



POLITECNICO DI TORINO

Corso di Laurea in Magistrale
in Ingegneria Biomedica

Tesi di Laurea Magistrale

**Caratterizzazione geometrica di
volumi TC della prostata per il
set-up dei parametri
nell'ottimizzazione di piani
radioterapeutici**

Relatori

Prof.ssa Samanta Rosati
Prof.ssa Gabriella Balestra
Christian Corrado Fiandra

Candidata

Giada Mengoni

Dicembre 2020

Sommario

La radioterapia a fasci esterni rappresenta una delle possibili terapie per la cura del tumore della prostata. Essa richiede la pianificazione di un piano radioterapeutico con l'obiettivo di irradiare la zona tumorale salvaguardando le regioni anatomiche sane vicine (organi a rischio).

In uno studio precedente si è dimostrata la validità clinica di piani terapeutici in cui la massima dose equivalente (MaxEUD) al retto e alla vescica è ottimizzata grazie all'implementazione di un algoritmo genetico avente integrata una rete RBFNN nella fitness. Per ciascun paziente sono stati identificati parametri specifici del GA e della RBFNN ed i risultati così ottenuti sono stati confrontati con quelli ricavati imponendo un set di parametri generico, valido per tutto il dataset.

Nel presente lavoro di tesi, un insieme di pazienti è stato suddiviso in sottogruppi sulla base di una caratterizzazione geometrica dei volumi TC della prostata e per ognuno dei cluster, sono stati selezionati dei parametri che potessero essere validi per ciascuno dei pazienti del sottogruppo.

Il confronto tra i risultati ottenuti con il set specifico per il cluster e quelli con il set selezionato nella fase di generalizzazione dello studio precedente evidenzia un miglioramento nelle performance e permette, dunque, di validare l'ipotesi che le caratteristiche geometriche della prostata contribuiscano alla definizione del piano di radioterapia.

Ringraziamenti

Ringrazio la Professoressa Samanta Rosati e la Professoressa Gabriella Balestra per avermi seguito con dedizione durante lo svolgimento del mio lavoro di tesi, il Fisico Medico Christian Fiandra e l'Ing. Stefania Zara per il tempo dedicatomi per il supporto nell'utilizzo dei software.

Un ringraziamento speciale va alla mia famiglia, che mi ha supportato durante tutto il percorso di studi. Dedico a mio padre, mia madre e mio fratello il raggiungimento di questo traguardo, perchè nonostante le molte difficoltà, mi sono sempre stati vicini.

Vorrei ringraziare anche le mie compagne di viaggio Marta, Caterina e Cristina. Abbiamo intrapreso questo percorso insieme, condividendo gioie e dolori, ci siamo supportate a vicenda e credo fortemente che senza di voi, tutto sarebbe risultato più difficile.

Un pensiero va a tutte le persone che in questi cinque anni, e non solo, mi sono state accanto. A Fulvio, Luca e Francesco per le risate fatte insieme, ad Agnese, perchè il legame che ci unisce è molto importante per me e so che posso contare su di lei, ad Alessia, perchè è l'amica che c'è sempre.

Indice

Elenco delle figure	6
Elenco delle tabelle	10
1 Introduzione	12
1.1 Descrizione del problema	12
1.2 Anatomia della prostata	13
1.2.1 Tumore della prostata	13
1.3 Radioterapia	14
1.3.1 Piano di trattamento	15
1.4 Caratterizzazione dei volumi della prostata: analisi della letteratura	17
2 Materiali e Metodi	19
2.1 Pazienti	19
2.2 Raystation	19
2.3 K-Nearest Neighbor	19
2.4 Clustering	20
2.4.1 Misure di similarità	21
2.4.2 Clustering gerarchico	22
2.4.3 Dendrogramma	23
2.4.4 Normalizzazione dei dati	24
2.5 Rete Neurale a Base Radiale (RBFNN)	25
2.5.1 Architettura della rete	25
2.5.2 Algoritmi di apprendimento	26
2.6 Algoritmi Genetici	27
3 Elaborazione delle immagini TC	30
3.1 Data Cleaning	30
3.2 Estrazione delle maschere	30
3.3 Analisi geometrica delle maschere estratte	31
3.4 Analisi della forma e estrazione di nuove maschere	33
3.5 Estrazione dei parametri geometrici	36

4	Clustering dei dati	39
4.1	Risultati del clustering	39
4.2	Associazione Paziente-Cluster	41
5	Ottimizzazione dei piani di radioterapia a partire dal clustering dei pazienti	44
5.1	Rete RBFNN	44
5.1.1	Estrazione dei dati di input della rete tramite Raystation	44
5.1.2	Allenamento della rete	45
5.1.3	Tuning dei parametri	45
5.1.4	Risultati Tuning RBFNN	45
5.1.5	Analisi dei singoli parametri	46
5.1.6	Analisi delle combinazioni	51
5.2	GA	59
5.2.1	Risultati dell'algorithmo genetico	60
5.3	Valutazione dei piani radioterapeutici tramite software	63
5.3.1	Confronto GPS e RBFNN+GA	65
5.3.2	Confronto GPS e RBFNN+GA (versione precedente[10])	65
5.3.3	Confronto RBFNN+GA e RBFNN+GA (versione precedente[10])	66
6	Conclusioni	68
A	Grafici relativi al clustering	69
B	Risultati associazione Paziente-Cluster	74
B.1	Distanza Paziente-Centroide	74
B.2	Associazione k-nn	75
C	Risultati tuning parametri della rete RBFNN-Analisi dei singoli parametri	78
C.1	Confronto target e valore di fitness restituito dalla rete	84
D	Risultati tuning parametri della rete RBFNN-Analisi delle combinazioni	115
E	Risultati GA-Valori delle MaxEUD vescica e retto	119
F	Valutazione dei piani di trattamento	126
	Bibliografia	133

Elenco delle figure

1.1	Istogramma dose-volume	16
1.2	PCAR [9]	17
1.3	Parametri estratti [2]	18
2.1	Diagramma del Clustering [6]	20
2.2	Clustering gerarchico e divisivo [6]	22
2.3	Esempio di dendrogramma [6]	23
3.1	Dimensioni in pixel dell'altezza dei Bounding Box delle regioni nelle slices dei 202 pazienti	31
3.2	Dimensioni in pixel della base dei Bounding Box delle regioni nelle slices dei 202 pazienti	32
3.3	Esempi di slices anomale	32
3.4	Confronto tra le dimensioni dell'altezza del BB delle 113 maschere complete e la media delle dimensioni dell'altezza del BB delle 86 maschere corrette	34
3.5	Confronto tra le dimensioni dell'altezza del BB delle 113 maschere modificate e la media delle dimensioni dell'altezza del BB delle 86 maschere corrette	35
3.6	Confronto tra le dimensioni della base del BB delle 113 maschere complete e la media delle dimensioni della base del BB delle 86 maschere corrette	35
3.7	Confronto tra le dimensioni della base del BB delle 113 maschere modificate e la media delle dimensioni della base del BB delle 86 maschere corrette	36
4.1	Dendrogramma con tagli a 2,3 e 5 cluster	39
4.2	Box plot delle distribuzioni del parametro 7 nelle varie clusterizzazioni	41
5.1	Pazienti 254-245-40	46
5.2	Bar Diagram dei parametri	47
5.3	Bar Diagram dei parametri	48
5.4	Bar Diagram dei parametri	49

5.5	Box-plot Cluster 1	50
5.6	Box-plot Cluster 2	51
5.7	Box-plot Cluster 3	51
5.8	Analisi percentili	52
5.9	Analisi percentili	53
5.10	Analisi percentili	54
5.11	Analisi delle combinazioni	55
5.12	Analisi delle combinazioni	55
5.13	Analisi delle combinazioni	56
5.14	Confronto tra le distribuzioni di mse delle combinazioni ottenute e le distribuzioni di mse relative alla combinazione ricavata in [10] . .	57
5.15	Confronto tra le distribuzioni di mse delle combinazioni scelte per il cluster e le distribuzioni di mse relative alle combinazioni degli altri cluster	58
5.16	Confronto tra le distribuzioni di mse dei pazienti del validation set .	59
5.17	Cluster 1 - Training set - Combinazione 471-449	61
5.18	Cluster 1 - Validation set - Combinazione 471-449	61
5.19	Cluster 2 - Training set - Combinazione 227-449	61
5.20	Cluster 2 - Validation set - Combinazione 227-449	61
5.21	Cluster 3 - Training set - Combinazione 499-449	62
5.22	Cluster 3 - Validation set - Combinazione 499-449	62
5.23	Distribuzioni degli scores GPS vs RBFNN+GA	63
5.24	Distribuzioni degli scores GPS vs RBFNN+GA (Validation Set) . .	64
5.25	Distribuzioni degli scores GPS vs RBFNN+GA vs RBFNN+GA[10]	64
A.1	Tre clusterizzazioni del dataset	69
A.2	Confronto variabilità intra-cluster	70
A.3	Box-plot	71
A.4	Box-plot	71
A.5	Box-plot	71
A.6	Box-plot	71
A.7	Box-plot	72
A.8	Box-plot	72
A.9	Box-plot	72
A.10	Box-plot	72
A.11	Box-plot	73
A.12	Box-plot	73
C.1	Pazienti 2-5-6	84
C.2	Pazienti 7-9-12	84
C.3	Pazienti 13-17-26	85
C.4	Pazienti 27-31-47	85
C.5	Pazienti 235-236-244	86

