paziente 116 (i) Paziente 118

(j) Paziente 120 (k) Paziente 121 (l) Paziente 122

Figura A.5: Valori MSE sul validation set dei pazienti 90, 92, 94, 96, 109, 112, 115, 116, 118, 120, 121 e 122



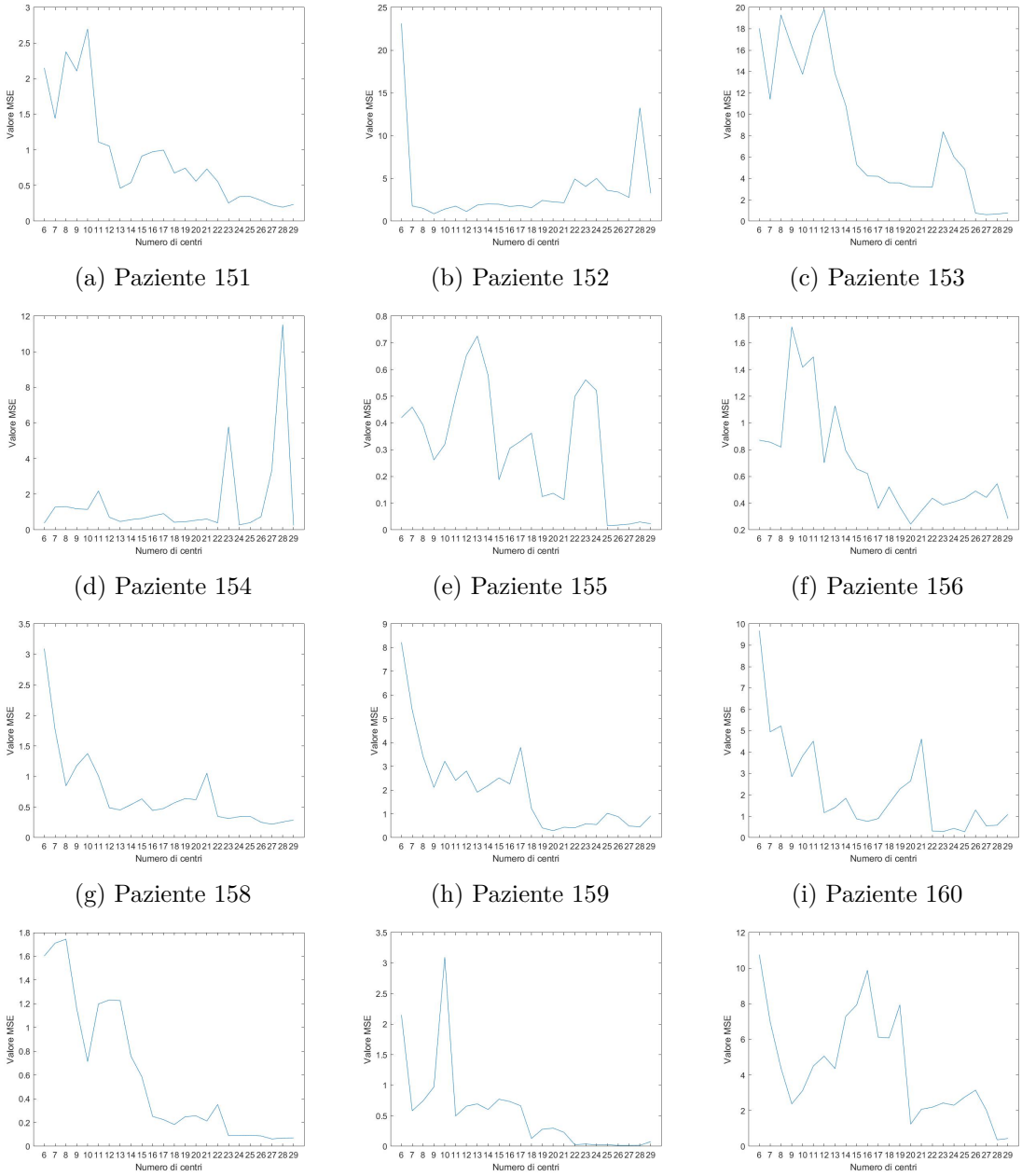
(a) Paziente 126 (b) Paziente 127 (c) Paziente 128

(d) Paziente 132 (e) Paziente 136 (f) Paziente 137

(g) Paziente 140 (h) Paziente 143 (i) Paziente 145

(j) Paziente 146 (k) Paziente 147 (l) Paziente 150

Figura A.6: Valori MSE sul validation set dei pazienti 126, 127, 128, 132, 136, 137, 140, 143, 145, 146, 147 e 150



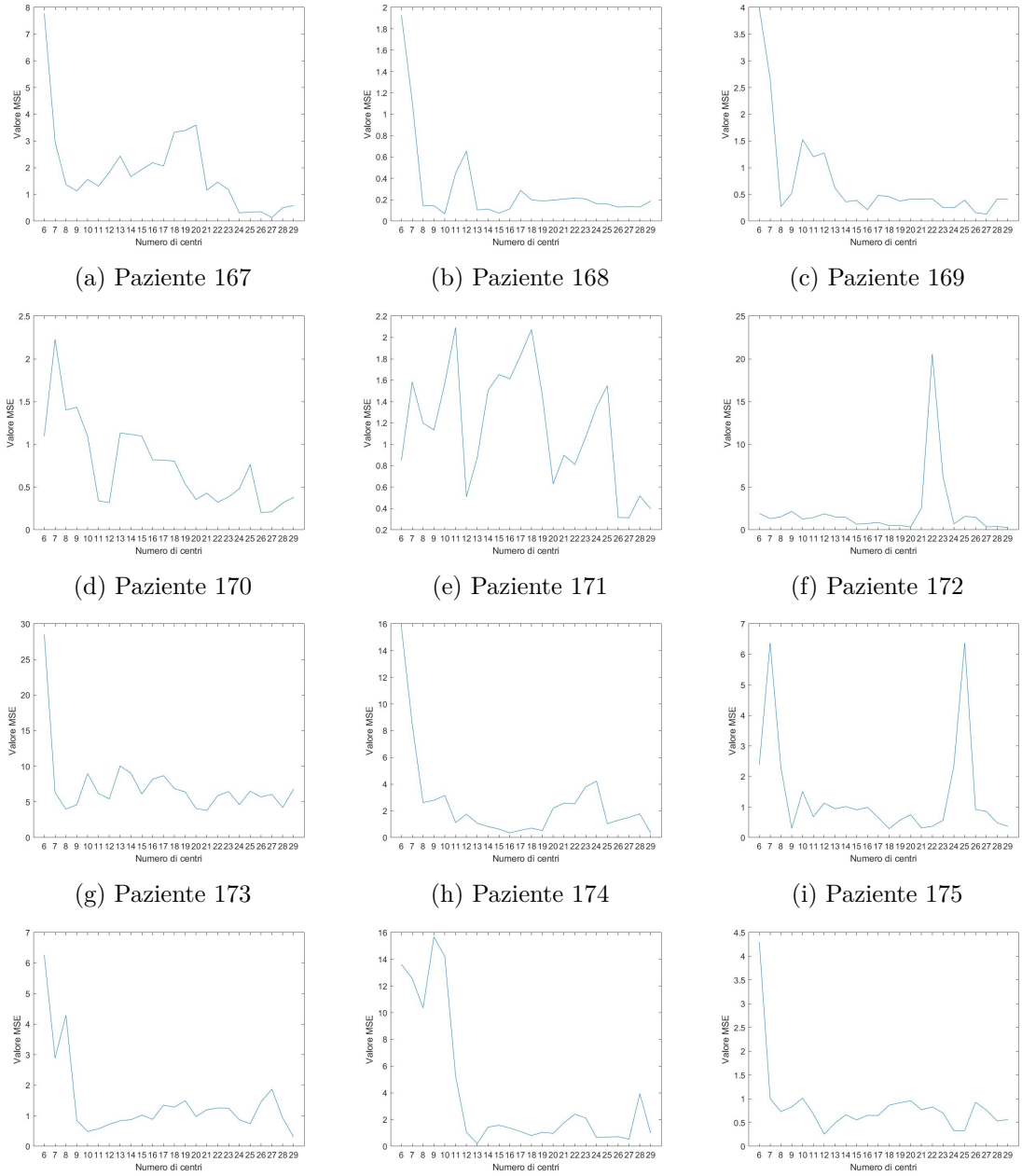
(a) Paziente 151 (b) Paziente 152 (c) Paziente 153

(d) Paziente 154 (e) Paziente 155 (f) Paziente 156

(g) Paziente 158 (h) Paziente 159 (i) Paziente 160

(j) Paziente 162 (k) Paziente 163 (l) Paziente 166

Figura A.7: Valori MSE sul validation set dei pazienti 151, 152, 153, 154, 155, 156, 158, 159, 160, 162, 163 e 166



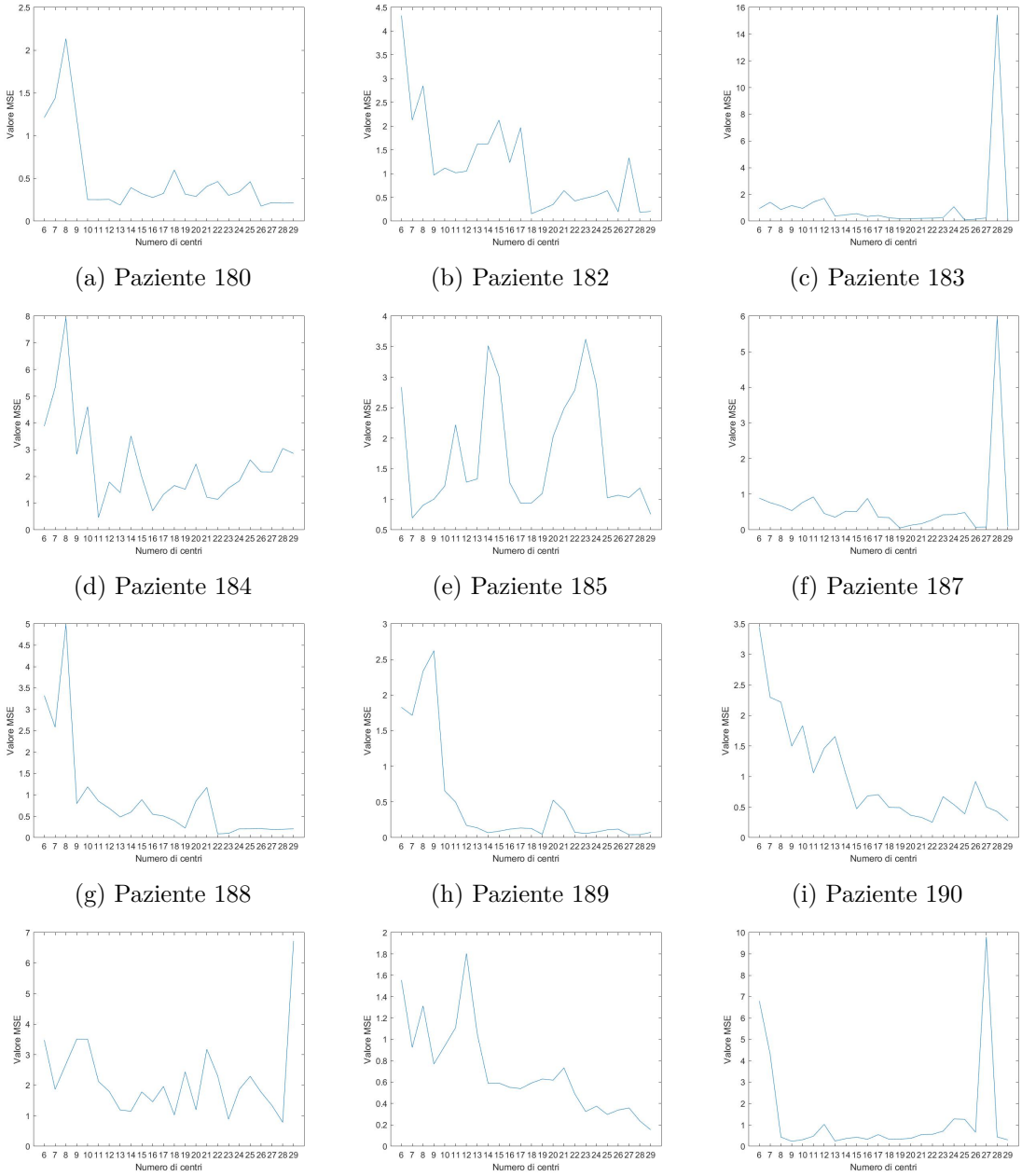
(a) Paziente 167 (b) Paziente 168 (c) Paziente 169

(d) Paziente 170 (e) Paziente 171 (f) Paziente 172

(g) Paziente 173 (h) Paziente 174 (i) Paziente 175

(j) Paziente 176 (k) Paziente 177 (l) Paziente 179

Figura A.8: Valori MSE sul validation set dei pazienti 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177 e 179



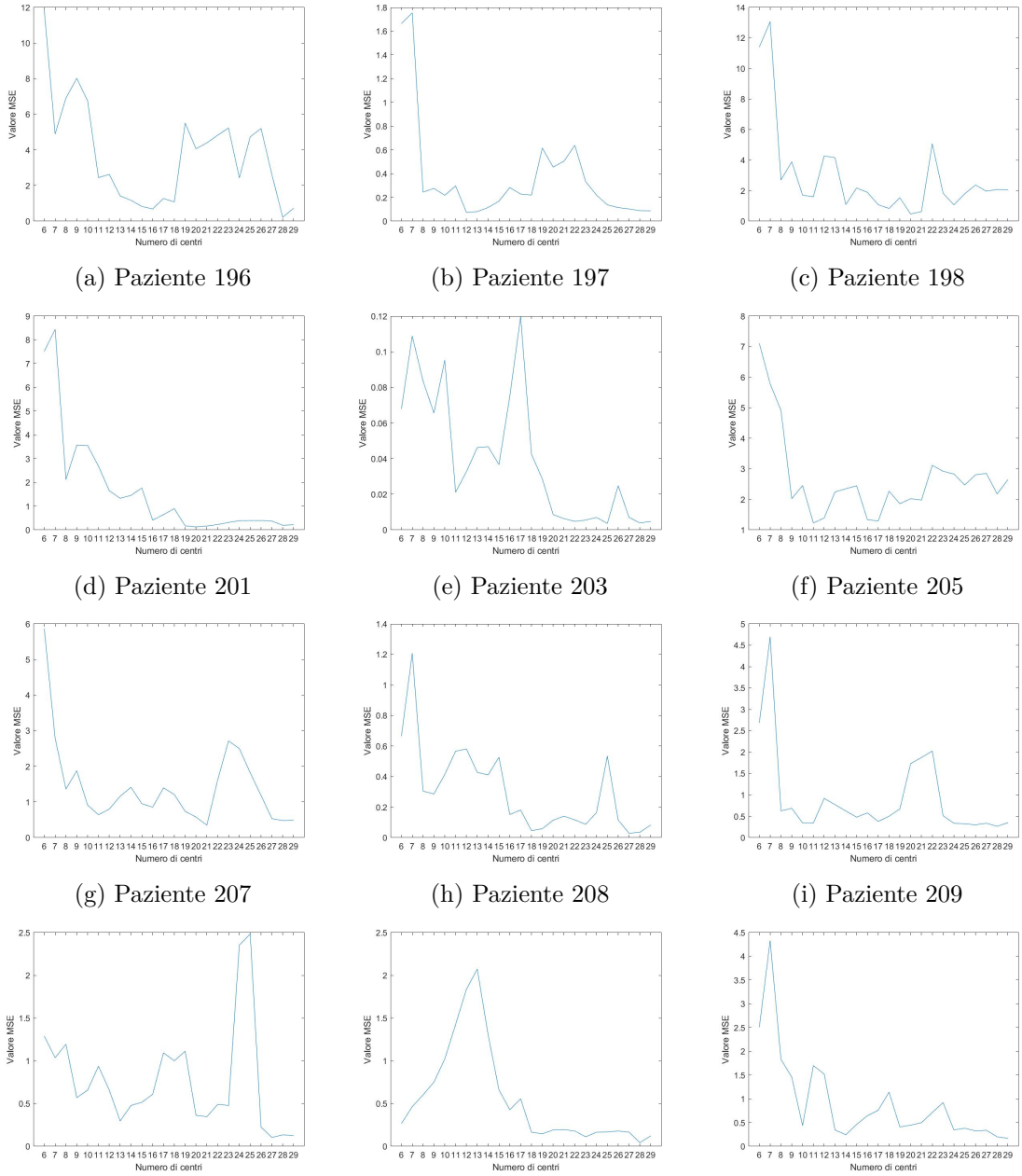
(a) Paziente 180 (b) Paziente 182 (c) Paziente 183

(d) Paziente 184 (e) Paziente 185 (f) Paziente 187

(g) Paziente 188 (h) Paziente 189 (i) Paziente 190

(j) Paziente 191 (k) Paziente 193 (l) Paziente 194

Figura A.9: Valori MSE sul validation set dei pazienti 180, 182, 183, 184, 185, 187, 188, 189, 190, 191, 193 e 194