Elenco delle figure

C.6 Pazienti 251-253-32	86
C.7 Pazienti 49-237-29	87
C.8 Pazienti 36-89-39	87
C.9 Pazienti 220-222-224	88
C.10 Pazienti 225-233-234	88
C.11 Pazienti 53-91-45	89
C.12 Pazienti 63-68-95	89
C.13 Pazienti 202-203-205	90
C.14 Pazienti 207-208-210	90
C.15 Pazienti 212-213-215	91
C.16 Pazienti 43-219-41	91
C.17 Pazienti 52-87-109	92
C.18 Pazienti 120-121-122	92
C.19 Pazienti 124-126-127	93
C.20 Pazienti 128-136-138	93
C.21 Pazienti 151-180-181	94
C.22 Pazienti 183-184-188	94
C.23 Pazienti 189-190-195	95
C.24 Pazienti 140-146-150	95
C.25 Pazienti 154-156-160	96
C.26 Pazienti 161-167-172	96
C.27 Pazienti 176-166	97
C.28 Pazienti 10-112-115	98
C.29 Pazienti 118-132-137	98
C.30 Pazienti 14-15-152	99
C.31 Pazienti 153-169-170	99
C.32 Pazienti 173-175-182	100
C.33 Pazienti 185-193-194	100
C.34 Pazienti 196-201-209	101
C.35 Pazienti 21-211-214	101
C.36 Pazienti 217-226-231	102
C.37 Pazienti 238-243-246	102
C.38 Pazienti 247-248-249	103
C.39 Pazienti 25-252-30	103
C.40 Pazienti 33-57-58	104
C.41 Pazienti 60-61-64	104
C.42 Pazienti 65-75-92	105
C.43 Pazienti 4-11-16	106
C.44 Pazienti 18-20-22	106
C.45 Pazienti 34-48-110	107
C.46 Pazienti 3-24-62	107
C.47 Pazienti 66-111-113	108

Elenco delle figure

C.48 Pazienti 114-117-129	108
C.49 Pazienti 130-131-139	109
C.50 Pazienti 142-162-163	109
C.51 Pazienti 168-73-8	110
C.52 Pazienti 94-1-119	110
C.53 Pazienti 123-143-155	111
C.54 Pazienti 158-159-96	111
C.55 Pazienti 23-164-177	112
C.56 Pazienti 178-187-191	112
C.57 Pazienti 192-197-198	113
C.58 Pazienti 199-200-229	113
C.59 Pazienti 90-223-230	114
C.60 Pazienti 240-242-250	114

Elenco delle tabelle

3.1	Parametri geometrici implementati	38
4.1	Clusterizzazione a 2 cluster	40
4.2	Clusterizzazione a 3 cluster	40
4.3	Clusterizzazione a 5 cluster	40
4.4	<i>Confusion Matrix k=11</i>	42
4.5	<i>Confusion Matrix k=11 Test set=Training Set</i>	43
4.6	<i>Confusion Matrix k=11 Test set=Validation Set</i>	43
5.1	Combinazioni selezionate	47
5.2	Combinazioni selezionate	48
5.3	Combinazioni selezionate	49
5.4	Combinazione scelta in [10]	50
5.5	Combinazioni scelte per ciascun cluster	56
5.6	Confronto scores dei piani di trattamento	65
5.7	Confronto scores dei piani di trattamento	65
5.8	Confronto scores dei piani di trattamento	66
5.9	Confronto scores dei piani di trattamento	66
5.10	Confronto scores dei piani di trattamento	67
5.11	Confronto scores dei piani di trattamento	67
B.1	<i>Confusion Matrix mediana-euclidea</i>	74
B.2	<i>Confusion Matrix mediana-L1</i>	74
B.3	<i>Confusion Matrix media-euclidea</i>	75
B.4	<i>Confusion Matrix media-L1</i>	75
B.5	<i>Confusion Matrix k=5</i>	75
B.6	<i>Confusion Matrix k=6</i>	75
B.7	<i>Confusion Matrix k=7</i>	76
B.8	<i>Confusion Matrix k=8</i>	76
B.9	<i>Confusion Matrix k=9</i>	76
B.10	<i>Confusion Matrix k=10</i>	76
B.11	<i>Confusion Matrix k=11</i>	77

B.12 <i>Confusion Matrix</i> $k=12$	77
B.13 <i>Confusion Matrix</i> $k=13$	77
C.1 Set parametri migliori	78
C.2 Set parametri migliori	80
C.3 Set parametri migliori	82
D.1 Set parametri migliori	115
D.2 Set parametri migliori	116
D.3 Set parametri migliori	117
E.1 Dosi ottimizzate	119
E.2 Dosi ottimizzate	122
E.3 Dosi ottimizzate	123
F.1 Scores dei piani di trattamento	126
F.2 Scores dei piani di trattamento	129
F.3 Scores dei piani di trattamento	130

Capitolo 1

Introduzione

1.1 Descrizione del problema

Il tumore della prostata è una delle patologie più comuni tra la popolazione maschile. Una volta diagnosticato, molto spesso, il trattamento scelto per la cura della neoplasia è la radioterapia. Essa consiste nell'esporre l'organo di interesse ad una specifica dose di radiazioni ionizzanti, con l'obiettivo di distruggere le cellule tumorali, salvaguardando le strutture sane adiacenti che, nel caso di tumore della prostata, coincidono con la vescica e il retto ("OAR, organi a rischio").

Per ottenere una terapia efficace, è dunque necessaria la creazione di un piano di radioterapia, nel quale devono essere definiti gli obiettivi, i vincoli di dose, le caratteristiche del fascio di radiazioni.

Nell'elaborazione del piano che, generalmente, richiede la collaborazione di diverse figure professionali, quali ad esempio il medico oncologo radioterapista e il fisico medico, hanno un ruolo fondamentale software di pianificazione automatizzata (TPS) che fungono da supporto alle decisioni dello specialista.

Nel lavoro di tesi precedentemente svolto [10], è stato implementato un algoritmo genetico la cui funzione di fitness è una Radial Basis Neural Network, con l'intento di ottimizzare la massima dose erogabile al retto e alla vescica durante un trattamento radioterapeutico per la cura del tumore della prostata. Il vantaggio derivante dall'associazione del GA con la RBFNN sta nel riuscire a valutare un grande numero di combinazioni di dosi al retto e alla vescica senza dover ricorrere al TPS, riducendo i tempi di elaborazione. Solo la coppia di dose individuata dall'algoritmo genetico viene settata nel software, il quale genera il piano ad essa associata.

I risultati ottenuti, per i singoli pazienti analizzati, dimostrano che i piani con la coppia di dose erogabile al retto e alla vescica, soluzione del GA, sono clinicamente validi.

A partire da questi risultati, si è pensato di valutare se le caratteristiche geometriche della prostata potessero contribuire allo sviluppo del piano di radioterapia.

Con il presente lavoro di tesi, infatti, si cerca di studiare la validità dell'approccio

appena esposto. Esso prevede più fasi di analisi. Nella prima fase sono stati implementati algoritmi di analisi della forma per estrarre, a partire da immagini TC di tumore della prostata, dei parametri geometrici che caratterizzino la geometria del volume di interesse. A partire da questi, si è applicato un metodo di clustering dei dati con l'obiettivo di suddividere i pazienti in sottogruppi sulla base delle proprietà geometriche della prostata e di poter stabilire dei parametri della RBFNN con i quali ottenere delle reti performanti per ciascuno dei pazienti del sottogruppo.

1.2 Anatomia della prostata

La prostata è una ghiandola appartenente all'apparato riproduttivo e urinario maschile. E' costituita da differenti tipi di cellule:

- cellule ghiandolari, che contribuiscono alla produzione dello sperma;
- cellule muscolari, che regolano il flusso d'urina e l'eiaculazione;
- cellule fibrose, di sostegno.

La prostata è ricoperta da uno strato di tessuto connettivo, la capsula prostatica. In condizioni normali ha una forma ovale, simile ad una noce. Si trova al di sotto della vescica e davanti al retto.

Il suo ruolo principale è la secrezione del liquido prostatico che contribuisce alla vitalità degli spermatozoi.

Inoltre, agisce nella regolarizzazione del flusso urinario in quanto le due valvole che regolano la minzione, sono situate nell'uretra laddove è presente anche la prostata, dunque il tessuto muscolare funge da rinforzo all'apertura e alla chiusura dei due sfinteri.

1.2.1 Tumore della prostata

Il tumore della prostata ha origine dalle cellule della ghiandola che iniziano a crescere e proliferare in maniera incontrollata.

È una delle patologie più comuni tra gli uomini. L'incidenza è in continua crescita ma il tasso di mortalità diminuisce perché, grazie al test del PSA (Antigene Prostatico Specifico), è possibile ottenere una diagnosi precoce. I principali fattori che aumentano il rischio di formazione di un tumore della prostata sono l'età, la storia genetica, la dieta ricca di grassi.

La diagnosi della neoplasia può essere ottenuta attraverso esami di screening quali:

- Dosaggio del PSA, PCA3, PHI (prostate health index)
- Esplorazione digito rettale (visita urologica)

- Biopsia prostatica: prelievo di campioni di prostata da analizzare al microscopio per valutare la presenza di cellule tumorali

La stadiazione, invece, può essere completata con tecniche di diagnostica per immagini:

- Risonanza magnetica
- TC addome pelvi con mezzo di contrasto
- PET

1.3 Radioterapia

La radioterapia è una delle tecniche utilizzate per il trattamento del tumore della prostata. Essa consiste nell'irradiare la regione tumorale con radiazioni ionizzanti ad alta energia che interagiscono con le cellule cancerogene, le quali, subendo un danno a livello del DNA, non sono più in grado di proliferare. L'effetto è dunque una riduzione progressiva del tumore. A seconda della tipologia di radiazione utilizzata, è possibile distinguere due tipologie di trattamento:

- Radioterapia a fasci esterni, in cui si fa uso di fasci di raggi X, prodotti da sorgenti esterne al paziente (acceleratori lineari)
- Brachiterapia, in cui la sorgente di radiazione è interna al paziente. In prossimità del tumore, si iniettano nuclei radioattivi che decadono rilasciando energia sottoforma di radiazione ionizzante.

Le tecniche di radioterapia a fasci esterni più utilizzate sono:

- Radioterapia conformazionale (3D-CRT):
Si utilizzano fasci di radiazioni adattati alla forma del volume di target.
- Radioterapia ad intensità modulata (IMRT):
L'intensità del fascio di radiazioni è modulata grazie alla presenza di un collimatore multilamellare (MLC) computerizzato. Le lamelle si muovono automaticamente seguendo una sequenza specifica con l'intento di colpire il tumore da diverse angolazioni e cercare di concentrare la maggior parte della radiazione sul target, limitando l'esposizione dei tessuti sani. È una tecnica che viene utilizzata, generalmente, quando si vogliono trattare tumori di forme complesse o situati in prossimità di organi particolarmente critici.
- Volumetric Modulated Arc Therapy (VMAT):
Evoluzione della tecnica IMRT, consente di irradiare il volume di interesse mediante archi multipli. La rotazione del gantry permette di erogare la radiazione in maniera continua e il movimento del complesso MLC, ne consente

la modulazione. Con questa tecnica si riducono i tempi dei trattamenti e si riesce ad ottenere la migliore focalizzazione della dose.

Il trattamento radioterapeutico può essere prescritto come:

- Terapia successiva all'asportazione chirurgica del tumore, per eliminare eventuali cellule tumorali residue e limitare la possibilità che il tumore si ripresenti
- Terapia preoperatoria, per ridurre le dimensioni del tumore e agevolarne l'asportazione
- Terapia a scopo curativo

L'intento della radioterapia è colpire in maniera selettiva le cellule cancerogene riducendo al minimo l'esposizione alle radiazioni degli organi circostanti sani (detti organi a rischio "OAR"). Da qui, la necessità della pianificazione di un piano di trattamento.

1.3.1 Piano di trattamento

Le figure professionali che si occupano della progettazione del piano di cura sono il medico oncologo radioterapista e il fisico medico. A seconda della tipologia di tumore, della sua localizzazione e delle sue dimensioni, viene scelta la tipologia di terapia, la dose da somministrare alla lesione e il frazionamento di essa. Il tutto a conclusione del primo step della pianificazione, ovvero la visita medica. L'equipé concorda poi con il paziente una TC di centratura per acquisire le immagini e per definire il posizionamento del paziente sul lettino che dovrà rimanere tale anche durante le sedute di radioterapia. Per garantire il corretto posizionamento, vengono effettuati dei tatuaggi permanenti sul corpo del paziente, punti di riferimento per le fasi successive del trattamento. Una volta eseguito l'esame, il tecnico radioterapista effettua, con l'ausilio di un software, il contouring degli organi a rischio vicini al volume da irradiare. Nello specifico caso di tumore della prostata, gli organi a rischio sono la vescica, il retto e le teste femorali destra e sinistra. Questo processo è fondamentale perché irradiare organi sani con dosi eccessive, che superano la dose massima erogabile, può indurre la generazione di gravi danni per il paziente. Il medico radio-oncologo, successivamente, contorna tre volumi:

- GTV (Gross Tumor Volume): il volume tumorale visibile
- CTV (Clinical Target Volume): è il volume bersaglio che comprende il GTV e le regioni anatomiche vicine, le quali potrebbero essere interessate dalla malattia
- PTV (Plannig Target Volume): è il volume bersaglio realmente irradiato, utilizzato per il calcolo della distribuzione della dose. È più grande del CTV

perché tiene in considerazione movimenti del paziente dovuti alla respirazione, alla deglutizione ed eventuali variazioni geometriche delle strutture anatomiche interessate (Internal Margin) o incertezze di posizionamento (Set-up Margin).

Treatment Planning System

Il Treatment Planning System (TPS) è un software di pianificazione automatica dei piani di radioterapia. Può essere utilizzato seguendo un approccio Inverse Planning, in cui si definiscono vincoli di dose per il volume target e gli OAR e alcune caratteristiche tecniche come ad esempio le direzioni del fascio. Successivamente un algoritmo di ottimizzazione determina i parametri del piano che meglio soddisfano i vincoli, utilizzando la minimizzazione di una funzione obiettivo.

Un altro approccio è il Forward Planning, in cui gli operatori scelgono, per ogni paziente, secondo la loro esperienza, la tecnica di irradiazione impostando il numero e la direzione di incidenza dei fasci, l'uso dei collimatori, la dose al PTV. Se la distribuzione della dose, calcolata dal TPS, con i parametri impostati non è soddisfacente, si procede alla modifica di alcuni di essi, secondo il principio del "try and see" fino a quando non si ottiene il piano di trattamento migliore per lo specifico caso clinico.

I vantaggi dell'Inverse Planning rispetto al Forward Planning risiedono nell'eliminare i tempi necessari a modificare i parametri qualora ce ne fosse l'esigenza e nel ridurre la dipendenza della bontà dei piani dall'esperienza dell'operatore.

Istogramma dose-volume (DHV)

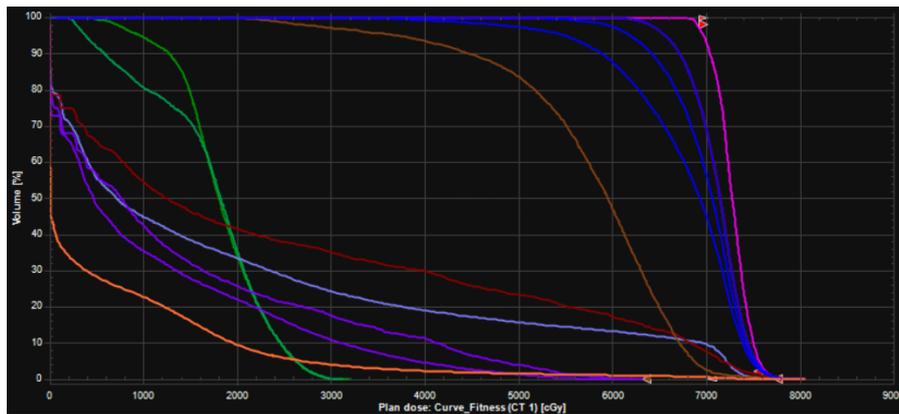


Figura 1.1: Istogramma dose-volume

Lo scopo della progettazione di un piano di trattamento è quello di ottenere una distribuzione ottimale della dose conforme agli obiettivi di dose prescritti per il volume target e le dosi di tolleranza dei tessuti sani circostanti.

Per valutare la distribuzione della dose si utilizza l'istogramma dose-volume, riportato in figura 1.1, il quale illustra la percentuale di volume (asse delle ordinate) che assorbe una certa dose di radiazioni (asse delle ascisse). Un piano terapeutico è soddisfacente se gli organi a rischio assorbono piccole dosi, quindi le curve che li rappresentano sono alla sinistra del grafico, e se, al contrario, una grande percentuale di volume del target assorbe la dose prescritta, quindi la curva deve trovarsi nella parte destra del grafico.

1.4 Caratterizzazione dei volumi della prostata: analisi della letteratura

È stata svolta una ricerca bibliografica su studi che presentassero dei parametri di tipo geometrico utili alla caratterizzazione di volumi anatomici. Matsugasumi et al. [9] hanno condotto uno studio sull'impatto dell'età del paziente sui cambiamenti anatomici della prostata. È stata svolta un'analisi morfometrica di immagini di risonanza magnetica di 307 pazienti statunitensi e giapponesi aventi un livello di PSA elevato. Successivamente alla segmentazione della ghiandola prostatica, sono stati estratti alcuni parametri geometrici tra i quali il volume della prostata, il volume della zona di transizione, il volume della zona periferica e il PCAR (Presumed Circle Area Ratio). Quest'ultimo parametro rappresenta il rapporto tra l'area della massima sezione assiale della prostata e l'area del cerchio avente circonferenza pari al perimetro della stessa (Figura 1.2).