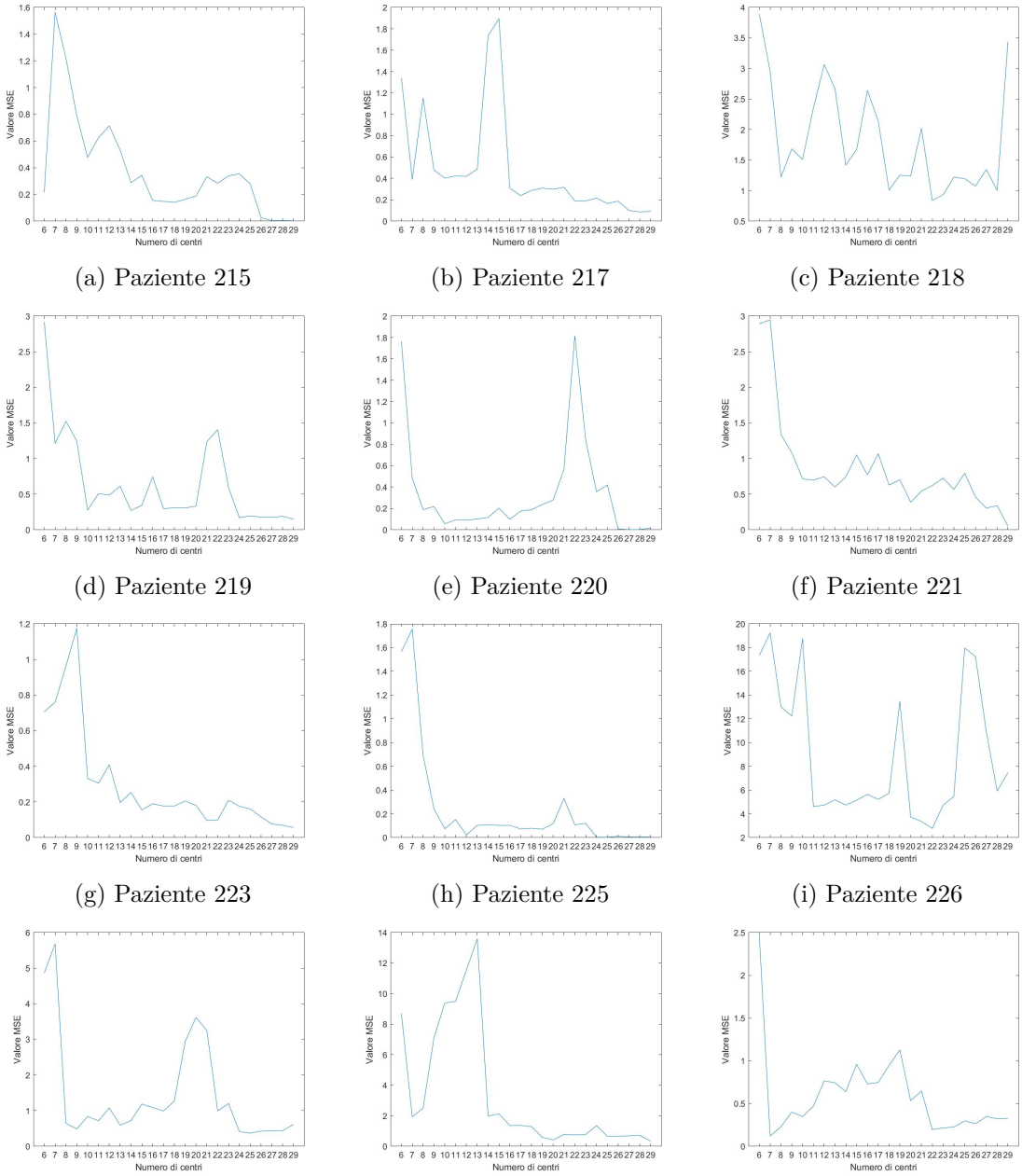
(a) Paziente 196 (b) Paziente 197 (c) Paziente 198

(d) Paziente 201 (e) Paziente 203 (f) Paziente 205

(g) Paziente 207 (h) Paziente 208 (i) Paziente 209

(j) Paziente 210 (k) Paziente 212 (l) Paziente 213

Figura A.10: Valori MSE sul validation set dei pazienti 196, 197, 198, 201, 203, 205, 207, 208, 209, 210, 212 e 213



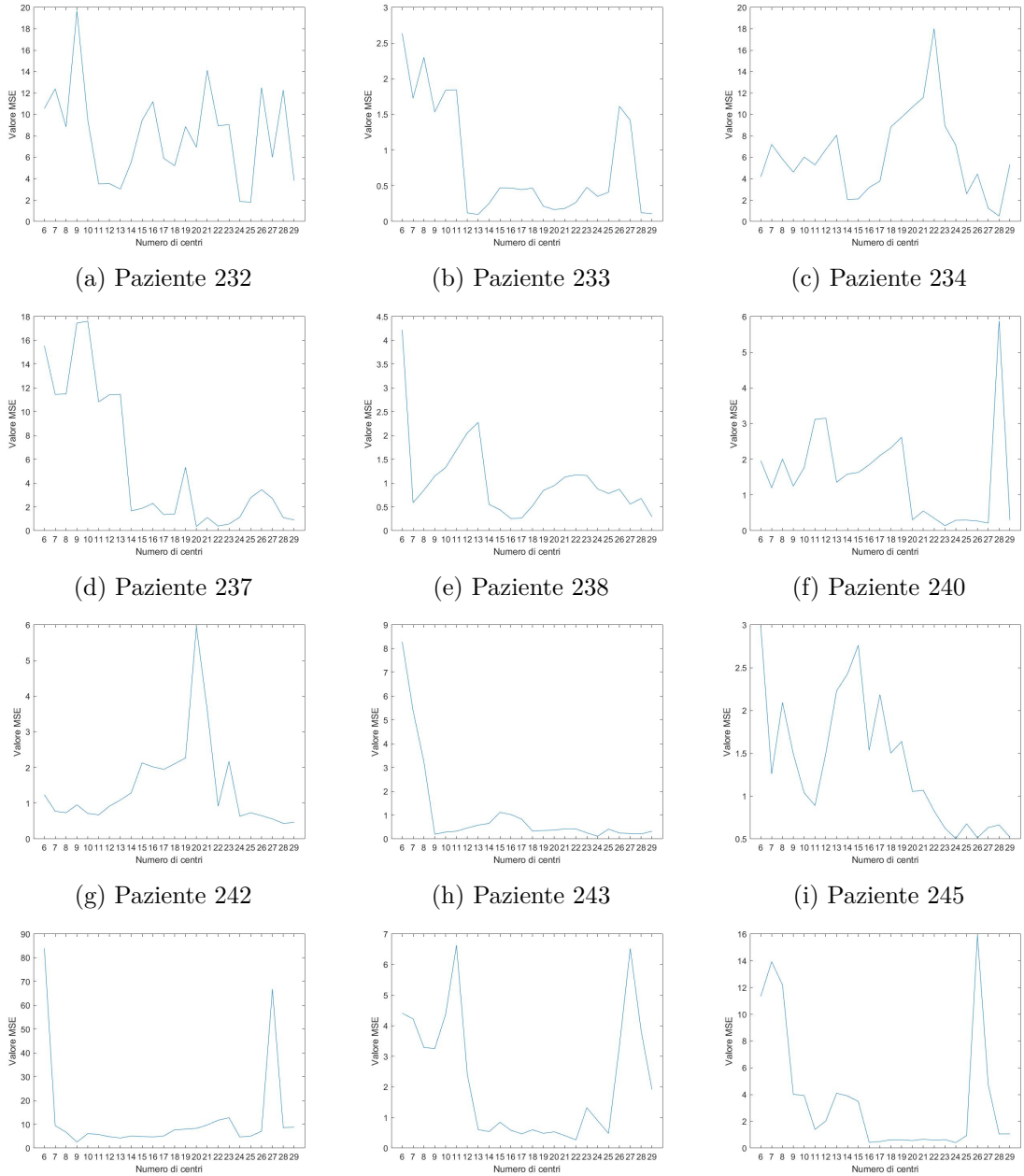
(a) Paziente 215 (b) Paziente 217 (c) Paziente 218

(d) Paziente 219 (e) Paziente 220 (f) Paziente 221

(g) Paziente 223 (h) Paziente 225 (i) Paziente 226

(j) Paziente 227 (k) Paziente 229 (l) Paziente 230

Figura A.11: Valori MSE sul validation set dei pazienti 215, 217, 218, 219, 220, 221, 223, 225, 226, 227, 229 e 230



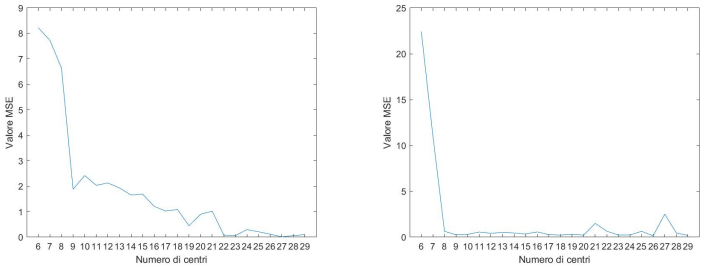
(a) Paziente 232 (b) Paziente 233 (c) Paziente 234

(d) Paziente 237 (e) Paziente 238 (f) Paziente 240

(g) Paziente 242 (h) Paziente 243 (i) Paziente 245

(j) Paziente 246 (k) Paziente 247 (l) Paziente 248

Figura A.12: Valori MSE sul validation set dei pazienti 232, 233, 234, 237, 238, 240, 242, 243, 245, 246, 247 e 248

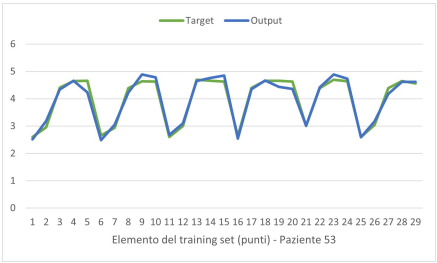


(a) Paziente 252 (b) Paziente 254

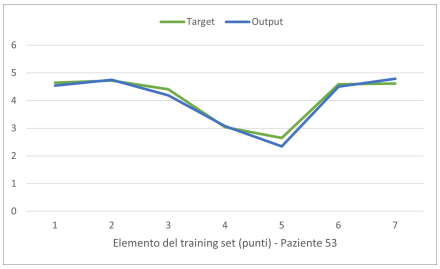
Figura A.13: Valori MSE sul validation set dei pazienti 252 e 254