È un buon indicatore della severità della BPH (iperplasia prostatica benigna), la

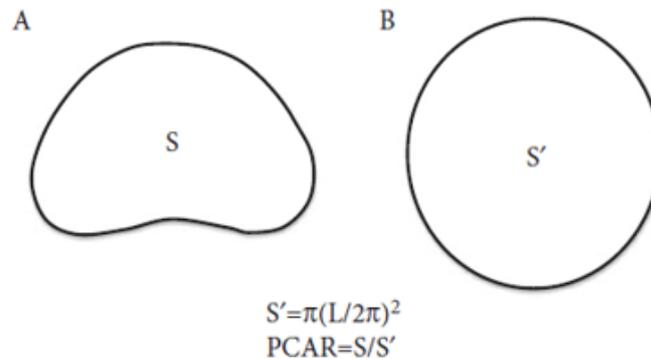


Figura 1.2: PCAR [9]

quale è caratterizzata da un'iperplasia della zona di transizione. Maggiore è il valore di PCAR, maggiore è la pressione intraprostatica, si ha un allargamento della zona di transizione e in accordo con il principio di Pascal, la forma della sezione prostatica diviene più rotonda. Il parametro è stato implementato nel lavoro di tesi.

Nello studio di Jan L. Bruse et al. [2] viene presentato un metodo statistico di analisi della forma, che permette di estrarre delle caratteristiche della forma anatomica 3D da immagini di risonanza magnetica cardiovascolare (CMR), senza necessità di un riferimento manuale. Lo studio prevede l'estrazione di parametri geometrici a partire da immagini di pazienti che hanno subito un intervento chirurgico per il trattamento della coartazione dell'aorta. Il modello (ovvero una "forma anatomica media") è stato ricostruito a partire da 20 immagini di archi aortici. Applicando ad esso delle deformazioni e trasformazioni adeguate è possibile ottenere il volume specifico del soggetto. Dal modello sono estratti i parametri di forma riportati in figura 1.3 e confrontati con quelli misurati in maniera convenzionale, trovando tra essi una buona correlazione.

Table 1
Morphometric parameters measured on the 3D surface models of the arches

2D measured parameters (manually)	3D measured parameters (VMTK)
<ul style="list-style-type: none"> • Arch height A [mm] • Arch width T [mm] • Ratio A/T • Diameters at ascending, transverse, isthmus and descending level of the aorta: D_{asc}, D_{trans}, D_{isth}, D_{desc} [mm] • Ratios D_{asc}/D_{desc}, D_{trans}/D_{desc} and D_{isth}/D_{desc} 	<ul style="list-style-type: none"> • Volume V [mm^3] • Surface Area A_{surf} [mm^2] • Ratio A_{surf}/V • Centreline length L_{CL} [mm] • Centreline Tortuosity To_{CL} • Median curvature along centreline C_{med} [1/mm] • Maximum, minimum and median diameter along centreline D_{max}, D_{min}, D_{med} [mm]

Figura 1.3: Parametri estratti [2]

La morfometria 2D convenzionale è stata eseguita manualmente sui dati di imaging CMR. I parametri di forma 3D sono stati calcolati semi-automaticamente dai modelli segmentati della superficie dell'arco usando The Vascular Modeling Toolkit (VMTK, Orobix, Bergamo, Italia; www.vmtk.org) in combinazione con Matlab.

Capitolo 2

Materiali e Metodi

2.1 Pazienti

Le immagini TC dei pazienti affetti da tumore della prostata, utilizzate nelle analisi condotte in questo lavoro di tesi, sono state fornite dal dipartimento di Oncologia dell'Ospedale delle Molinette di Torino. Le cartelle, contenenti gli esami diagnostici in formato *dicom*, sono state ordinate assegnando un codice identificativo da 1 a 254, garantendo così l'anonimato dei pazienti.

L'analisi delle immagini, l'estrazione dei parametri geometrici, le analisi di clustering, la rete RBFNN e l'algoritmo genetico sono stati implementate in Matlab.

2.2 Raystation

Raystation è il software TPS, commercializzato da RaySearch Laboratories, utilizzato in questo lavoro di tesi per la generazione del piano radioterapeutico migliore che permetta di soddisfare i vincoli di dose imposti. Raystation integra in un'unica piattaforma tecniche di machine learning e deep learning per il contouring automatico e semi-automatico dei volumi anatomici e permette di ottenere piani di trattamento con tecniche di IMRT, VMAT, 3D-CRT. Implementa algoritmi di ottimizzazione multi-criterio (MCO), con i quali è possibile esplorare tutti i possibili piani che soddisfano le condizioni di vincolo impostate e dà la possibilità al medico di scegliere il piano più adatto allo specifico caso clinico.

2.3 K-Nearest Neighbor

L'algoritmo k-Nearest Neighbor è un metodo di classificazione supervisionato, utilizzato nel riconoscimento di pattern. La determinazione della classe di un elemento è ottenuta sulla base della somiglianza di quest'ultimo ai k elementi vicini. Infatti,

all'elemento è assegnata la classe che è più rappresentata dai k vicini. Il classificatore necessita di due set di dati: il training set, composto dagli elementi di cui già si conosce la classe e il test set, in cui si ha l'elemento da classificare. L'algoritmo prevede:

1. Impostare un valore di k
2. Calcolare le distanze tra l'elemento da classificare e gli elementi del training set. La metrica più utilizzata è la distanza euclidea
3. Ordinare le distanze in ordine crescente
4. Selezionare le prime k distanze a cui corrispondono k elementi del training set
5. Associare all'elemento la classe più ricorrente tra le classi dei k elementi selezionati

La scelta del valore di k influenza la bontà della classificazione. Confrontando le classificazioni ottenute per diversi valori di k , è possibile selezionare quello che permette di ridurre gli errori di classificazione e il numero di elementi non classificati. Generalmente, se viene scelto un k troppo piccolo, la predizione è influenzata dal rumore dei dati, mentre se k è troppo grande, è più probabile che l'elemento venga classificato nella classe che presenta il numero maggiore di elementi (aumentano gli errori di classificazione).

2.4 Clustering

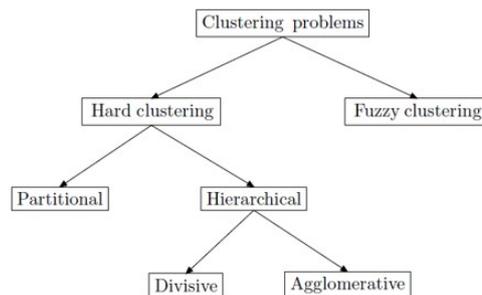


Figura 2.1: Diagramma del Clustering [6]

Il clustering dei dati [6],[3] è un insieme di tecniche di data mining che permettono di suddividere un dataset in gruppi, detti cluster, in modo che gli elementi appartenenti ad un cluster siano simili tra loro e diversi dagli elementi di altri cluster. Ciò che è alla base del raggruppamento è dunque una misura di similarità.

2.4.1 Misure di similarità

A seconda della tipologia di dati da raggruppare, sono rese disponibili diverse metriche per il calcolo delle “distanze” all’interno e all’esterno dei cluster. Trattando dataset di tipo numerico, la metrica più utilizzata è la distanza euclidea, definita:

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

Altre metriche utilizzabili sono:

- Distanza di Manhattan (“cityblock distance” o L1 norm)

$$d(x, y) = \sum_{j=1}^d |(x_j - y_j)| \quad (2.2)$$

- Distanza massima (o L_∞ norm)

$$d(x, y) = \max_{1 \leq j \leq d} |(x_j - y_j)| \quad (2.3)$$

- Distanza di Mahalanobis

$$d(x, y) = \sqrt{(x - y) \sum (x - y)^T} \quad (2.4)$$

- Distanza di Minkowsky (generalizzazione della distanza euclidea)

$$d(x, y) = \left(\sum_{j=1}^d (x_j - y_j)^p \right)^{1/p} \quad (2.5)$$

In generale, un problema di clustering può essere affrontato seguendo due diversi approcci (Figura 2.1): l’hard clustering e il fuzzy clustering. Nell’ hard clustering o clustering esclusivo ogni elemento del dataset può appartenere ad un solo cluster, mentre nel fuzzy clustering o clustering non esclusivo, i cluster possono avere elementi in comune; infatti, un elemento può far parte, con un certo grado di appartenenza, di più di un cluster. Gli algoritmi di clustering esclusivo, a loro volta, possono dividersi in clustering partizionale e clustering gerarchico.