Appendice B

Confronto tra target e output della rete con 15 centri

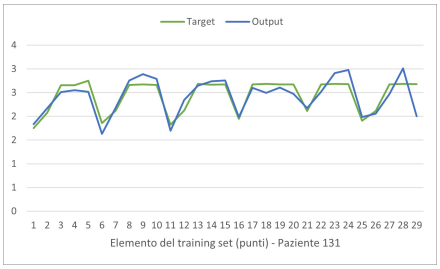


(a) Training set

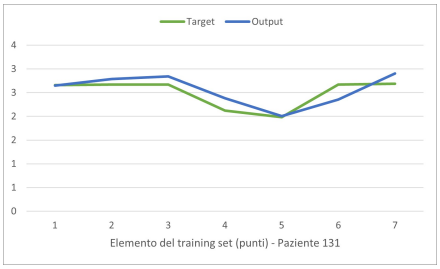


(b) Validation set

Figura B.1: Confronto target e valore di output restituito - paziente 53

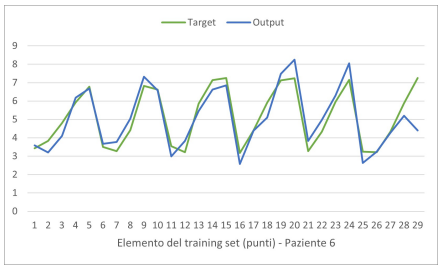


(a) Training set

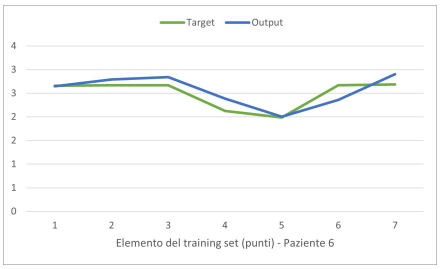


(b) Validation set

Figura B.2: Confronto target e valore di output restituito - paziente 131

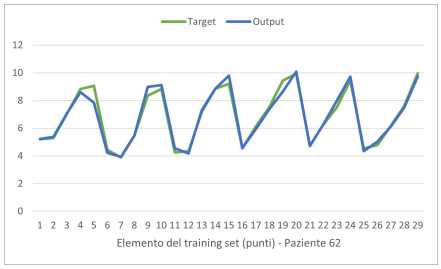


(a) Training set

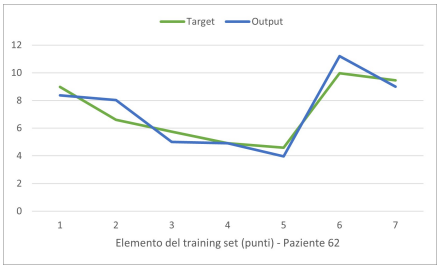


(b) Validation set

Figura B.3: Confronto target e valore di output restituito - paziente 6

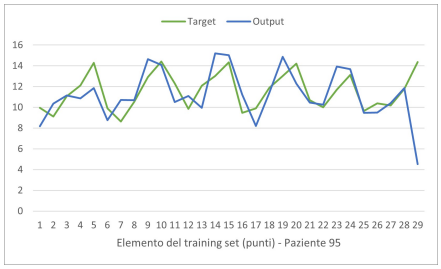


(a) Training set

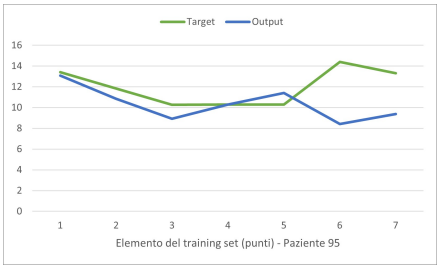


(b) Validation set

Figura B.4: Confronto target e valore di output restituito - paziente 62



(a) Training set

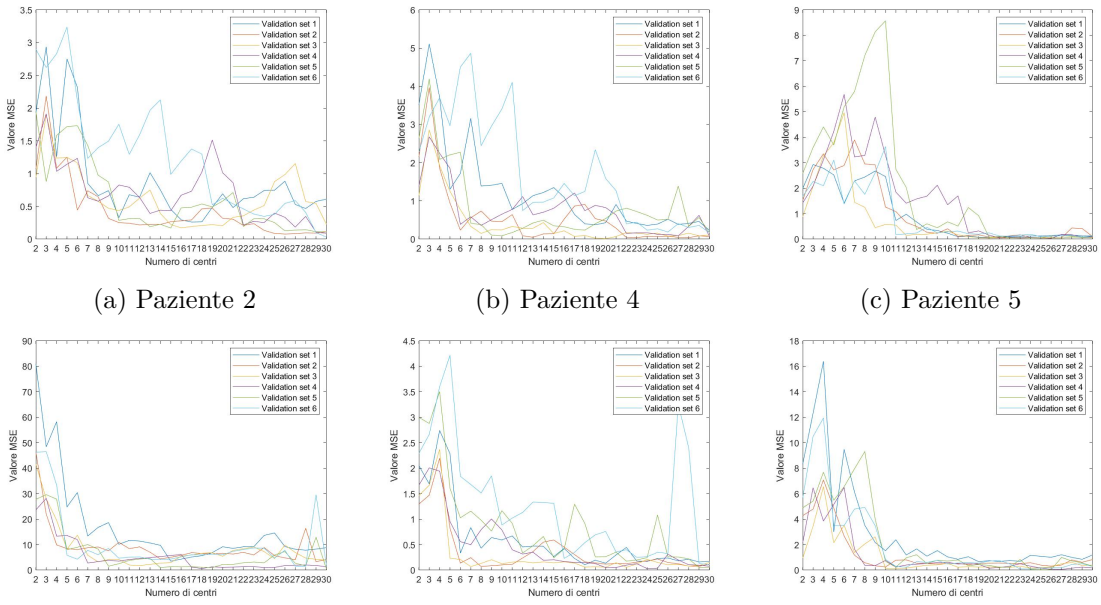


(b) Validation set

Figura B.5: Confronto target e valore di output restituito - paziente 95

Appendice C

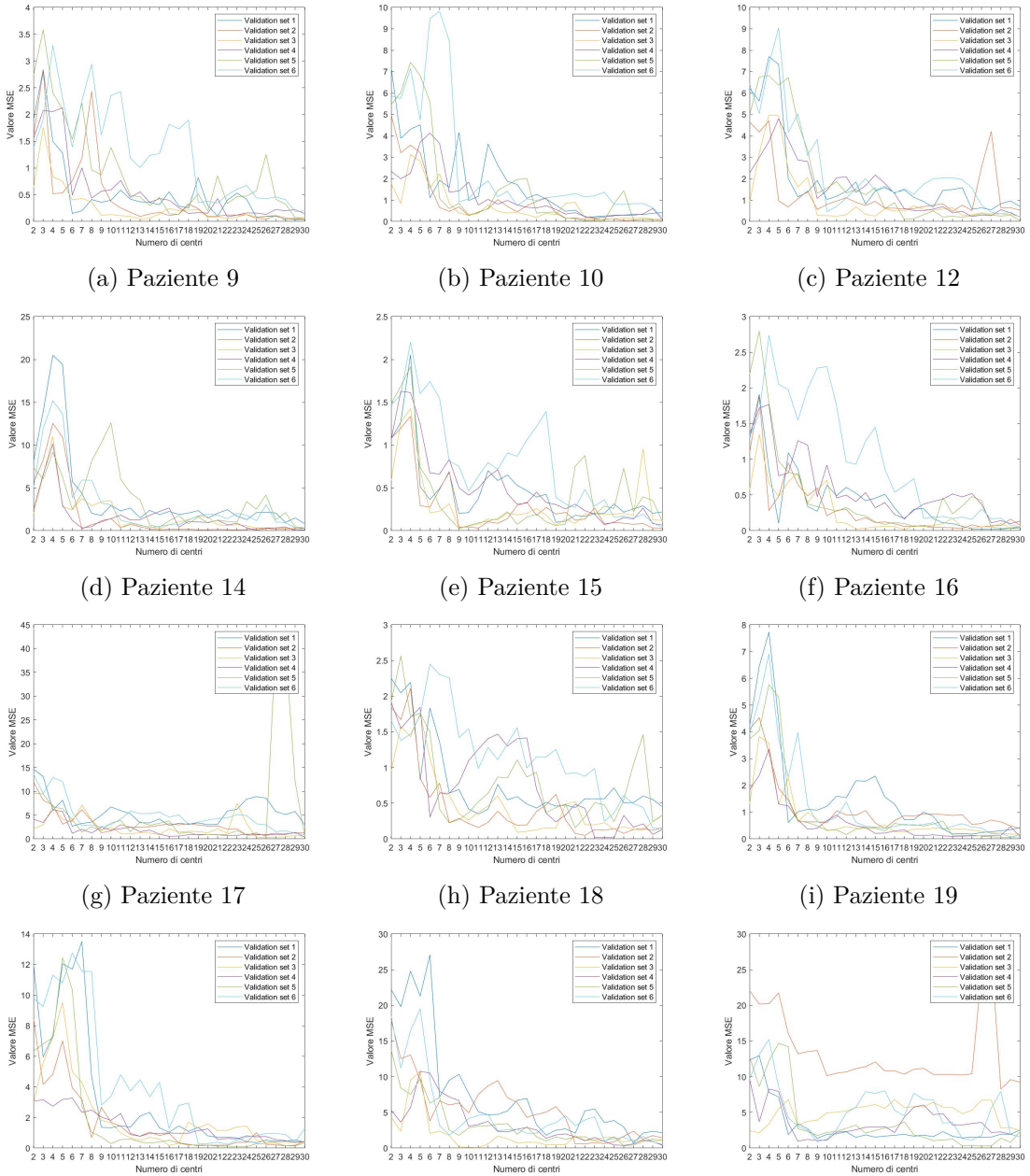
Valori MSE: allenamento RBFNN con il metodo k-fold



(a) Paziente 2 (b) Paziente 4 (c) Paziente 5

(d) Paziente 6 (e) Paziente 7 (f) Paziente 8

Figura C.1: Valori MSE sul validation set dei pazienti 2, 4, 5, 6, 7 e 8



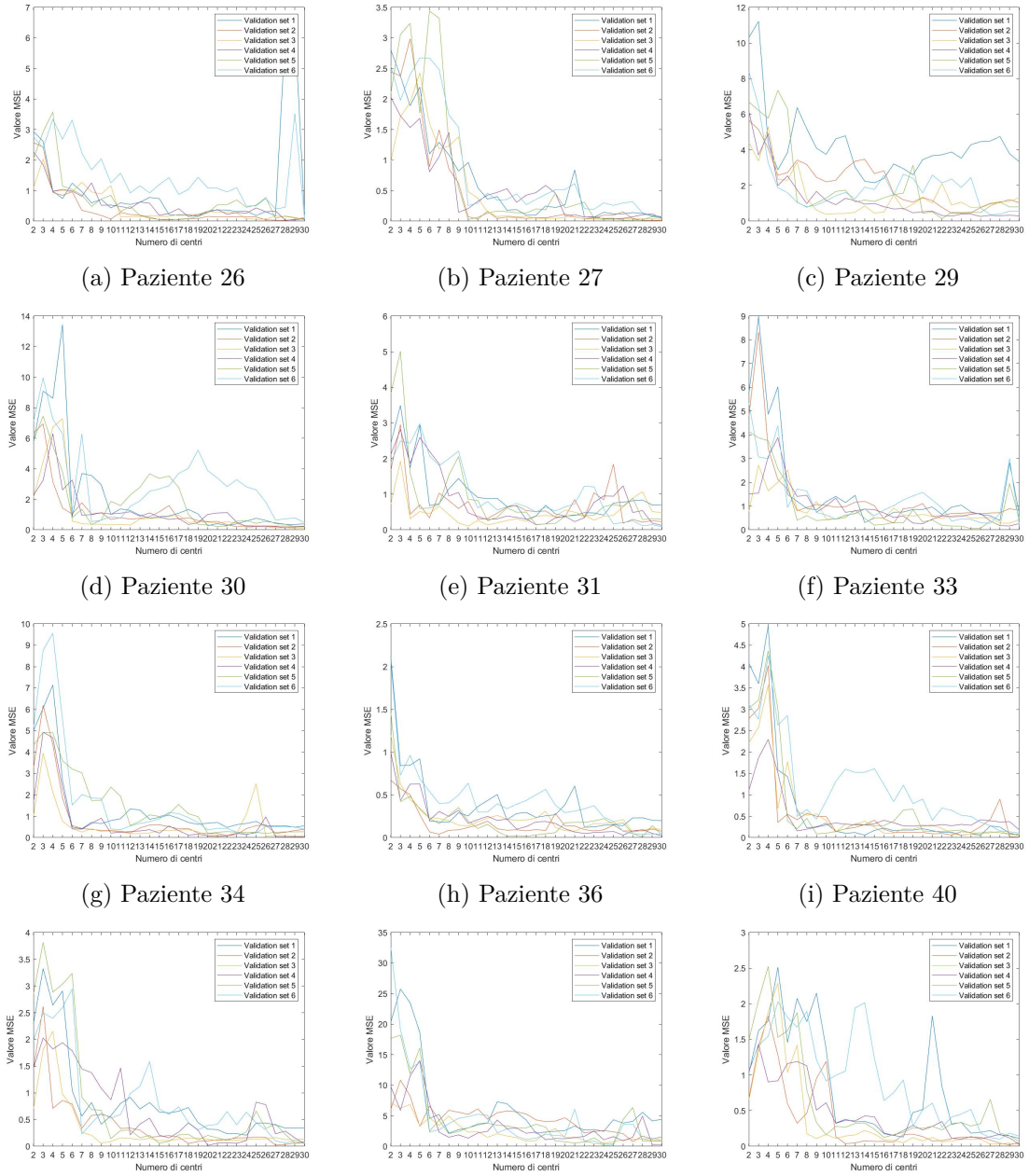
(a) Paziente 9 (b) Paziente 10 (c) Paziente 12

(d) Paziente 14 (e) Paziente 15 (f) Paziente 16

(g) Paziente 17 (h) Paziente 18 (i) Paziente 19

(j) Paziente 20 (k) Paziente 22 (l) Paziente 25

Figura C.2: Valori MSE sul validation set dei pazienti 9, 10, 12, 14, 16, 17, 18, 19, 20, 22 e 25



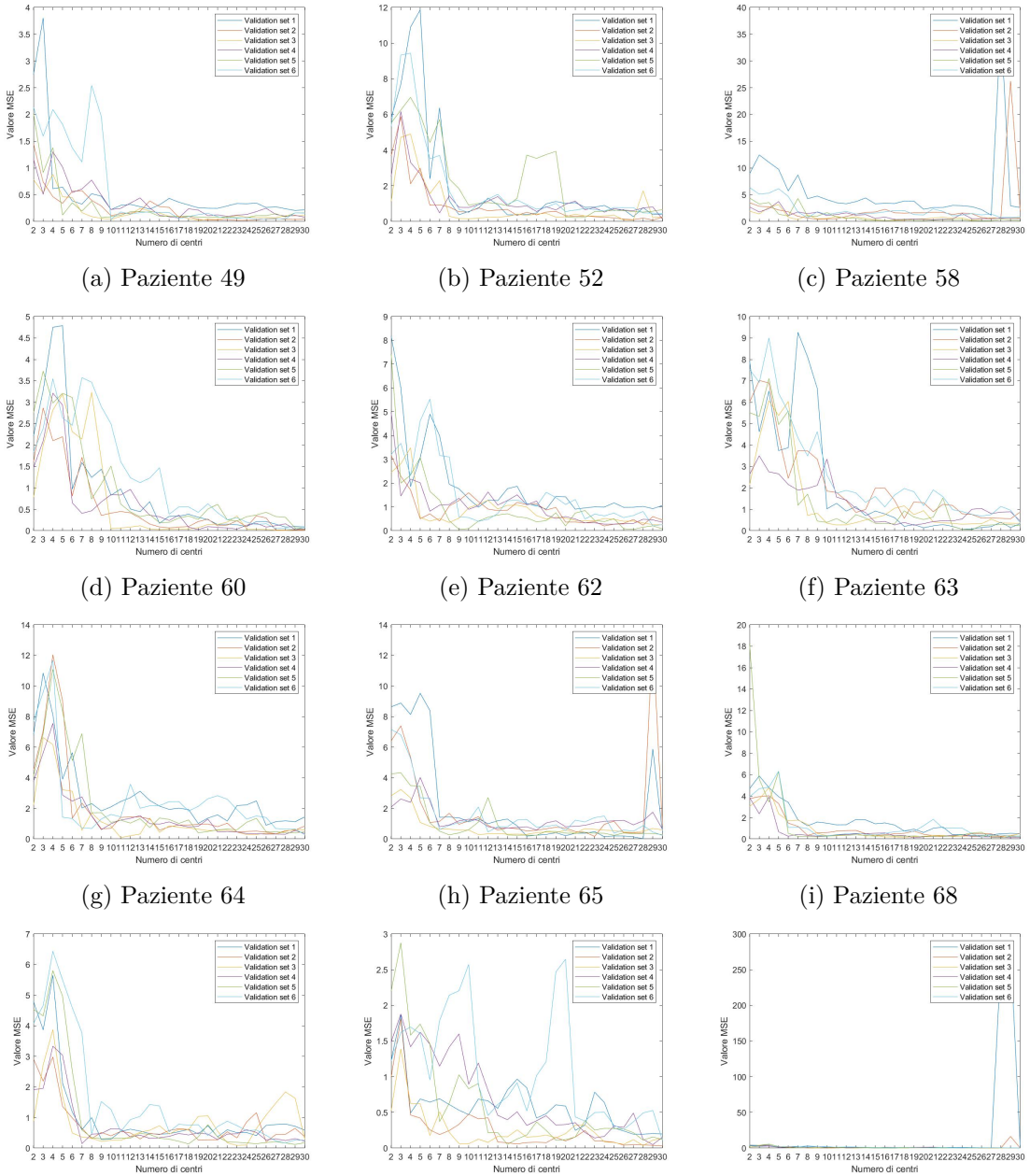
(a) Paziente 26 (b) Paziente 27 (c) Paziente 29

(d) Paziente 30 (e) Paziente 31 (f) Paziente 33

(g) Paziente 34 (h) Paziente 36 (i) Paziente 40

(j) Paziente 43 (k) Paziente 45 (l) Paziente 48

Figura C.3: Valori MSE sul validation set dei pazienti 26, 27, 29, 30, 31, 33, 34, 36, 40, 43, 45 e 48



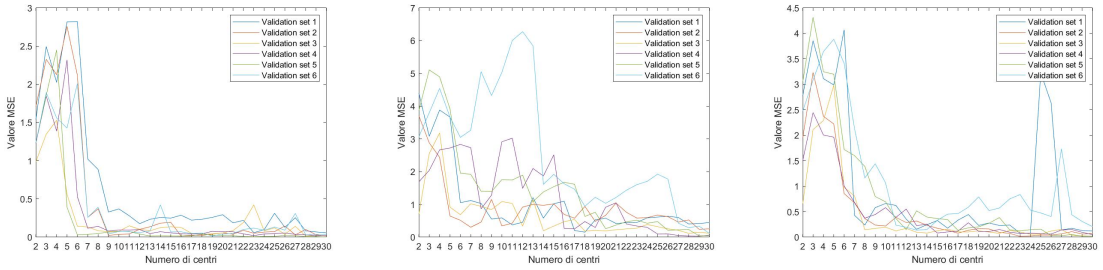
(a) Paziente 49 (b) Paziente 52 (c) Paziente 58

(d) Paziente 60 (e) Paziente 62 (f) Paziente 63

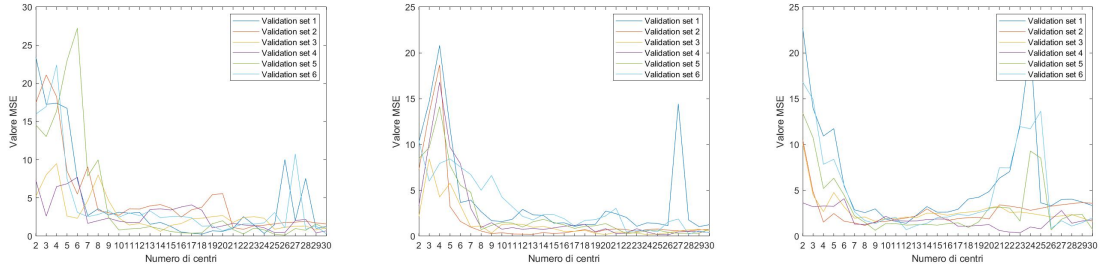
(g) Paziente 64 (h) Paziente 65 (i) Paziente 68

(j) Paziente 73 (k) Paziente 87 (l) Paziente 89

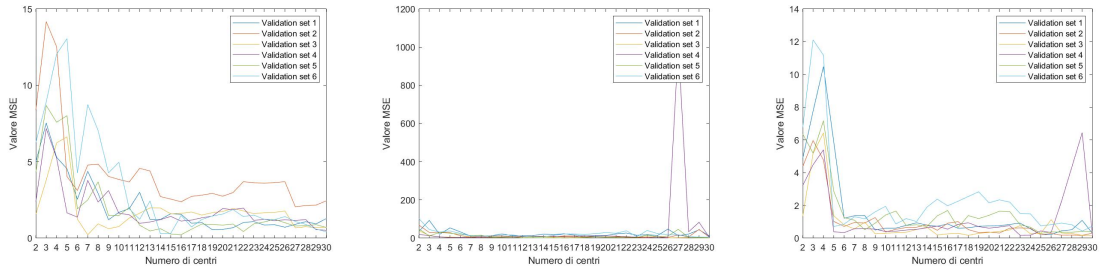
Figura C.4: Valori MSE sul validation set dei pazienti 49, 52, 58, 60, 62, 63, 64, 65, 68, 73, 87 e 89



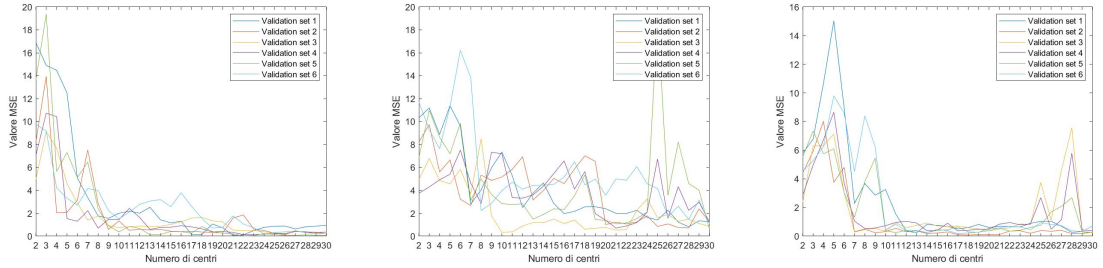
(a) Paziente 90 (b) Paziente 92 (c) Paziente 94