2.4.2 Clustering gerarchico

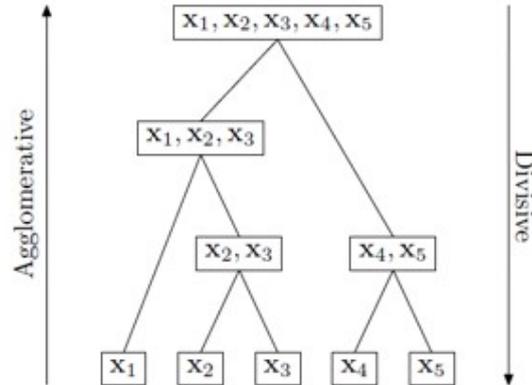


Figura 2.2: Clustering gerarchico e divisivo [6]

In un algoritmo di clustering gerarchico, il set di dati è diviso in una sequenza di partizioni annidate, senza la necessità di sapere a priori il numero di cluster che si vogliono ottenere. È possibile distinguere due tipologie di clustering gerarchico: agglomerativo e divisivo, entrambe di tipo iterativo. Lo schema dei due processi è rappresentato nella figura 2.2.

Il metodo divisivo è definito top-down poiché alla prima iterazione tutti gli elementi sono uniti in un unico cluster. Ad ogni iterazione, si ha la divisione in cluster più piccoli. Il metodo agglomerativo, invece, è definito bottom-up in quanto alla prima iterazione si hanno tanti cluster quanti sono gli elementi da raggruppare. Ad ogni iterazione, per mezzo di una misura di similarità, si calcola la distanza tra i cluster e si uniscono quelli la cui distanza è minima. Il processo di unione prosegue fino all'ottenere un unico cluster contenente tutti gli elementi.

I metodi di annidamento utilizzati sono:

- Single link (legame singolo): la distanza tra due cluster (C e C') è definita come la minima tra tutte le distanze calcolabili a due a due degli elementi dei due gruppi.

$$D(C, C') = \min_{x \in C, y \in C'} d(x, y) \quad (2.6)$$

Con questo metodo si predilige la differenza tra i cluster piuttosto che l'omogeneità degli elementi all'interno del gruppo.

- Complete link (legame completo): la distanza tra due cluster (C e C') è definita come la massima tra tutte le distanze calcolabili a due a due degli elementi dei due gruppi.

$$D(C, C') = \max_{x \in C, y \in C'} d(x, y) \quad (2.7)$$

A differenza del legame singolo, con questo metodo si dà maggiore importanza all'omogeneità degli elementi all'interno del cluster.

- Averaged link (legame medio): la distanza tra due cluster (C e C') è definita come la media di tutte le distanze calcolabili a due a due degli elementi dei due gruppi.

$$D(C, C') = \frac{1}{|C||C'|} \sum_{x \in C, y \in C'} d(x, y) \quad (2.8)$$

- Centroid (metodo del centroide): la distanza tra due cluster è definita come la distanza tra i centroidi dei due gruppi.

La metrica del calcolo delle distanze può essere scelta tra quelle precedentemente elencate.

L'indice di valutazione dell'omogeneità interna al cluster è la variabilità intra-cluster definita come la media delle distanze di ciascun elemento del cluster dal centroide del cluster stesso:

$$VarIntra_j = \frac{\sum_{elemento \in cluster_j} dist(elemento, c_j)}{n_j} \quad (2.9)$$

La distanza inter-cluster, invece, permette di valutare l'eterogeneità esterna ai cluster ed è ottenuta calcolando le distanze tra tutte le coppie dei centroidi.

2.4.3 Dendrogramma

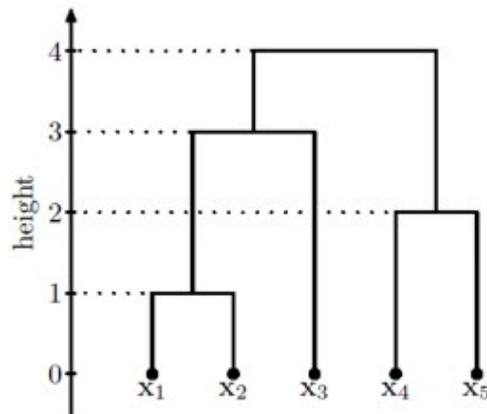


Figura 2.3: Esempio di dendrogramma [6]

Il dendrogramma è un diagramma ad albero (Figura 2.3) utilizzato per la visualizzazione del processo di clusterizzazione ottenuto mediante un algoritmo gerarchico. Sull'asse delle ascisse sono collocati tutti gli elementi del dataset, mentre l'asse

delle ordinate indica la distanza alla quale gli elementi o cluster sono uniti. Il primo step per la costruzione del dendrogramma è il calcolo della matrice delle distanze attraverso l'utilizzo del metodo e della metrica stabiliti. Tutti gli elementi del dataset sono rappresentati visivamente nella parte inferiore del dendrogramma. Seguendo un processo di agglomerazione, ad ogni iterazione (livello di annidamento nel dendrogramma) vengono uniti gli elementi a formare i cluster con conseguente aggiornamento della matrice delle distanze. Il procedimento si ripete fino a quando non si ottiene il singolo cluster che contiene tutti gli elementi, l'apice del dendrogramma.

Per visualizzare un raggruppamento con uno specifico numero di cluster è necessario tagliare il dendrogramma ad un certo livello dell'albero. Il taglio può essere effettuato ad un valore di distanza scelto dall'utente oppure laddove il gap tra due unioni successive è il maggiore ("metodo della distanza massima").

2.4.4 Normalizzazione dei dati

Quando si dispone di un set di dati numerici, i cui elementi sono caratterizzati da variabili con range molto diversi tra loro, è preferibile standardizzare i dati in quanto il calcolo delle distanze è influenzato dalle differenze di ampiezza degli intervalli delle variabili. Un metodo di standardizzazione è il min-max scaling:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.10)$$

dove x è il vettore di valori assunti dalla variabile, $\min(x)$ e $\max(x)$ sono rispettivamente il minimo e il massimo valore che essa assume. In questo modo x' sarà definito nell'intervallo $[0,1]$.

Un'alternativa è la normalizzazione media:

$$x' = \frac{x - \text{mean}(x)}{\max(x) - \min(x)} \quad (2.11)$$

in cui le variabili sono riscalate in un intervallo $[-1,1]$.

Si può utilizzare anche la standardizzazione o normalizzazione Z-score in cui si ridimensionano i dati in modo che questi assumano le proprietà di una distribuzione gaussiana di media nulla e deviazione standard pari ad 1.

$$x' = \frac{x - \mu}{\sigma} \quad (2.12)$$

dove μ è la media dei valori della variabile considerata e σ la sua deviazione standard.

2.5 Rete Neurale a Base Radiale (RBFNN)

Le reti neurali a base radiale sono delle reti neurali artificiali supervisionate di tipo feedforward, ovvero reti neurali in cui l'informazione procede univocamente dall'input all'output. Nascono con l'intento di risolvere problemi di approssimazione di curve in uno spazio ad alta dimensione. Alla base del principio di funzionamento delle RBFNN vi è il teorema di Cover, nel quale si afferma che *“un problema di classificazione complesso, formulato attraverso una trasformazione non-lineare dei dati in uno spazio ad alta dimensionalità, ha maggiore probabilità di essere linearmente separabile che in uno spazio a bassa dimensionalità.”*

2.5.1 Architettura della rete

Una rete neurale a base radiale è costituita da 3 layer: il layer di input, il layer nascosto e il layer di output. Ogni layer è formato da uno specifico numero di neuroni. I collegamenti tra i neuroni di input e i neuroni del layer nascosto sono rappresentati da archi non pesati e la trasformazione che li caratterizza è non lineare, mentre i collegamenti tra i neuroni del layer nascosto e i neuroni di output sono archi pesati e la trasformazione da uno spazio all'altro è lineare. Ogni neurone del layer nascosto implementa una funzione a base radiale (Φ) (RBF), ovvero una funzione a valori reali che dipende solamente dalla distanza tra l'input e un punto fisso, chiamato centro (μ), unico massimo della funzione. La distanza è calcolata come norma euclidea ($(\|\bullet\|)_2$).

Oltre al centro, alcune RBF sono caratterizzate dal parametro σ che definisce, per ogni unità nascosta, l'ampiezza del campo ricettivo della funzione nello spazio di input.

L'output del neurone del layer nascosto indica quanto l'input (z) è prossimo al centro della funzione a base radiale.

Esistono diverse tipologie di funzioni a base radiale:

- Lineare

$$\Phi(\|z - \mu\|_2) = \|z - \mu\|_2 \quad (2.13)$$

- Cubica

$$\Phi(\|z - \mu\|_2) = \|z - \mu\|_2^3 \quad (2.14)$$

- Multiquadratica

$$\Phi(\|z - \mu\|_2, \sigma) = \sqrt{\|z - \mu\|_2^2 + \sigma^2} \quad (2.15)$$

- Multiquadratica inversa

$$\Phi(\|z - \mu\|_2, \sigma) = \frac{1}{\sqrt{\|z - \mu\|_2^2 + \sigma^2}} \quad (2.16)$$

- Gaussiana

$$\Phi(\|z - \mu\|_2, \sigma) = e^{-\frac{\|z - \mu\|_2^2}{2\sigma^2}} \quad (2.17)$$

- Logistica

$$\Phi(\|z - \mu\|_2, \sigma) = \frac{1}{1 + e^{-\frac{\|z - \mu\|_2^2}{2\sigma^2} - \theta}} \quad (2.18)$$

I neuroni di output, invece, implementano una funzione di attivazione lineare, dunque l'output della rete è una combinazione lineare delle funzioni a base radiale, in cui i pesi w sono i coefficienti.