(d) Paziente 96 (e) Paziente 109 (f) Paziente 112

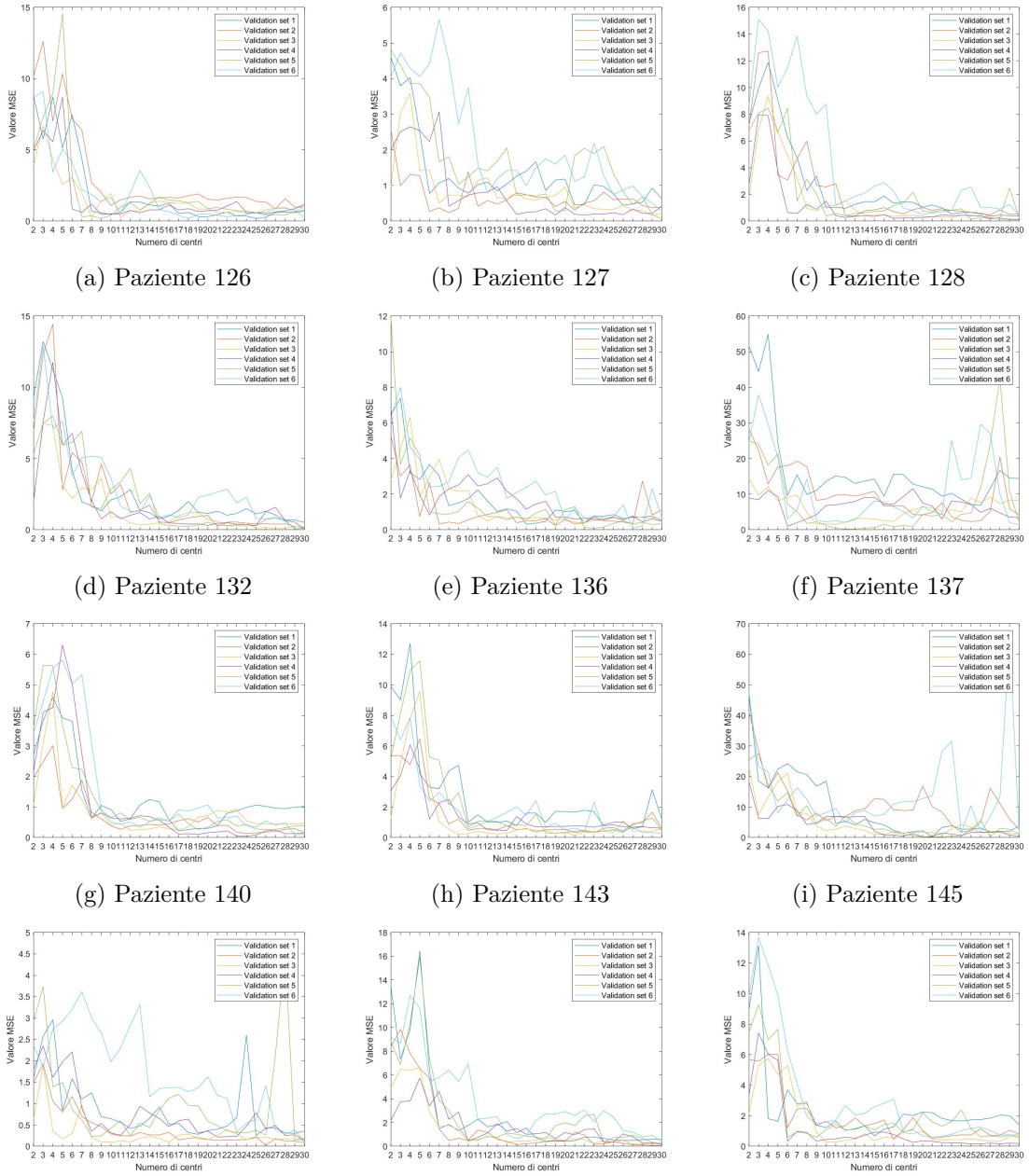


(g) Paziente 115 (h) Paziente 116 (i) Paziente 118



(j) Paziente 120 (k) Paziente 121 (l) Paziente 122

Figura C.5: Valori MSE sul validation set dei pazienti 90, 92, 94, 96, 109, 112, 115, 116, 118, 120, 121 e 122



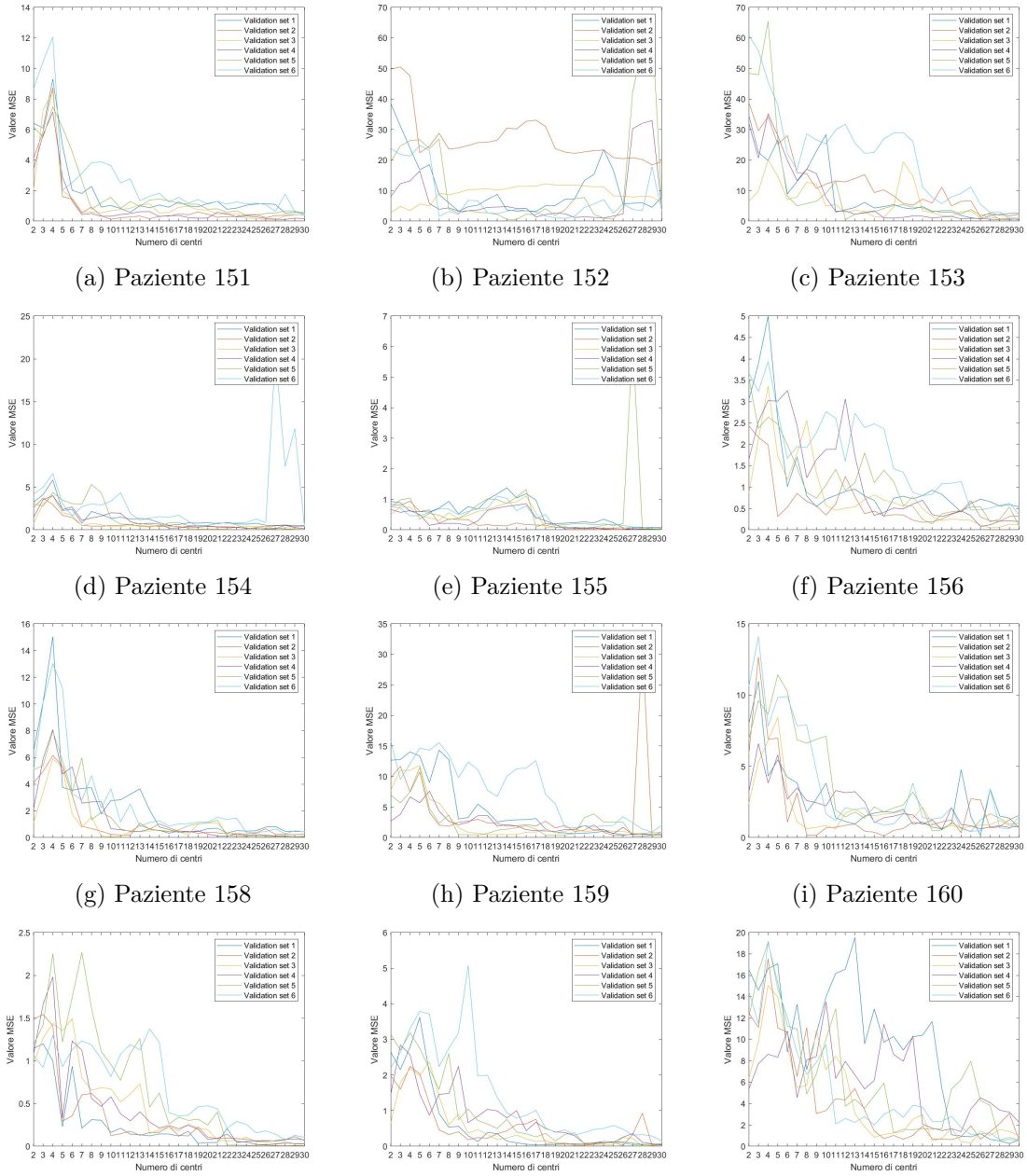
(a) Paziente 126 (b) Paziente 127 (c) Paziente 128

(d) Paziente 132 (e) Paziente 136 (f) Paziente 137

(g) Paziente 140 (h) Paziente 143 (i) Paziente 145

(j) Paziente 146 (k) Paziente 147 (l) Paziente 150

Figura C.6: Valori MSE sul validation set dei pazienti 126, 127, 128, 132, 136, 137, 140, 143, 145, 146, 147 e 150



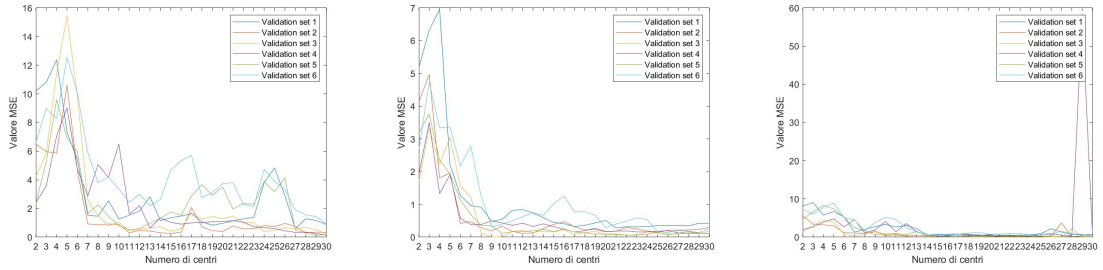
(a) Paziente 151 (b) Paziente 152 (c) Paziente 153

(d) Paziente 154 (e) Paziente 155 (f) Paziente 156

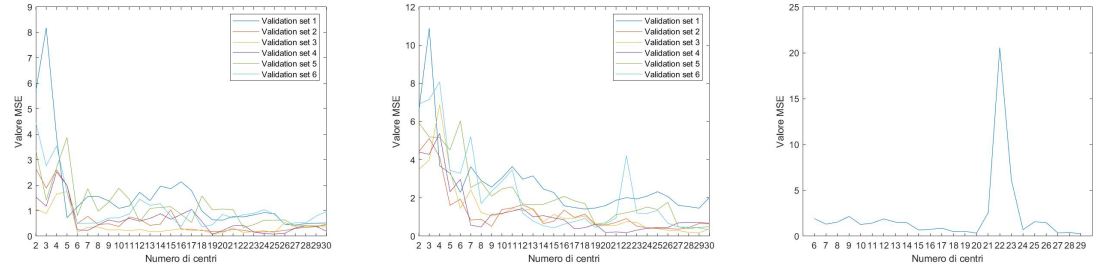
(g) Paziente 158 (h) Paziente 159 (i) Paziente 160

(j) Paziente 162 (k) Paziente 163 (l) Paziente 166

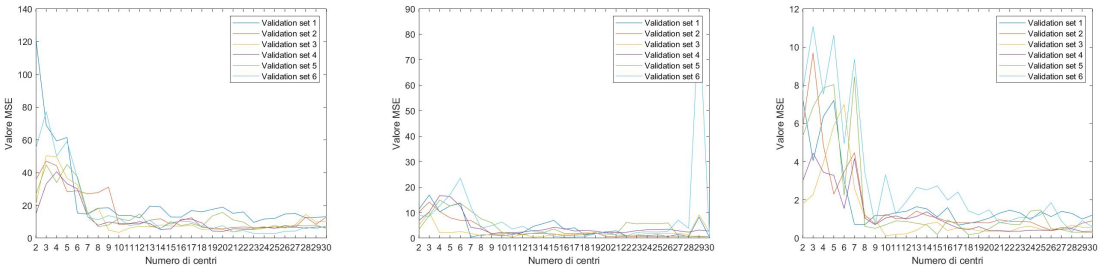
Figura C.7: Valori MSE sul validation set dei pazienti 151, 152, 153, 154, 155, 156, 158, 159, 160, 162, 163 e 166



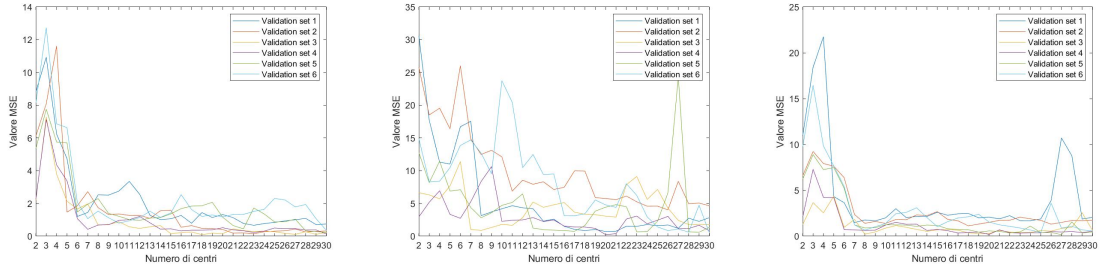
(a) Paziente 167 (b) Paziente 168 (c) Paziente 169



(d) Paziente 170 (e) Paziente 171 (f) Paziente 172

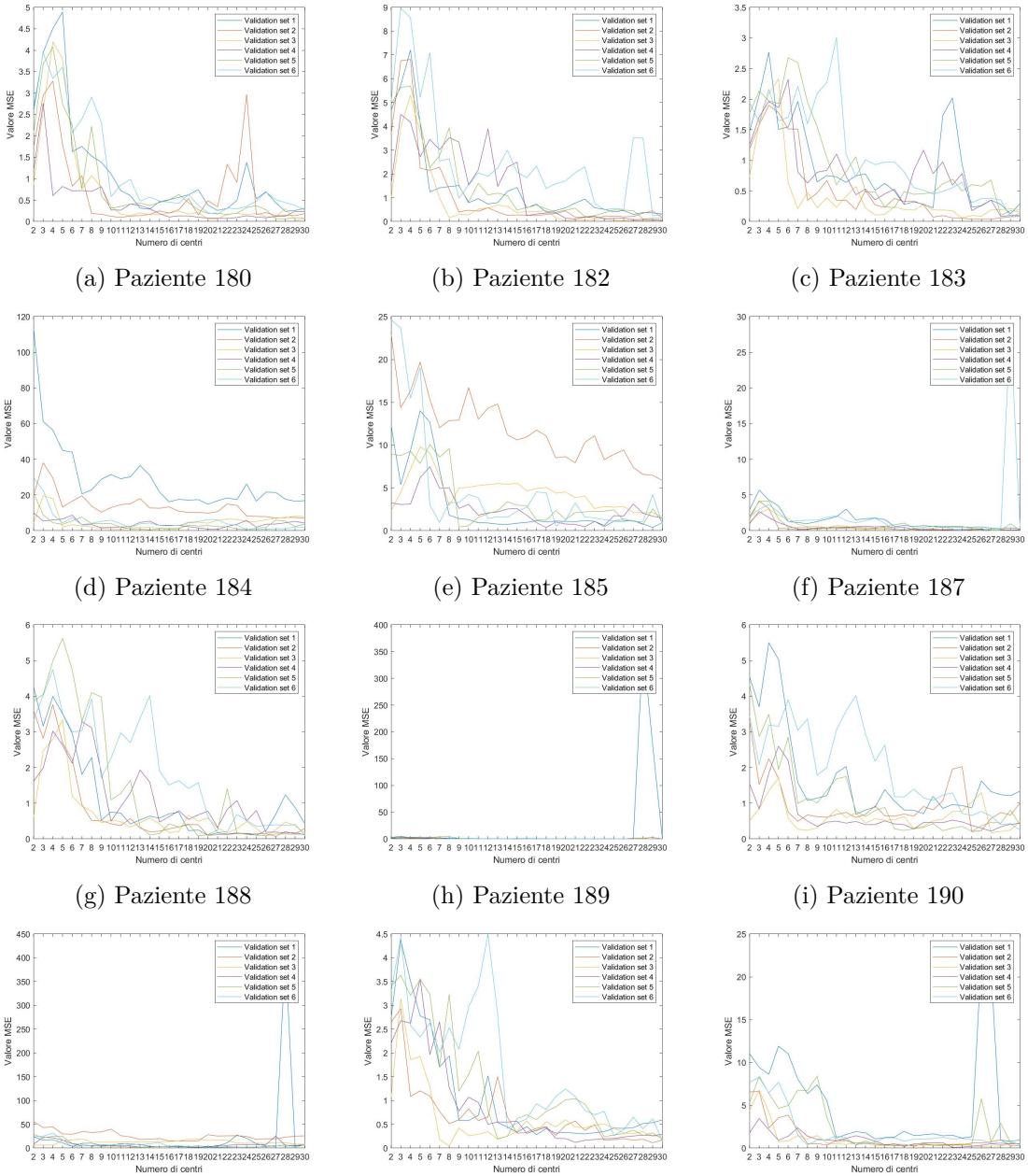


(g) Paziente 173 (h) Paziente 174 (i) Paziente 175



(j) Paziente 176 (k) Paziente 177 (l) Paziente 179

Figura C.8: Valori MSE sul validation set dei pazienti 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177 e 179



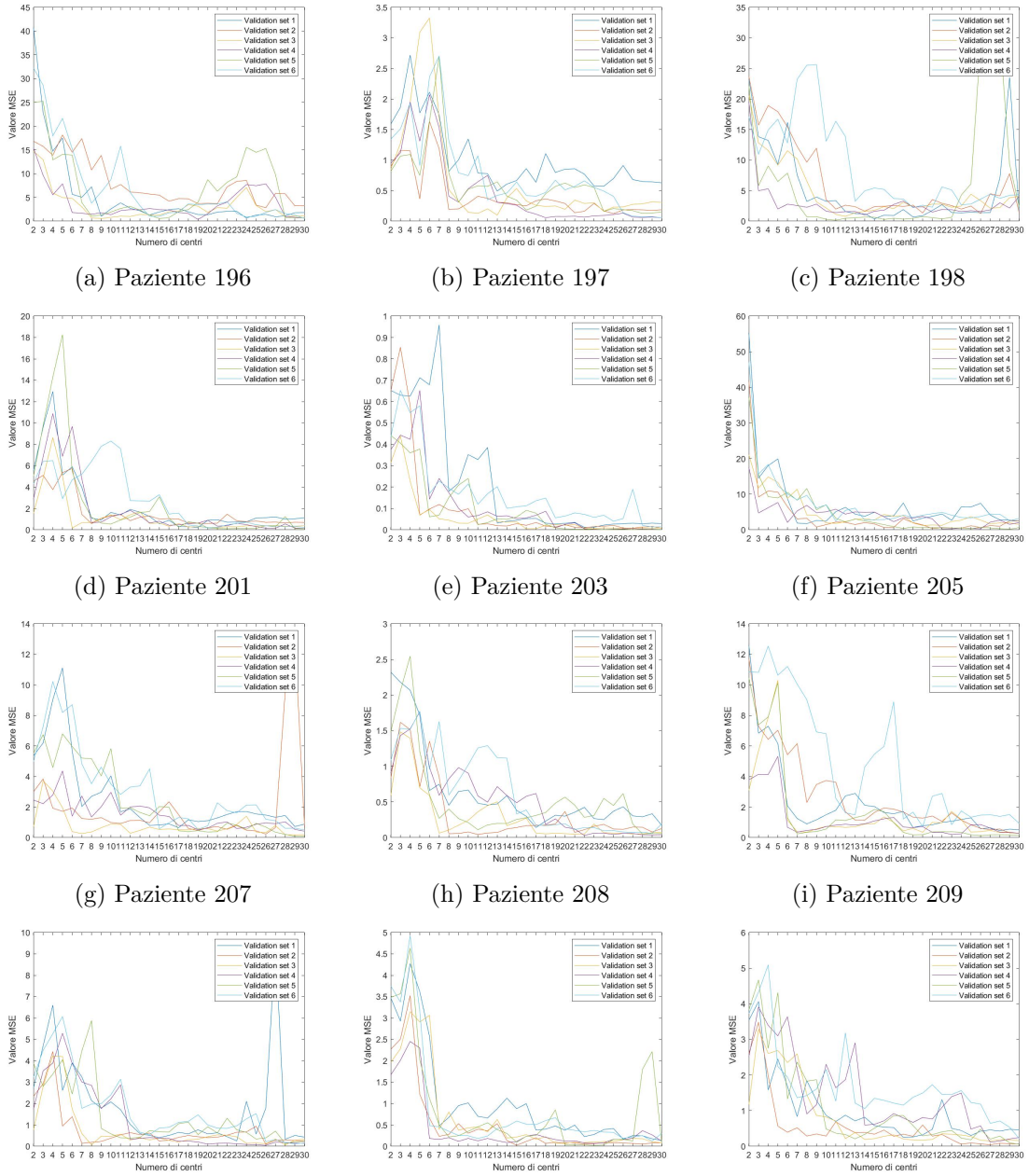
(a) Paziente 180 (b) Paziente 182 (c) Paziente 183

(d) Paziente 184 (e) Paziente 185 (f) Paziente 187

(g) Paziente 188 (h) Paziente 189 (i) Paziente 190

(j) Paziente 191 (k) Paziente 193 (l) Paziente 194

Figura C.9: Valori MSE sul validation set dei pazienti 180, 182, 183, 184, 185, 187, 188, 189, 190, 191, 193 e 194



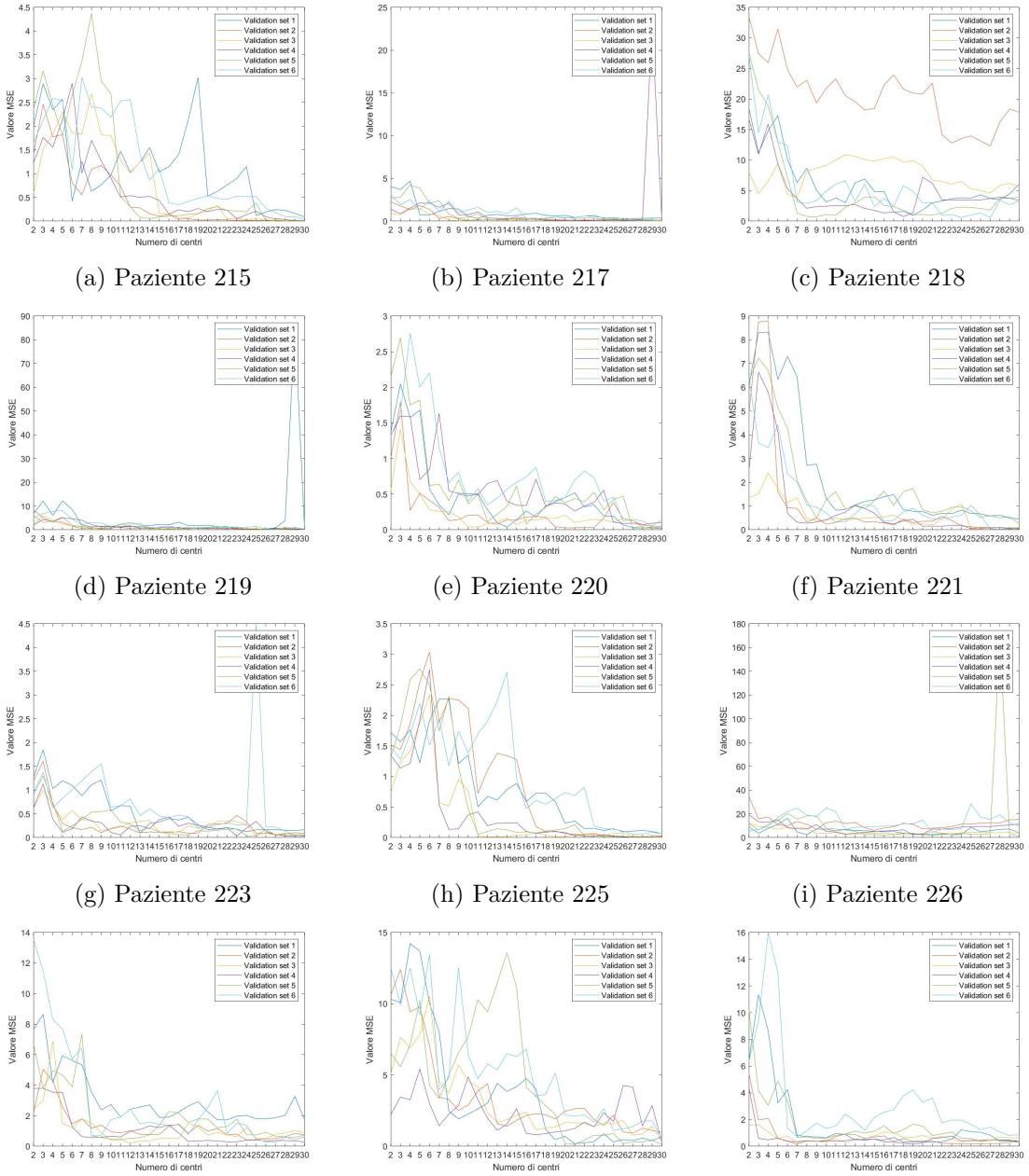
(a) Paziente 196 (b) Paziente 197 (c) Paziente 198

(d) Paziente 201 (e) Paziente 203 (f) Paziente 205

(g) Paziente 207 (h) Paziente 208 (i) Paziente 209

(j) Paziente 210 (k) Paziente 212 (l) Paziente 213

Figura C.10: Valori MSE sul validation set dei pazienti 196, 197, 198, 201, 203, 205, 207, 208, 209, 210, 212 e 213



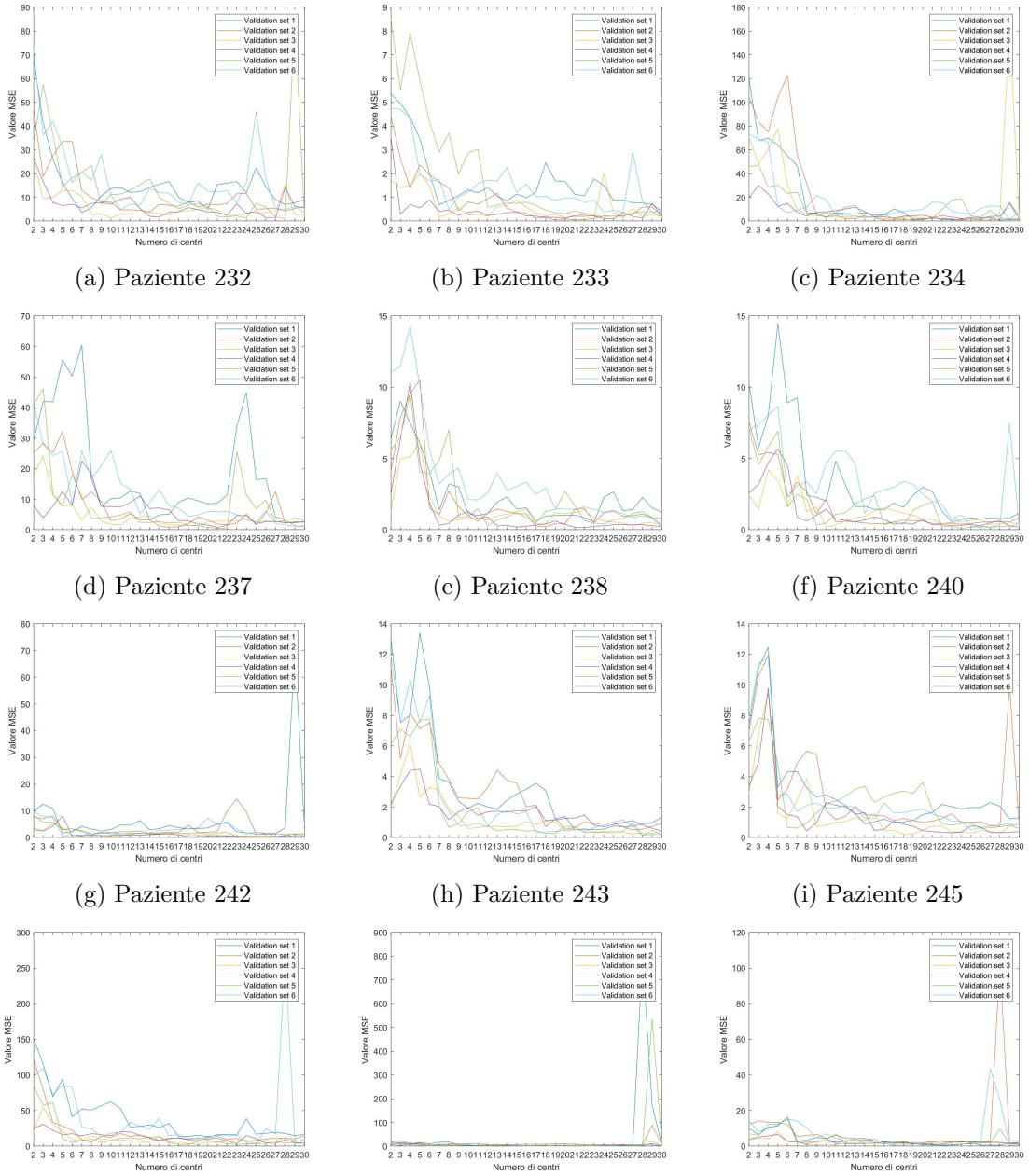
(a) Paziente 215 (b) Paziente 217 (c) Paziente 218

(d) Paziente 219 (e) Paziente 220 (f) Paziente 221

(g) Paziente 223 (h) Paziente 225 (i) Paziente 226

(j) Paziente 227 (k) Paziente 229 (l) Paziente 230

Figura C.11: Valori MSE sul validation set dei pazienti 215, 217, 218, 219, 220, 221, 223, 225, 226, 227, 229 e 230



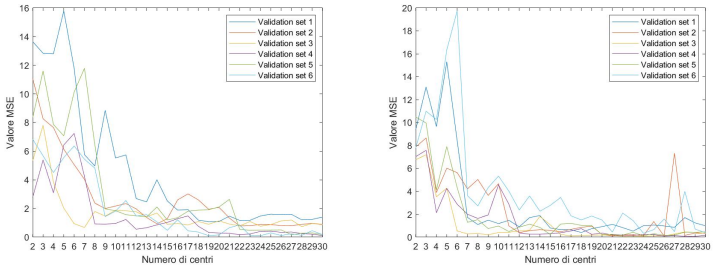
(a) Paziente 232 (b) Paziente 233 (c) Paziente 234

(d) Paziente 237 (e) Paziente 238 (f) Paziente 240

(g) Paziente 242 (h) Paziente 243 (i) Paziente 245

(j) Paziente 246 (k) Paziente 247 (l) Paziente 248

Figura C.12: Valori MSE sul validation set dei pazienti 232, 233, 234, 237, 238, 240, 242, 243, 245, 246, 247 e 248