La costruzione di una rete RBFNN richiede tre set di dati:

- Training set, per allenare la rete
- Validation set, per valutare l'andamento dell'errore quadratico medio durante l'allenamento in modo da sapere quando fermarlo
- Test set, per verificare le performance della rete

2.5.2 Algoritmi di apprendimento

Sono stati sviluppati diversi algoritmi di apprendimento per allenare una rete RBFNN, alcuni in cui vengono modificati solo i pesi tra il layer nascosto e il layer di output (Algoritmo dei centri fissi) e altri in cui si aggiustano più parametri quali i pesi, i centri e le deviazioni (Algoritmo dei centri adattivi).

Il training con i centri fissi prevede di selezionare i centri in maniera random dal training set e mantenerli fissi durante l'addestramento. Le deviazioni sono calcolate, nel caso di RBF gaussiana come

$$\sigma = \frac{d_{max}}{\sqrt{J}} \quad (2.19)$$

dove d_{max} è la massima distanza euclidea tra i centri e J è il numero di neuroni del layer nascosto.

I pesi sono inizializzati in modo che essi assumano valori random piccoli. Successivamente è calcolato l'output per ogni unità di output tramite l'equazione 2.20 e modificati i pesi utilizzando l'equazione 2.21.

$$o = \sum_{j=1}^J w_j y_j \quad \text{dove} \quad y_j = \Phi(\|z - \mu_j\|_2) \quad (2.20)$$

$$w = (\Phi^T \Phi)^{-1} \Phi^T t \quad (2.21)$$

dove Φ è la matrice delle funzioni RBF e t il vettore dei targets.

Il training con il gradiente discendente, invece, prevede gli stessi step di inizializzazione dei centri, delle deviazioni e dei pesi, ma durante l'addestramento, questi

parametri vengono aggiornati in funzione dell'errore tra l'output della rete e il target, scalato rispetto ai rispettivi learning rate η_c , η_σ ed η_w fino a quando non si raggiunge una condizione di stop, la quale può essere legata al valore dell'errore o al raggiungimento del numero totale di epoche di addestramento [5].

$$w(t+1) = w(t) - \eta_w \frac{\partial E}{\partial w} \quad (2.22)$$

$$\mu(t+1) = \mu(t) - \eta_c \frac{\partial E}{\partial \mu} \quad (2.23)$$

$$\sigma(t+1) = \sigma(t) - \eta_\sigma \frac{\partial E}{\partial \sigma} \quad (2.24)$$

Un'altra tipologia di allenamento con i centri adattivi è il training Two-Phase, in cui si possono distinguere due fasi: l'allenamento non supervisionato dei centri, ottenibile con il k-means e l'allenamento supervisionato dei pesi, utilizzando il metodo del gradiente discendente.

Il k-means è un algoritmo di clustering partizionale, in cui la scelta del numero di cluster che devono essere generati è fatta a priori. Con un meccanismo di tipo iterativo viene definita l'appartenenza di ogni elemento del dataset ad un gruppo tramite la misura di una distanza tra l'elemento stesso e un punto rappresentativo del cluster.

L'allenamento della rete viene interrotto prima che le prestazioni sul validation set inizino a peggiorare, ovvero quando si raggiunge il minimo nella curva dell'errore quadratico medio (mse) relativa al set di validazione.

2.6 Algoritmi Genetici

Gli algoritmi genetici sono un metodo euristico di ottimizzazione che, nella sua costruzione, prende esempio dalla teoria evolutiva darwiniana. Egli, infatti, afferma che l'evoluzione della specie si basa sul principio della selezione naturale, secondo cui l'organismo più adattabile al contesto di sviluppo prevale su quello più debole. Holland propose l'algoritmo genetico come un metodo in grado di emulare questo principio. Per questo motivo, rientrano nell'insieme di metodi raccolti sotto il nome di Evolutionary Computing (EC).

Si tratta di un algoritmo stocastico, dunque non è garantito l'ottenimento della soluzione migliore del problema ma, generalmente, si arriva a trovare delle soluzioni definibili buone.

Un problema di ottimizzazione consiste nell'individuare tutte le possibili soluzioni e tra queste, selezionare quella più adatta alla risoluzione dello stesso.

In un algoritmo genetico, si parte da una popolazione iniziale di soluzioni, le quali sono soggette ad operatori di selezione e riproduzione durante un processo iterativo

che simula l'evoluzione delle soluzioni verso quella migliore.

Nel GA, ogni soluzione è rappresentata attraverso un vettore binario, definito cromosoma. La codifica/decodifica della soluzione può essere di diverse tipologie, Yes/No Variables, Codified Variables o Integer Variables e la si sceglie in base al tipo di problema da affrontare.

Ad ogni soluzione è associato un valore di fitness, ovvero una funzione che permette di valutarne la bontà. L'obiettivo dell'algoritmo è quello di arrivare all'ottimo, ovvero la soluzione che minimizza la fitness.

Definite la funzione obiettivo e la codifica delle soluzioni, l'algoritmo prevede la definizione della popolazione iniziale, ovvero l'insieme di soluzioni che, per prime, saranno sottoposte ad analisi. Generalmente viene creata selezionando randomicamente delle soluzioni dallo spazio di ricerca. Ha un ruolo importante la dimensione della popolazione. È necessario trovare un compromesso tra il tempo computazionale e la volontà di avere una popolazione che sia il più possibile rappresentativa dell'intero insieme di soluzioni possibili.

Gli operatori di selezione permettono di individuare, a partire dalla popolazione iniziale, il sottoinsieme di soluzioni (genitori) che vengono fatte riprodurre per generarne delle nuove (figlie) per mezzo degli operatori di riproduzione. Inoltre, sono utilizzati anche nella fase successiva alla riproduzione, per creare la nuova popolazione da cui far partire un nuovo ciclo "evolutivo", la quale può essere formata da sole soluzioni figlie o da un insieme di soluzioni figlie e genitori.

Le soluzioni figlie sono generate grazie agli operatori di crossover e mutazione. Il crossover consiste nel generare una nuova soluzione ricombinando i geni di una coppia di soluzioni. Il crossover può essere:

- ad un punto di taglio: si sceglie un punto in cui dividere il cromosoma e le due parti di ciascuna soluzione vengono scambiate,
- a due punti di taglio: si scelgono due punti di divisione e le parti centrali vengono scambiate,
- uniforme: i geni da combinare sono scelti in maniera random

L'operatore di mutazione, invece, consiste nell'alterare il valore dei geni di una soluzione genitore. Se con il crossover, si generano soluzioni vicine a quelle di partenza, con la mutazione, si ha la possibilità di ampliare l'esplorazione delle soluzioni dell'intero spazio di ricerca. Gli operatori di riproduzione sono regolati da due parametri, la probabilità di crossover (pc) e la probabilità di mutazione (pm), che definiscono la frequenza con cui le soluzioni subiscono crossover e mutazione. Generalmente, si sceglie un valore alto per pc e uno basso per pm .

Un algoritmo genetico prevede dunque l'iterazione dei seguenti step:

1. Generazione della popolazione iniziale di soluzioni

2. Assegnazione dei valori di fitness alle soluzioni della popolazione per valutarne la bontà
3. Selezione delle soluzioni da sottoporre agli operatori di riproduzione
4. Generazione della nuova popolazione a partire dalle soluzioni figlie ottenute

Il ciclo si interrompe al raggiungimento della condizione di stop, ovvero quando si sono raggiunte il numero massimo di iterazioni o quando non si verificano più cambiamenti nelle soluzioni.

Capitolo 3

Elaborazione delle immagini TC

Per ciascun paziente, si ha a disposizione una documentazione di tipo dicom comprensiva di slices TC addome, ovvero immagini bidimensionali di dimensioni 512x512 (coordinate x e y) e un file denominato “RS”, in cui sono salvate le coordinate dei contorni dei distretti anatomici di interesse, tra cui il CTV, il PTV, la vescica, il retto. I nomi delle slices presentano la coordinata z a cui fanno riferimento.

3.1 Data Cleaning

È stato effettuato un primo controllo circa la corretta lettura dei files da parte del software Matlab, con l'intento di evitare l'utilizzo di pazienti la cui documentazione presentasse immagini corrotte o assenza dei contorni necessari allo sviluppo delle successive analisi. A fronte dei risultati ottenuti, il dataset si compone di 202 pazienti dei 254 a disposizione.

3.2 Estrazione delle maschere

I parametri relativi alla geometria della prostata sono stati estratti a partire dalla maschera binaria del volume di interesse di dimensioni 512x512xN dove N è il numero di slices in cui il volume è presente. Ogni sottomatrice di ciascuna maschera rappresenta una slice, ovvero una sezione trasversale bidimensionale del volume. La maschera è un'immagine i cui pixel possono assumere solamente due valori, 0 e 1. La regione che appartiene al volume sarà bianca (1), mentre tutto ciò che non fa parte di esso sarà di colore nero (0). Per ottenerla, è stato necessario implementare diversi step di processing.