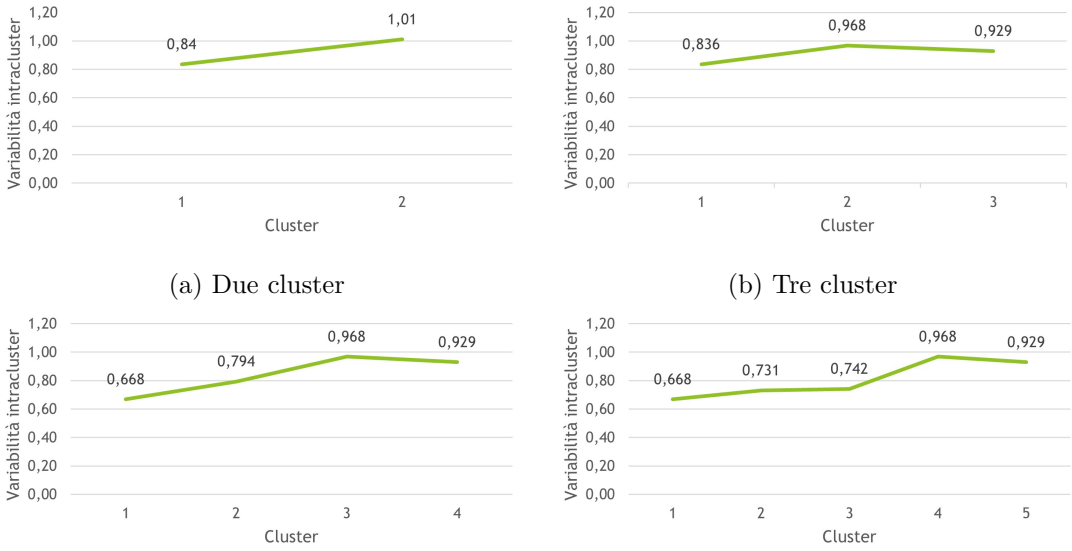


(a) Paziente 252 (b) Paziente 254

Figura C.13: Valori MSE sul validation set dei pazienti 252 e 254

Appendice D

Risultati delle analisi di clustering



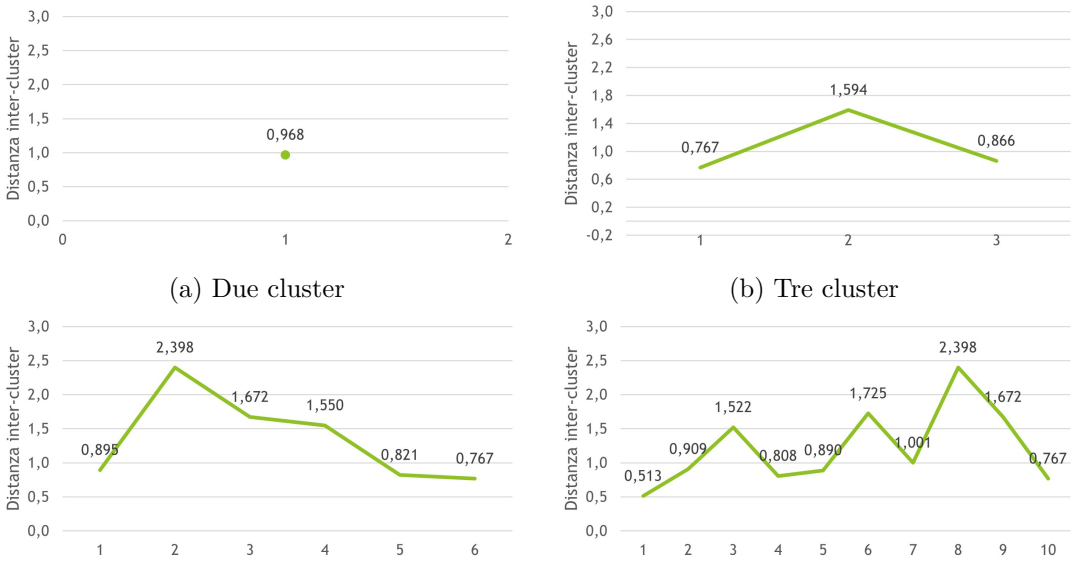
(a) Due cluster

(b) Tre cluster

(c) Quattro cluster

(d) Cinque cluster

Figura D.1: Variabilità intra-cluster



(a) Due cluster

(b) Tre cluster

(c) Quattro cluster

(d) Cinque cluster

Figura D.2: Distanze inter-cluster

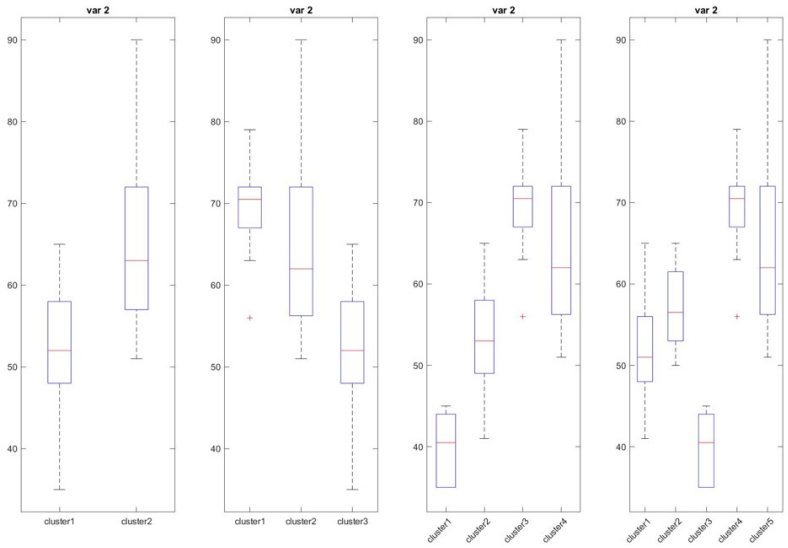


Figura D.3: Variabile 2

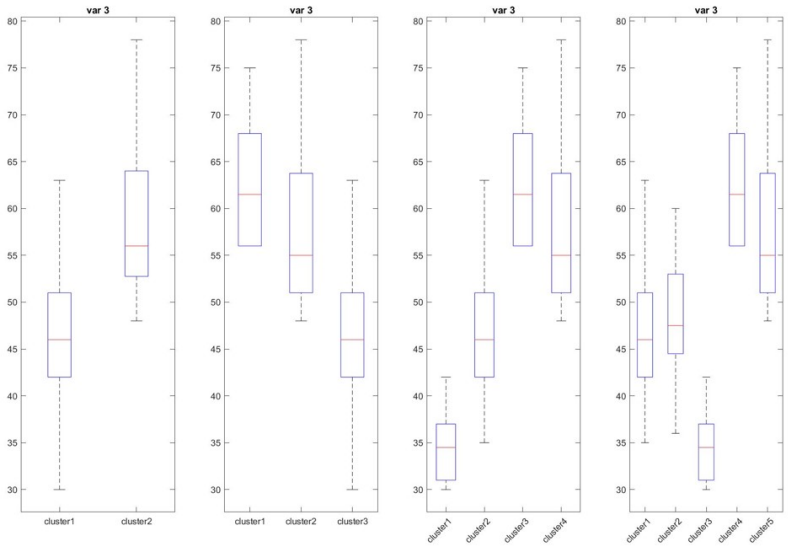


Figura D.4: Variabile 3

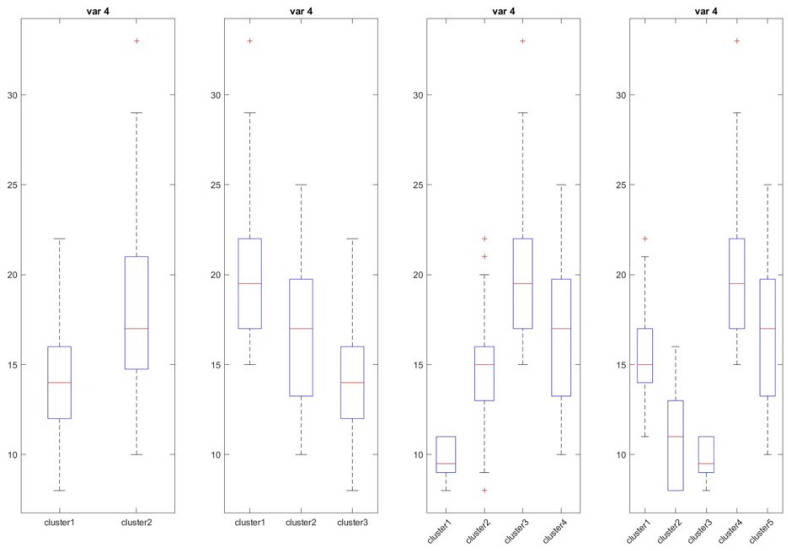


Figura D.5: Variabile 4

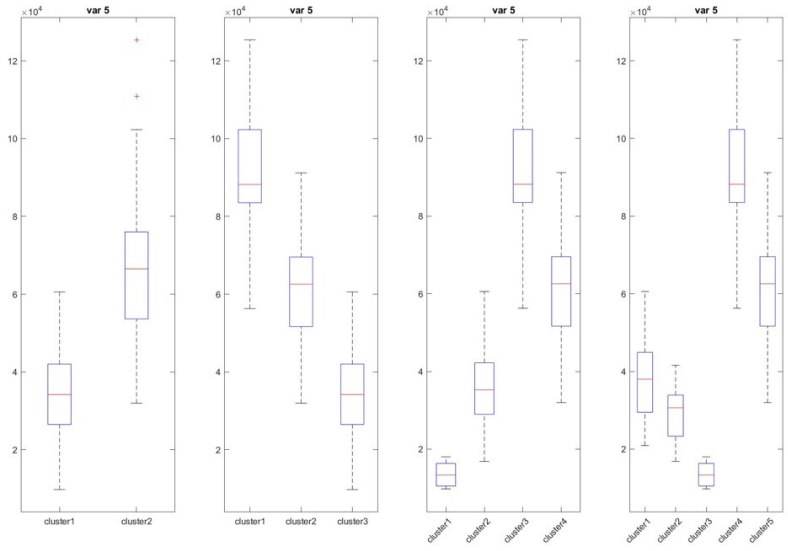


Figura D.6: Variabile 5

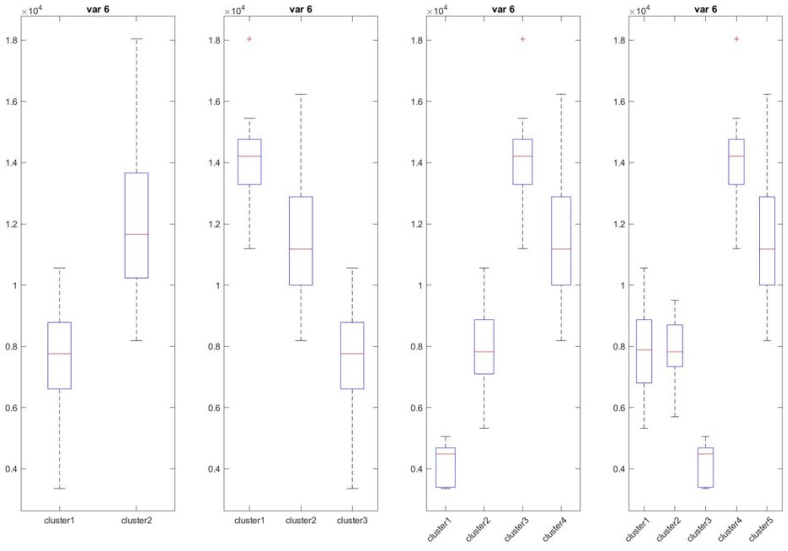


Figura D.7: Variabile 6

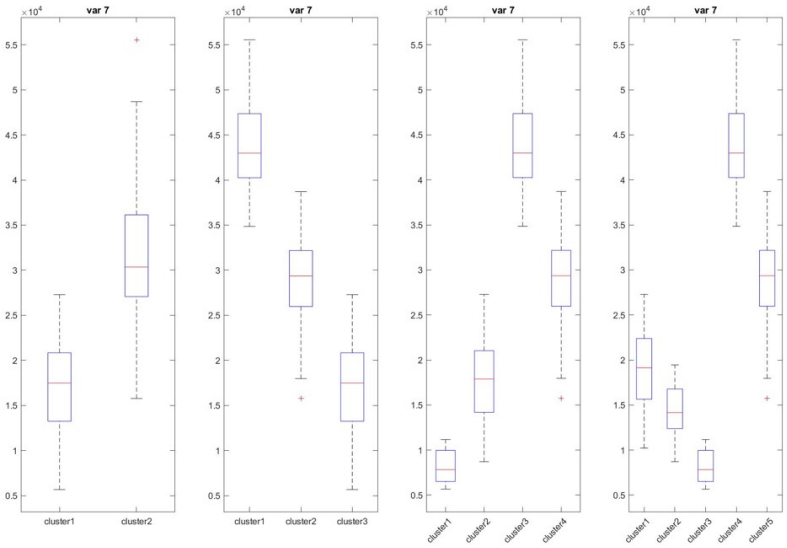


Figura D.8: Variabile 7

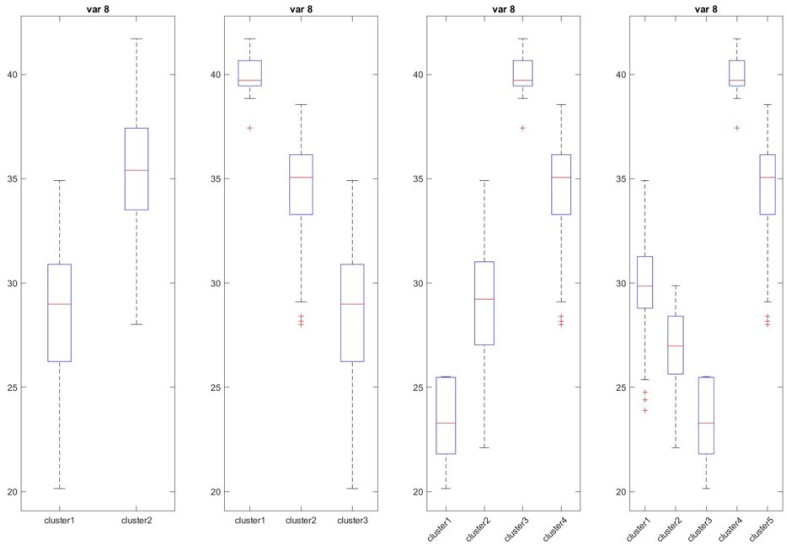


Figura D.9: Variabile 8

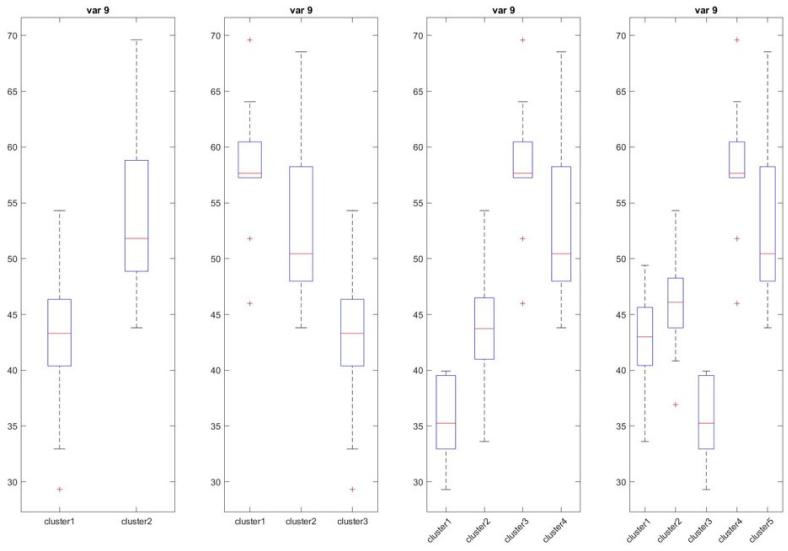


Figura D.10: Variabile 9

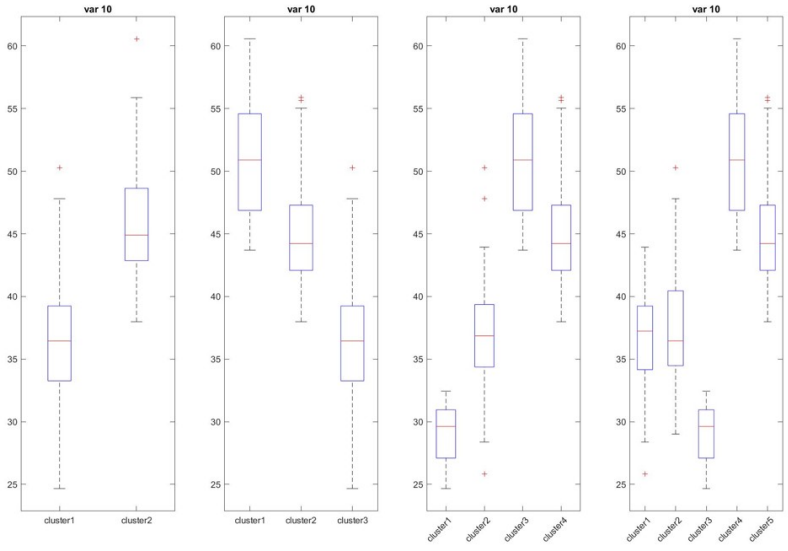


Figura D.11: Variabile 10

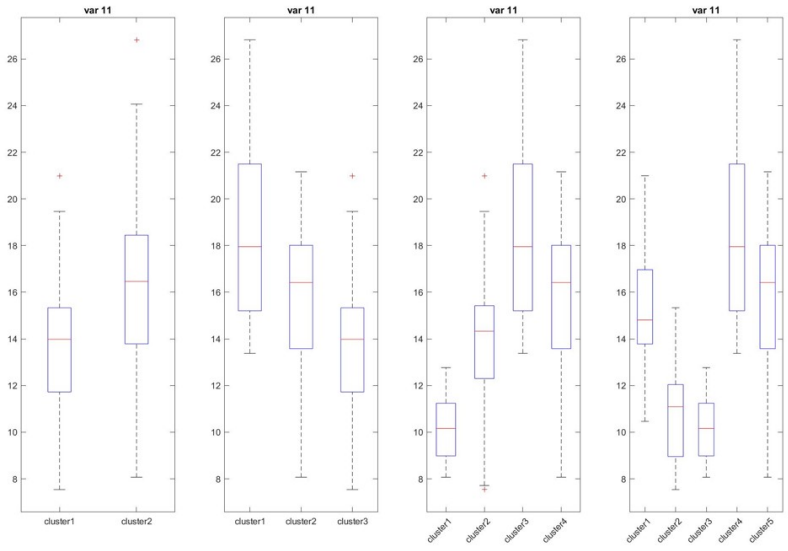


Figura D.12: Variabile 11

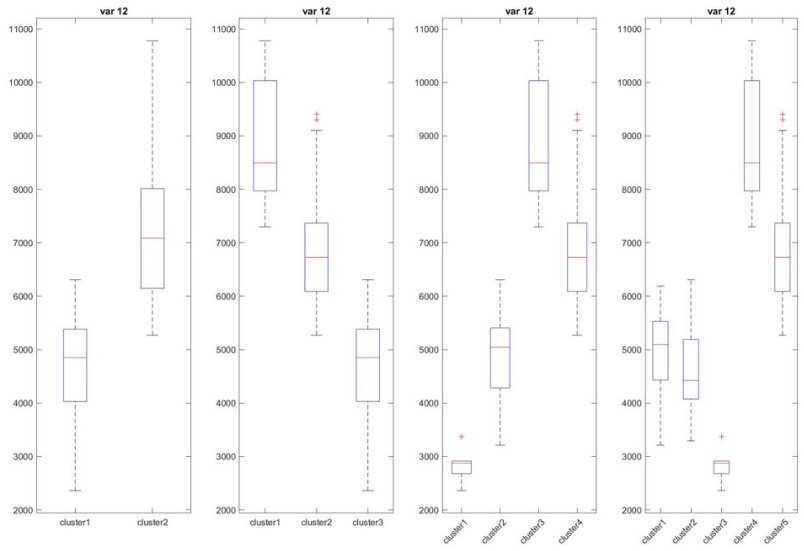


Figura D.13: Variabile 12

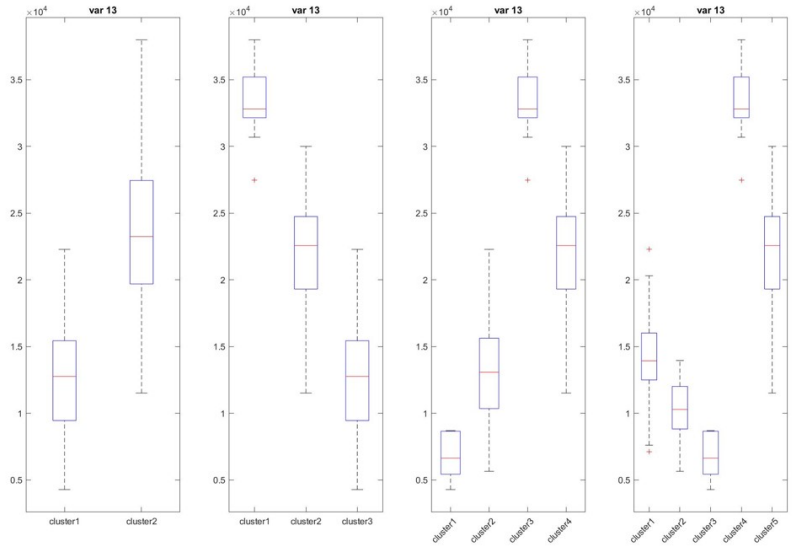


Figura D.14: Variabile 13

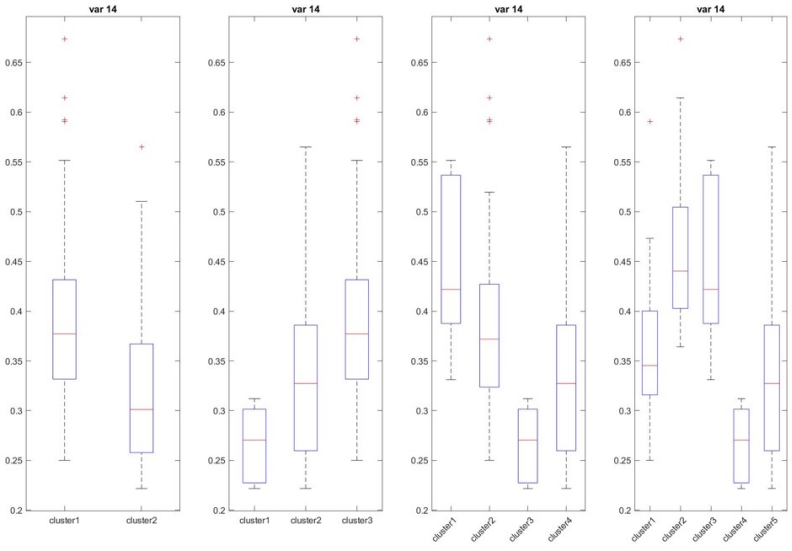


Figura D.15: Variabile 14

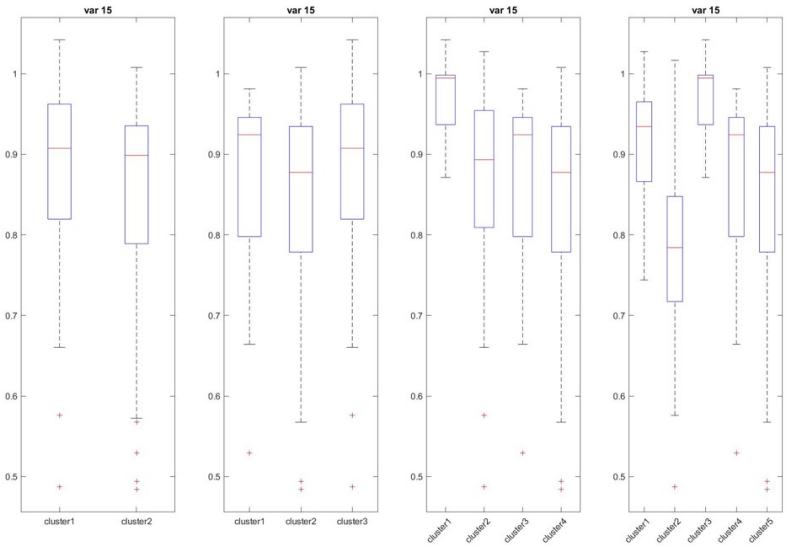


Figura D.16: Variabile 15

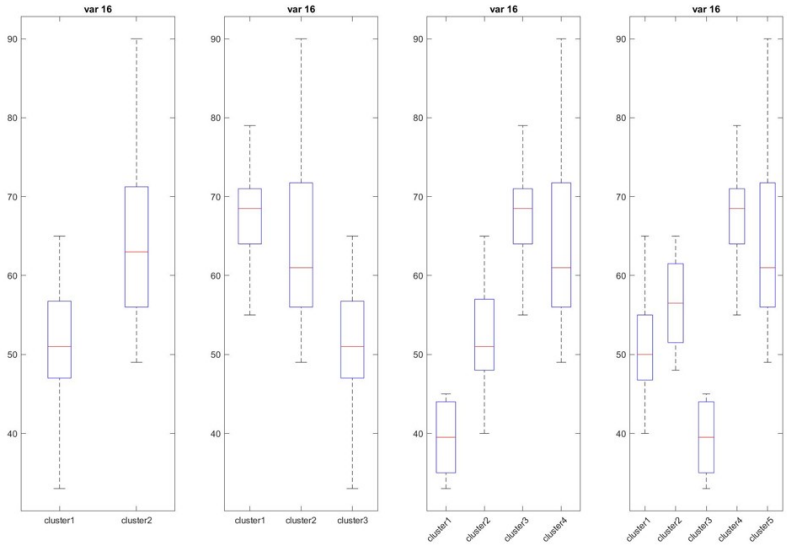


Figura D.17: Variabile 16

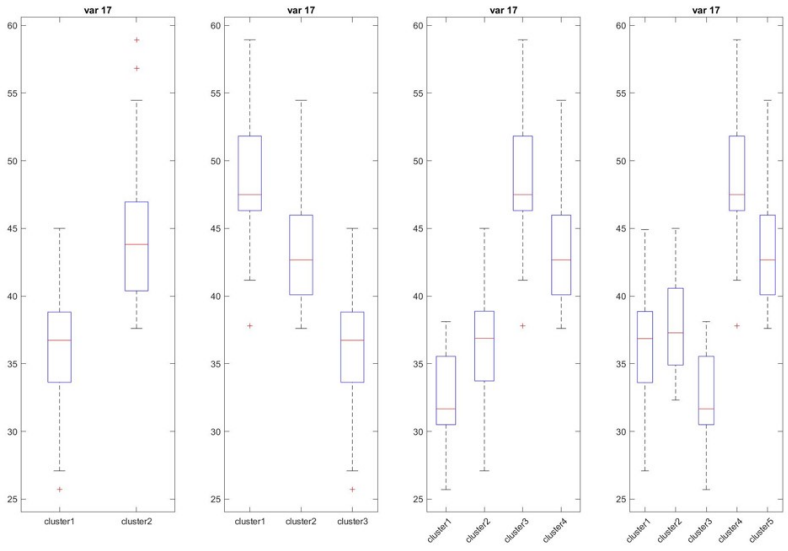


Figura D.18: Variabile 17

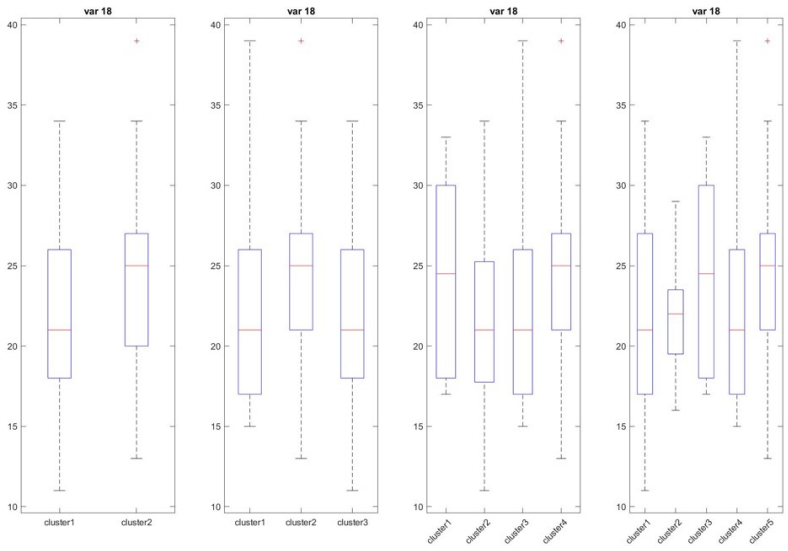


Figura D.19: Variabile 18

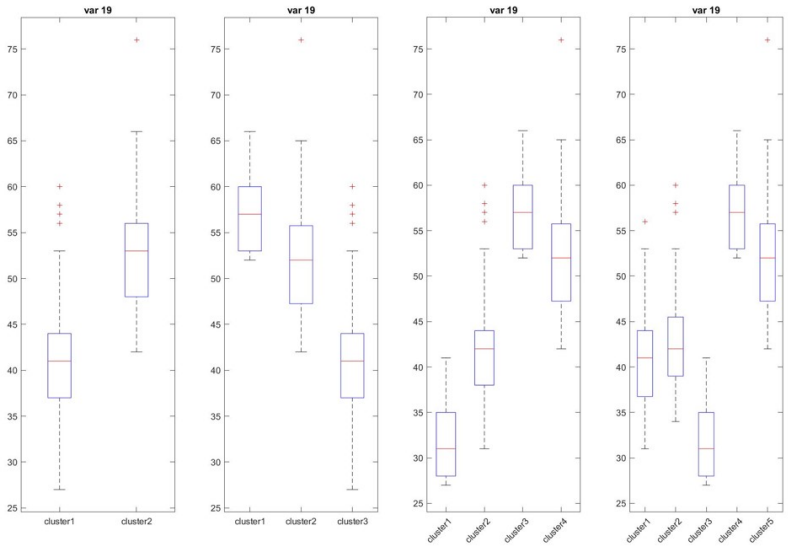


Figura D.20: Variabile 19

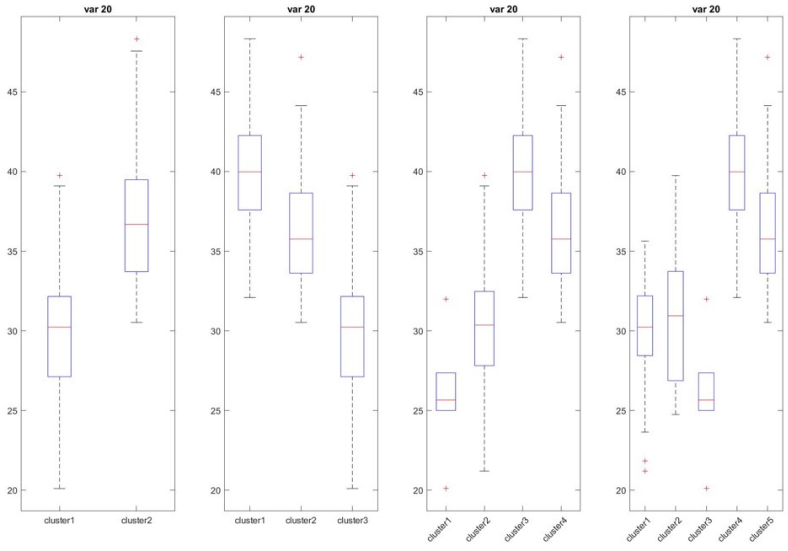


Figura D.21: Variabile 20

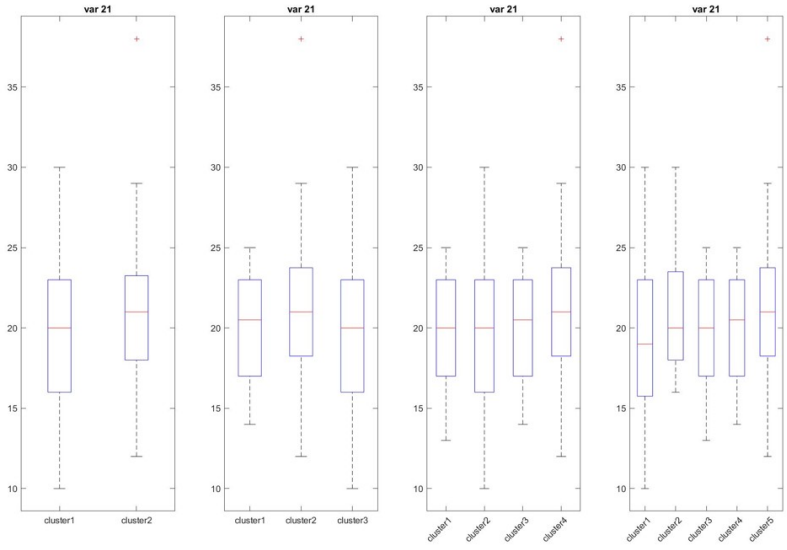


Figura D.22: Variabile 21

Appendice E

Assegnazione dei pazienti ai cluster

ID	Cluster assegnato
2	1
4	1
5	1
6	1
7	1
8	1
9	1
10	2
12	1
14	2
15	2
16	1
17	1
18	1
19	2
20	1
22	1
25	2
26	2
27	1

ID	Cluster assegnato
29	2
30	2
31	2
33	2
34	1
36	1
40	1
43	2
45	2
48	1
49	2
52	2
58	2
60	2
62	1
63	1
64	2
65	2
68	2
73	1

ID	Cluster assegnato
87	1
89	2
90	1
92	2
94	1
96	1
109	1
112	2
115	2
116	2
118	2
120	1
121	1
122	1
126	1
127	1
128	2
132	2
136	1
137	2

Tabella E.1: Assegnazione dei pazienti del training set al relativo cluster di appartenenza

ID	Cluster assegnato	ID	Cluster assegnato	ID	Cluster assegnato
140	1	176	1	215	1
143	2	177	1	217	2
145	1	179	1	218	1
146	1	180	1	219	1
147	2	182	2	220	1
150	1	183	1	221	1
151	1	184	1	223	1
152	2	185	2	225	1
153	2	187	1	226	2
154	1	188	1	227	2
155	1	189	1	229	1
156	1	190	2	230	1
158	1	191	1	232	2
159	1	193	2	233	2
160	2	194	2	234	1
162	1	196	2	237	1
163	1	197	1	238	2
166	1	198	1	240	1
167	1	201	2	242	1
168	1	203	1	243	2
169	2	205	1	245	1
170	2	207	2	246	2
171	1	208	1	247	2
172	1	209	2	248	2
173	2	210	1	252	2
174	1	212	1	254	1
175	2	213	1		

Tabella E.2: Assegnazione dei pazienti del training set al relativo cluster di appartenenza

Appendice F

Risultati piani di radioterapia ottimizzati

	Senza cluster		**Con cluster**		**GPS**	
ID	Vescica	Retto	Vescica	Retto	Vescica	Retto
2	500	500	4180	500	1231	1027
4	980	500	1200	500	610	1830
5	500	1040	1840	940	667	1478
6	3340	1260	2620	1320	1086	3634
7	3560	760	5000	500	996	1275
8	3700	1200	4180	1320	1242	1710
9	500	500	500	500	1229	1324
10	2940	500	500	500	1633	1955
12	1360	1080	1540	1280	856	1738
14	1240	500	1520	500	1462	2001
15	5000	500	2920	500	596	1023
16	3920	500	5000	500	392	1171
17	1460	1000	1420	1300	1727	1986
18	3380	500	5000	500	1366	834
19	1360	500	1360	600	1506	1730
20	1780	1280	1840	1260	1720	1926
22	1420	1140	1420	1120	2146	2516
25	1400	1120	1400	1080	1643	1455
26	5000	500	1420	500	926	1204
27	5000	500	500	500	611	1341
29	1620	500	1520	500	1721	1247
30	3300	500	1700	500	1239	1578
31	3700	500	500	500	1582	1214
33	1520	1320	1360	1260	1547	972
34	4260	1020	1660	1200	1300	1347

	Senza cluster		**Con cluster**		**GPS**	
ID	Vescica	Retto	Vescica	Retto	Vescica	Retto
36	1600	500	1580	500	773	547
40	3540	720	5000	500	1100	1401
43	1380	500	4980	500	1076	1012
45	1480	1040	2040	500	1957	1161
48	1260	500	1300	500	964	768
49	1420	500	1500	500	1008	752
52	1480	1340	1380	960	1039	949
58	3720	1240	3720	1260	736	1626
60	2540	500	2520	500	686	1453
62	1360	500	1360	500	1661	1553
63	1240	1220	1400	1140	653	1294
64	1560	1180	1660	500	2197	1641
65	1540	1260	1340	1320	691	1705
68	1260	500	1360	500	938	1279
73	1980	1180	1380	1200	630	1474
87	1460	500	1300	740	1247	922
89	1500	500	1740	500	671	1813
90	5000	500	2100	500	361	1205
92	2080	500	1880	500	714	1356
94	3240	1080	3120	960	552	1565
96	1620	500	1980	500	3255	1403
109	3640	1100	1300	500	1748	1984
112	3560	1280	3140	1220	824	1713
115	1380	960	1340	980	1189	1328
116	1400	1320	3180	1340	3145	2538
118	1500	500	1300	1040	1764	1314
120	1620	500	1600	500	3646	1222
121	1500	1120	5000	500	1760	2110
122	5000	820	3220	500	1885	1746
126	1860	500	1860	500	2942	1393
127	3180	980	5000	960	1043	1663
128	1640	500	1440	860	1522	1828
132	2000	1420	3700	1340	1782	1878
136	1440	720	3760	500	1528	1906
137	1400	1400	1380	1340	3298	2098
140	5000	500	5000	500	771	1069
143	5000	1480	3700	1040	1504	2108
145	1020	1420	500	1460	1005	2327
146	660	500	1220	1000	1180	1052
147	500	500	500	500	1510	2095
150	1640	500	2120	500	2160	1394