In questa analisi, il volume d'interesse è il CTV, che nel caso del tumore della prostata coincide con il GTV. Tra le ROI contornate, spesso, è possibile trovare diverse nomenclature per il CTV e in alcuni casi ne sono presenti di due tipologie, CTV1.1 e CTV1.2. Ciò che li diversifica è la dimensione del volume. Infatti, uno dei due volumi è più grande perché contiene anche altre strutture anatomiche quali, ad esempio, le vescicole seminali.

È stato implementato un algoritmo in grado di estrarre, per ogni paziente, la maschera binaria del CTV con il volume più piccolo, nell'ipotesi che la struttura anatomica principale contornata sia la prostata, in modo da ottenere parametri che ne definiscano la geometria, limitando su di essi l'influenza di altre strutture vicine.

3.3 Analisi geometrica delle maschere estratte

Una volta estratte le maschere binarie, è stata eseguita un'analisi della geometria delle varie sezioni della prostata. Essendo la ghiandola prostatica di forma ovale, sono state confrontate le dimensioni della base e dell'altezza del Bounding Box delle regioni presenti in ognuna delle slices. Il Bounding Box è la regione rettangolare più piccola che racchiude l'oggetto presente nella maschera. Confrontare queste lunghezze, definite in termini di numero di pixel che formano il contorno, permette di capire se ci sono sezioni con dimensioni che si discostano particolarmente dalle corrispettive sezioni nelle maschere degli altri pazienti e che per questo motivo, necessitano di analisi più approfondite.

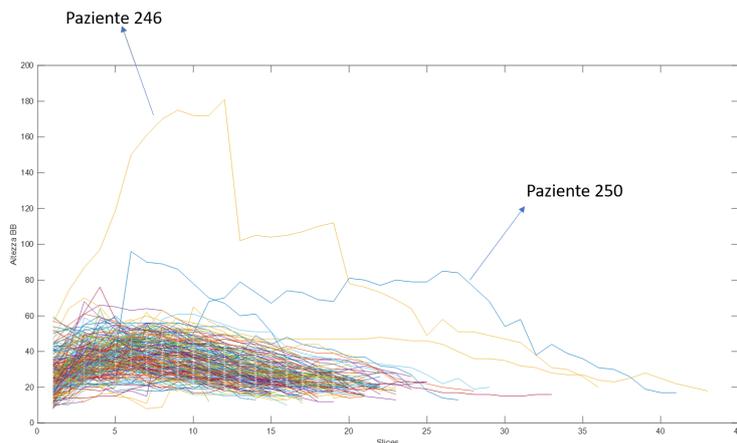


Figura 3.1: Dimensioni in pixel dell'altezza dei Bounding Box delle regioni nelle slices dei 202 pazienti

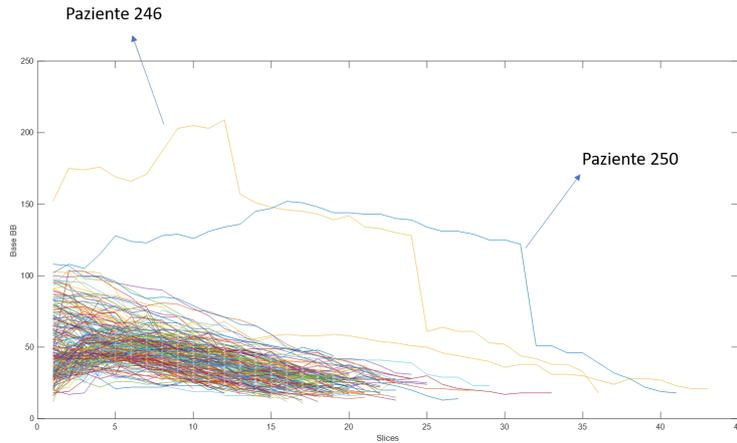


Figura 3.2: Dimensioni in pixel della base dei Bounding Box delle regioni nelle slices dei 202 pazienti

I grafici riportati in figura 3.1 e 3.2 mostrano, come ci si aspettava, un andamento discendente delle curve, in quanto le sezioni della prostata diminuiscono in area man mano che la coordinata z diminuisce. Tuttavia, permettono di identificare, in maniera evidente, la presenza di alcune sezioni del volume caratterizzate da dimensioni che si discostano molto dalle altre, in particolare nei pazienti 246 e 250.

A seguito di un'analisi visiva delle slices dei pazienti, si è visto che, pur avendo

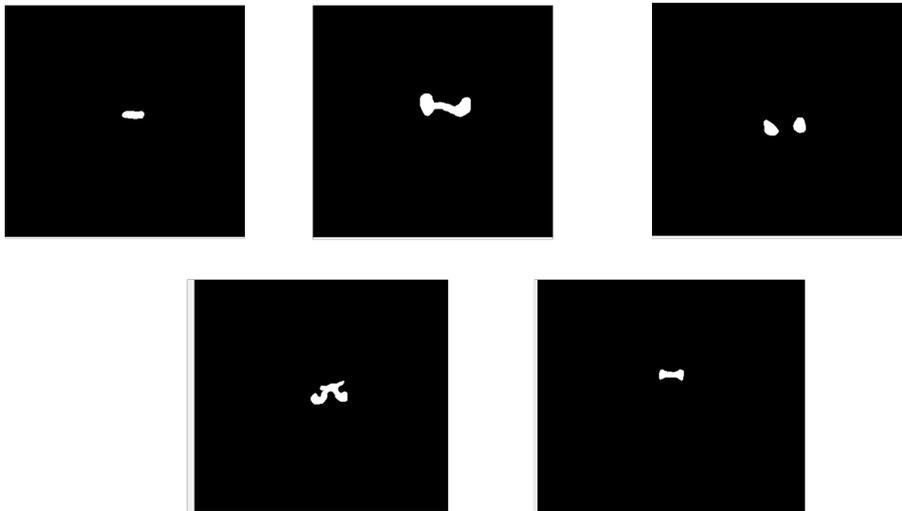


Figura 3.3: Esempi di slices anomale

selezionato il CTV più piccolo, in alcuni di essi sono comunque presenti altre strutture anatomiche, come evidente nelle maschere riportate in figura 3.3.

Sono stati, inoltre, identificati, per ogni slice, gli indici dei pazienti che sono considerati outliers rispetto alla distribuzione dei valori di base e altezza dei Bounding Box.

3.4 Analisi della forma e estrazione di nuove maschere

Per ottenere parametri geometrici più veritieri, è stato implementato un algoritmo che, per ciascun paziente, estrae dalla maschera del volume solamente le slices le cui sezioni hanno eccentricità minore di 0.9 e circolarità maggiore di 0.4, facendo attenzione che queste siano consecutive. Regioni caratterizzate da un'elevata eccentricità presentano una forma allungata e assottigliata, non tipica del volume di interesse, mentre l'indice di circolarità permette di non considerare quelle slices le cui regioni hanno una forma non allungata ma particolarmente complessa e diversa dalla forma della prostata.

Delle 202 maschere iniziali, 116 sono state modificate andando ad eliminare le sezioni che non rispettavano i controlli di forma imposti, mentre 86 non hanno subito variazioni.

Tra le 116 modificate, le maschere dei pazienti 107, 44 e 148 sono state escluse dalle successive analisi perchè alcune delle loro sezioni, pur avendo forme anomale, non vengono eliminate.

I grafici riportati in figura 3.4,3.5,3.6 e 3.7 permettono di confrontare le dimensioni dell'altezza e della base dei Bounding Box delle regioni delle maschere prima e dopo la modifica (curve colorate), con il valor medio e deviazione standard dell'altezza e della base delle regioni delle maschere che non necessitano di correzioni (curva verde). E' evidente come, dopo aver eliminato le slices anomale, le curve sono più vicine all'intervallo individuato dalla distribuzione media.