	Senza cluster		**Con cluster**		**GPS**	
ID	Vescica	Retto	Vescica	Retto	Vescica	Retto
151	1160	500	960	500	1874	1523
152	1280	1260	1720	980	2642	1474
153	1700	500	1520	500	4362	2627
154	1380	1120	1460	1120	1333	788
155	500	500	500	500	444	797
156	1680	1080	1580	500	1290	912
158	4660	1260	3580	1300	1637	1640
159	960	1400	1360	1320	1733	2060
160	1440	1400	1440	1340	1489	1373
162	3420	500	2000	500	1221	989
163	2400	1020	2360	1000	351	1484
166	1780	1520	2240	1560	1363	2934
167	3460	940	500	500	1591	2233
168	3440	1020	3460	1080	1171	1739
169	3480	1180	3460	1200	881	1486
170	3420	940	1140	500	1619	1012
171	1440	500	1420	500	2660	2083
172	4840	1380	3560	1520	450	1354
173	1540	1580	1560	1620	4752	2100
174	1400	1040	500	500	1800	2025
175	2000	500	1900	500	1879	1245
176	1440	800	1620	500	1467	1770
177	1380	1100	1280	1360	2534	2424
179	1320	1140	1360	1740	1755	1423
180	3660	1160	3160	1180	1058	1482
182	5000	500	1420	500	1181	1171
183	5000	500	5000	500	1130	967
184	1140	1380	1200	1720	3165	2713
185	1400	1160	1260	1020	1840	1808
187	3540	1020	1540	1200	1138	1463
188	3540	1120	3280	1100	676	1472
189	4200	1180	3000	1060	962	1515
190	2040	1120	2400	1160	656	1295
191	1620	900	1680	1180	3294	1965
193	1520	500	1400	500	1046	1470
194	2080	500	5000	1420	1369	1910
196	1200	1160	1120	1140	2012	2456
197	1160	500	1280	500	1523	967
198	1040	1360	1100	1360	2576	2026
201	1380	820	1480	500	2049	1722
203	500	500	500	500	490	645

	Senza cluster		**Con cluster**		**GPS**	
ID	Vescica	Retto	Vescica	Retto	Vescica	Retto
205	1360	1460	1680	1520	1394	2766
207	1580	1300	1360	1220	1275	1538
208	1380	500	1120	760	1047	1022
209	1580	500	1400	500	2170	2220
210	1600	500	1780	500	1931	1107
212	3220	500	3220	640	787	1659
213	1360	500	3600	500	1131	1468
215	500	500	500	520	859	1216
217	1360	900	1860	640	702	1707
218	1620	500	1520	1040	875	1270
219	5000	500	500	500	1286	2174
220	2460	500	2920	500	670	1059
221	3400	500	1940	500	1142	1402
223	3520	860	1560	1120	1148	868
225	4240	500	3200	500	749	1189
226	500	500	2780	500	4389	1959
227	1900	500	1420	1260	2282	1282
229	500	1400	500	1540	1881	2042
230	1020	500	1080	500	2182	1824
232	1200	1440	1120	1320	1758	2517
233	1340	500	1300	500	1726	1662
234	1460	1440	1580	1420	1920	1940
237	1580	1400	1560	1580	2931	3279
238	3900	1120	3880	1200	1238	1775
240	1220	500	1460	500	1228	2160
242	1360	760	1440	500	3041	1037
243	1360	1140	1300	1220	1322	1732
245	1620	500	1640	500	2118	863
246	1100	1340	1020	1220	2797	3609
247	1520	980	1560	1080	2974	1286
248	1660	500	1380	1040	1612	1935
252	3640	500	1080	1300	1771	2213
254	1480	500	2120	500	2926	1896

Tabella F.1: Piani ottimizzati per i pazienti del training set

	Senza cluster		**Con cluster**		**GPS**	
ID	Vescica	Retto	Vescica	Retto	Vescica	Retto
3	4960	500	2400	500	2197	1148
13	3220	1140	3540	1140	4360	760
23	500	1140	4420	500	1075	1736
24	1500	500	1120	500	2016	1287
39	1520	840	1540	900	793	1068
41	1440	500	1480	500	596	1023
47	1820	520	1420	1100	750	1353
53	1500	500	3380	500	1168	947
75	1500	500	1420	1220	797	797
88	1220	500	1500	1060	938	1279
93	1320	1100	1700	1080	909	1291
124	740	500	1180	500	1256	2079
125	4360	500	3140	700	1824	1682
138	1300	500	1180	500	2271	1148
161	1440	1100	3380	1140	1009	1409
164	3280	1120	540	500	1221	1441
165	1520	500	1500	500	2244	2374
178	1260	1000	1260	980	1275	1538
181	500	580	1160	1260	1672	1368
192	3560	1200	3540	1280	1228	1290
195	1480	720	1660	980	2408	1694
199	760	1080	500	920	1279	684
200	3700	1220	1780	1040	1120	1892
202	1460	980	1840	500	646	1126
211	1320	1080	1360	1060	2925	1600
214	1200	1020	1140	1060	2570	1647
222	1500	500	1460	500	2002	1289
224	1540	780	1560	500	2531	1051
231	500	580	1940	860	1389	2091
235	1620	1140	3940	1100	3177	1035
236	1220	1060	1360	1100	1205	1744
244	1280	940	1500	980	2066	1574
249	500	960	3400	1200	1542	2049
251	540	500	4320	500	2035	1157
253	1500	500	1500	500	1426	2291

Tabella F.2: Piani ottimizzati per i pazienti del test set

Appendice G

Score percentuali dei piani radioterapeutici

	Score %		
ID	Senza cluster	Con cluster	GPS
2	100,00	100,00	100,00
4	90,15	90,24	90,17
5	97,31	96,72	98,43
6	90,57	90,59	90,42
7	100,00	99,94	100,00
8	90,08	90,08	90,12
9	99,86	99,86	100,00
10	97,60	97,91	90,11
12	100,00	98,89	89,48
14	90,15	90,17	90,21
15	100,00	100,00	100,00
16	99,99	99,99	100,00
17	90,15	90,16	90,11
18	100,00	100,00	100,00
19	99,88	99,94	100,00
20	89,98	89,99	90,00
22	88,28	88,35	88,13
25	100,00	99,99	100,00
26	100,00	100,00	100,00
27	100,00	100,00	100,00
29	99,98	99,98	100,00
30	100,00	100,00	96,73
31	100,00	99,82	100,00
33	100,00	100,00	100,00
34	100,00	100,00	100,00

	Score %		
ID	Senza cluster	Con cluster	GPS
36	90,91	100,00	100,00
40	100,00	100,00	100,00
43	99,99	99,99	100,00
45	98,52	98,82	98,76
48	100,00	100,00	100,00
49	100,00	100,00	100,00
52	100,00	100,00	99,98
58	89,74	89,75	89,78
60	100,00	100,00	98,63
62	99,01	99,01	90,16
64	89,99	98,57	90,11
65	100,00	100,00	96,66
68	100,00	100,00	100,00
73	100,00	100,00	100,00
87	100,00	100,00	100,00
89	100,00	100,00	90,06
90	100,00	100,00	98,54
92	99,99	100,00	100,00
94	90,01	99,01	90,03
96	100,00	84,69	75,76
109	88,49	90,08	88,53
112	100,00	100,00	100,00
115	100,00	100,00	100,00
116	81,13	66,22	65,99
118	99,99	100,00	100,00
120	73,33	73,33	73,33
121	90,19	99,86	90,13
122	97,14	98,52	88,71
126	98,91	98,91	75,76
127	100,00	100,00	98,14
128	99,89	98,37	90,00
132	90,12	90,14	90,12
136	97,28	98,21	90,09
137	65,82	65,78	65,84
140	99,99	99,99	100,00
143	90,13	90,13	90,13
145	89,88	89,99	89,94
146	99,97	100,00	100,00
147	98,00	98,00	90,26
150	99,88	99,88	100,00
151	99,96	99,94	98,57

	Score %		
ID	Senza cluster	Con cluster	GPS
152	99,87	99,95	98,48
153	66,11	66,05	66,06
154	100,00	100,00	100,00
155	100,00	100,00	100,00
156	100,00	100,00	100,00
158	96,45	96,62	98,03
159	98,18	97,63	89,63
160	99,99	99,98	99,98
162	100,00	100,00	100,00
163	100,00	100,00	99,99
166	90,27	90,34	90,36
167	90,11	98,00	90,05
168	90,14	90,03	90,21
169	100,00	100,00	100,00
170	100,00	100,00	100,00
171	80,95	81,12	65,97
172	100,00	100,00	100,00
173	65,63	65,65	65,88
174	90,02	98,05	90,08
175	99,99	99,98	100,00
176	100,00	99,95	98,11
177	89,03	88,74	88,26
179	99,99	100,00	100,00
180	90,01	89,98	90,04
182	100,00	99,99	100,00
183	100,00	100,00	100,00
184	80,57	80,75	66,13
185	89,95	98,83	89,97
187	98,00	89,62	89,65
188	100,00	100,00	100,00
189	89,99	89,90	89,98
190	100,00	100,00	100,00
191	65,84	65,86	66,04
193	100,00	100,00	100,00
194	99,88	89,57	89,75
196	81,06	80,98	81,13
197	99,98	99,98	100,00
198	87,93	87,82	88,69
201	87,95	97,41	88,08
203	100,00	100,00	100,00
205	90,26	90,27	90,21

	Score %		
ID	Senza cluster	Con cluster	GPS
207	100,00	100,00	96,95
208	100,00	100,00	100,00
209	87,96	88,14	86,37
210	100,00	100,00	100,00
212	99,91	99,97	97,82
213	99,98	99,97	97,54
215	99,10	99,09	99,68
217	98,48	100,00	89,88
218	100,00	100,00	100,00
219	90,35	90,18	90,28
220	100,00	100,00	100,00
221	99,82	99,80	99,97
223	100,00	100,00	100,00
225	100,00	100,00	98,78
226	62,60	65,84	65,96
227	99,92	99,97	100,00
229	88,78	88,78	90,30
230	98,00	97,96	98,15
232	88,69	88,88	90,52
233	99,95	99,95	97,95
234	65,88	65,90	65,91
237	66,28	66,38	66,19
238	89,96	98,99	90,03
240	98,23	97,56	90,26
242	90,91	90,89	75,76
243	98,38	98,45	98,02
245	100,00	100,00	100,00
246	88,44	88,47	68,14
247	90,12	90,38	75,76
248	99,91	100,00	89,46
252	89,81	90,18	90,22
254	97,39	94,98	75,02

Tabella G.1: Punteggi dei piani di trattamento per i pazienti del training set

	Score %		
ID	Senza cluster	Con cluster	GPS
3	100,00	100,00	100,00
13	98,22	100,00	100,00
23	97,82	99,92	89,85
24	100,00	99,94	100,00
39	100,00	100,00	100,00
41	97,96	97,99	99,03
47	99,90	100,00	100,00
75	100,00	100,00	98,28
88	100,00	100,00	100,00
93	100,00	100,00	96,39
124	89,96	90,03	90,16
125	98,38	90,11	90,12
138	100,00	100,00	100,00
161	97,48	97,78	98,67
164	99,96	97,52	100,00
165	88,87	88,71	81,12
178	99,97	99,95	99,98
181	97,76	100,00	98,40
192	98,55	97,58	97,55
195	99,94	98,23	98,45
199	98,55	98,46	100,00
200	90,25	90,25	90,30
202	100,00	100,00	100,00
211	98,19	98,01	73,74
214	89,52	89,19	80,69
222	99,98	99,98	100,00
224	99,21	99,01	99,18
231	90,28	90,22	90,12
235	90,91	88,89	89,10
236	98,27	98,20	89,87
244	98,43	98,50	98,58
249	90,07	90,18	90,18
251	99,91	100,00	100,00
253	87,46	87,46	88,03

Tabella G.2: Punteggi dei piani di trattamento per i pazienti del test set

Bibliografia

[1] Tonini, G. *et al.* (2020). "La prostata". *I numeri del cancro in Italia*: 60-61.

[2] Humanitas. <https://www.humanitas.it/malattie/tumore-della-prostata/> [ultimo accesso: 04/09/2022]

[3] Alessi, A. *et al.* (2018). Il cancro della prostata. 9. Milano: AIMAC.

[4] Huggins, C. (1967). "Endocrine-induced regression of cancers". *Science*, 156(3778): 1050-4.

[5] <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-della-prostata> [ultimo accesso 04/09/2022]

[6] <https://www.airc.it/cancro/affronta-la-malattia/guida-alle-terapie/radioterapia> [ultimo accesso 04/09/2022]

[7] Hyun, M.K., Suk, H.K., Bo, S.K. (2018). "Radioprotective effects of delphinidin on normal human lung cells against proton beam exposure". *Nutrition Researc and Practice*, 12(1): 41-46.

[8] RaySearch Laboratories. <https://www.raysearchlabs.com/raystation/> [ultimo accesso 04/09/2022]

[9] Tecnologie Avanzate. <https://tecnologieavanzate.com/radioterapia/tps/raystation/> [ultimo accesso 04/09/2022]

[10] Omran, M.G.H., Engelbrecht, A.P. and Salman, A. (2007). "An overview of clustering methods". *Intelligent Data Analysis*, 11(6): 583–605.

[11] Hinton, G.E. (1992). “How neural networks learn from experience”. *Scientific American*:145–151

[12] Gullo, G. (2018). "Deep learning svelato: ecco come funzionano le reti neurali artificiali". *italiancoders.it*, 15 settembre. <https://italiancoders.it/deep-learning-svelato-ecco-come-funzionano-le-reti-neurali-artificiali/> [ultimo accesso 04/09/2022]

[13] Medici, P. (2022). "Reti neurali". *ce.unipr.it*, 9 marzo. <http://www.ce.unipr.it/ medici/geometry/node110.html> [ultimo accesso 04/09/2022]

[14] <https://nerdthecoder.wordpress.com/2019/02/03/recurrent-neural-net/> [ultimo accesso 04/09/2022]

[15] Cover, T.M. (1965). Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. *doi:10.1109/pgec.1965.264137*, EC-14(3): 326–334.

[16] Hashemi F.A., Madanifar, F. and Abbasi, M. (2018). Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. Petroleum. *Petroleum*, 8(2).

[17] Bansal, R. (2017). "Radial basis function network". *.hackerearth.com*, 6 febbraio. <https://www.hackerearth.com/blog/developers/radial-basis-function-network/> [ultimo accesso: 04/09/2022]

[18] Broomhead, D.S. and Lowe, D. (1988). "Multivariate functional interpolation and adaptive networks". *Complex systems*, 2:321-355.

[19] Darken, C. and Moody, J. (1990). "Note on learning rate schedules for stochastic optimization". *NIPS'90: Proceedings of the 3rd International Conference on Neural Information Processing Systems*, 832–838.

[20] Poggio, T. and Girosi, F. (1990). "Networks for Approximation and Learning". *Proceedings of the IEEE*, 78(9): 1481-1497.

[21] Berrar, D. (2018). "Cross-Validation". *Elsevier*.

[22] Zhou, B. *et al.* (2019). "Analysis of Factors Affecting Hit-and-Run and Non-Hit-and-Run in Vehicle-Bicycle Crashes: A Non-Parametric Approach Incorporating Data Imbalance Treatment". *Sustainability*, 11.

[23] Chatterjee, C. and Das, S. (2021). "Recent changes in Indian monsoon in light of regionalization based on various rain features". *Theoretical and Applied Climatology*, 146.

[24] Fiandra, C. *et al.* (2018). "Automated heuristic optimization of prostate VMAT treatment planning". *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology*, 07(03):414–425.

[25] Perolini, V. (2018). *Utilizzo di algoritmi genetici per l'ottimizzazione di piani per radioterapia = Use of genetic algorithms for the optimization of radiotherapy plans.* Tesi di dottorato. Politecnico di Torino.

[26] Mengoni, G. (2020). *Caratterizzazione geometrica di volumi TC della prostata per il set-up di parametri nell'ottimizzazione di piani radioterapeutici = Geometrical caracterization of prostate CT volumes for a parameter set-up in the optimization of radiotherapy plans.* Tesi di Laurea Magistrale. Politecnico di Torino.

[27] Liuzzi, J. (2021). *Ottimizzazione di piani radioterapeutici per il tumore alla prostata = Optimization of radiotherapy plans for prostate cancer.* Tesi di Laurea Magistrale. Politecnico di Torino.