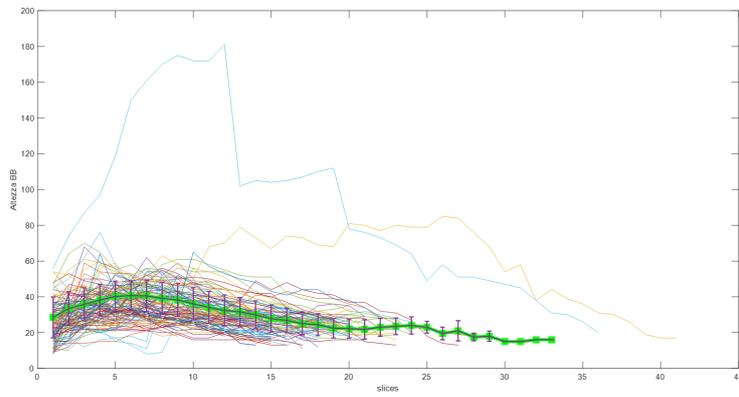


Figura 3.4: Confronto tra le dimensioni dell'altezza del BB delle 113 maschere complete e la media delle dimensioni dell'altezza del BB delle 86 maschere corrette

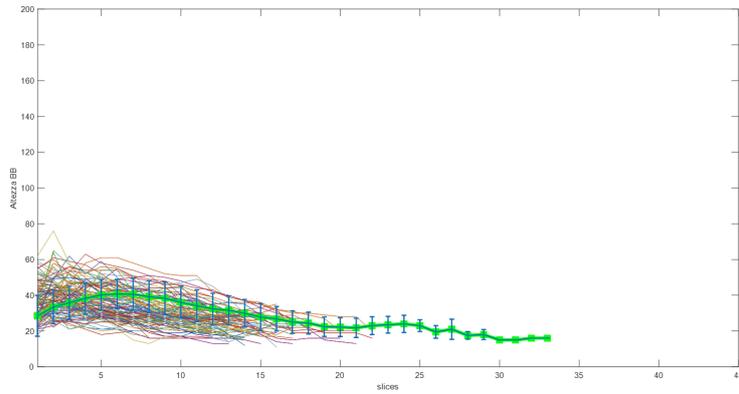


Figura 3.5: Confronto tra le dimensioni dell'altezza del BB delle 113 maschere modificate e la media delle dimensioni dell'altezza del BB delle 86 maschere corrette

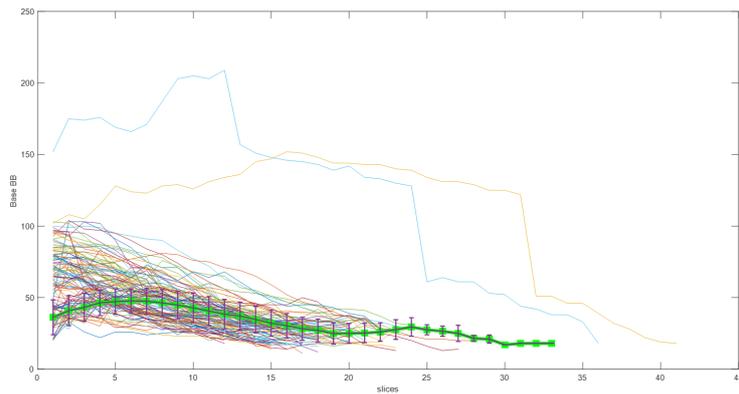


Figura 3.6: Confronto tra le dimensioni della base del BB delle 113 maschere complete e la media delle dimensioni della base del BB delle 86 maschere corrette

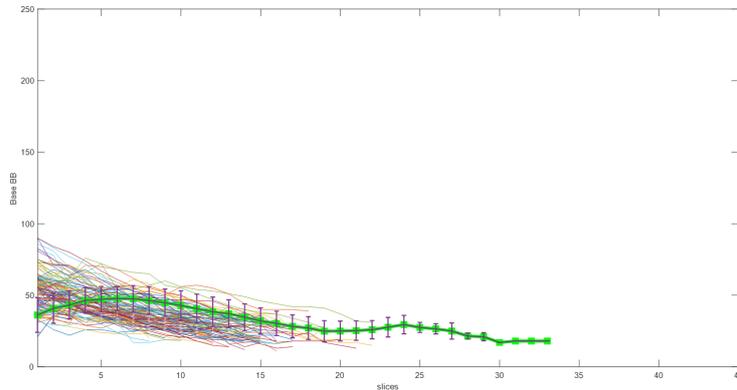


Figura 3.7: Confronto tra le dimensioni della base del BB delle 113 maschere modificate e la media delle dimensioni della base del BB delle 86 maschere corrette

3.5 Estrazione dei parametri geometrici

Ottenute le nuove maschere per i 199 pazienti, è stato possibile ricavare una matrice che contiene i valori dei parametri geometrici tridimensionali e bidimensionali implementati.

I parametri relativi alla geometria tridimensionale del volume di interesse sono stati ottenuti a partire dalla funzione “*regionprops3*”, già implementata in Matlab. Di seguito, è presentato un elenco dei parametri utilizzati.

- Lunghezza (b) del Bounding Box
- Altezza (h) del Bounding Box
- Profondità (s) del Bounding Box
- Superficie del Bounding Box

$$S_b = 2 \cdot (b \cdot h + h \cdot s + b \cdot s) \quad (3.1)$$

- Volume del Bounding Box

$$V_b = b \cdot h \cdot s \quad (3.2)$$

Il Bounding Box è il più piccolo cuboide che contiene il volume.

- Convex Volume: volume del più piccolo poligono convesso che contiene la regione di interesse

- Diametro equivalente: diametro di una sfera avente lo stesso volume della regione di interesse

$$D_{eq} = \left(\frac{6V}{\pi}\right)^{\frac{1}{3}} \quad (3.3)$$

- Lunghezza assi principali: lunghezza in voxel degli assi maggiori dell'ellissoide che ha lo stesso secondo momento centrale normalizzato della regione di interesse
- Superficie: stima della superficie della regione di interesse
- Volume: numero di voxel che compongono la regione di interesse I parametri implementati in studi trovati in letteratura sono:
- Rapporto Superficie/Volume

$$\frac{S}{V} \quad (3.4)$$

- PCAR
- Valore minimo, massimo e medio delle basi del Bounding Box 2D delle sezioni di prostata
- Valore minimo, massimo e medio delle altezze del Bounding Box 2D delle sezioni di prostata.

Le dimensioni del Bounding Box 2D sono uno degli output della funzione “*region-props*” di Matlab. Ai 20 parametri implementati, si aggiunge una colonna che contiene il numero identificativo del paziente.

Nella tabella 3.1 è riportata l'associazione tra le variabili e i parametri estratti.

Tabella 3.1: Parametri geometrici implementati

Variabile	Parametro Geometrico
1	Id Paziente
2	Lunghezza (b) 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

Capitolo 4

Clustering dei dati

4.1 Risultati del clustering

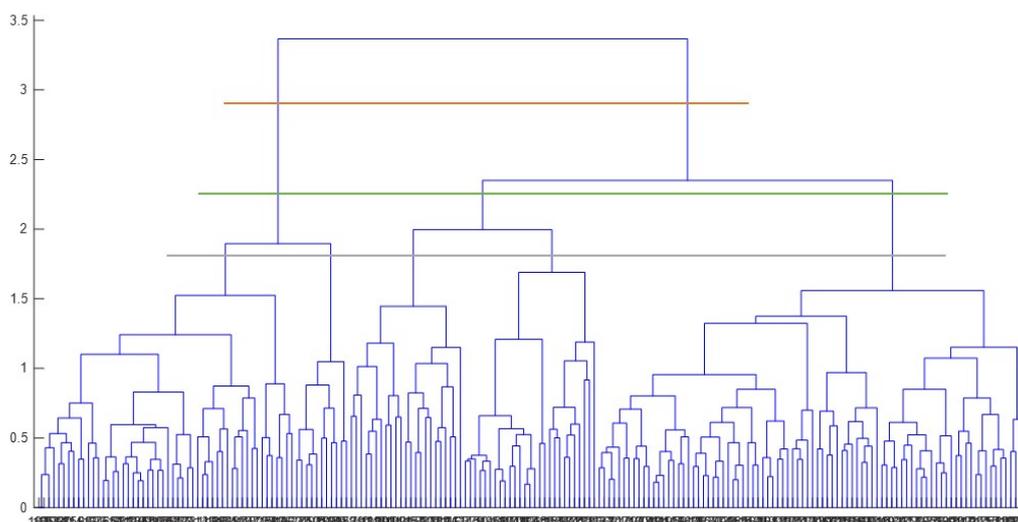


Figura 4.1: Dendrogramma con tagli a 2,3 e 5 cluster

Uno degli obiettivi del lavoro di tesi è capire se sia possibile suddividere i pazienti in sottogruppi, ognuno dei quali caratterizzato da proprietà geometriche simili della prostata.

È stato, dunque, utilizzato un algoritmo gerarchico agglomerativo per eseguire il clustering dei dati geometrici a disposizione. I dati, prima di essere processati, sono stati normalizzati con il metodo del min-max scaling, in quanto i range di definizione dei parametri sono molto diversi tra loro.

Per ottenere la suddivisione, si è scelto il metodo del legame completo e, come metrica, la distanza euclidea. Il dendrogramma, riportato in figura 4.1, ha permesso la visualizzazione del processo.

Si è scelto di analizzare tre diversi raggruppamenti dei pazienti, in particolare a due, tre e cinque cluster. Nelle tabelle 4.1, 4.2 e 4.3 sono riportate le suddivisioni dei pazienti.

Tabella 4.1: Clusterizzazione a 2 cluster

Cluster	Numero Pazienti
1	63
2	136

Tabella 4.2: Clusterizzazione a 3 cluster

Cluster	Numero Pazienti
1	86
2	50
3	63

Tabella 4.3: Clusterizzazione a 5 cluster

Cluster	Numero Pazienti
1	11
2	52
3	23
4	27
5	86

Per confrontare le tre clusterizzazioni, per ognuna di esse, sono state calcolate le variabilità intra-cluster e i box plot di ciascun parametro, riportati nell'appendice A. Questi ultimi permettono di valutare la distribuzione dei valori di ognuno dei parametri all'interno dei cluster relativi alle diverse clusterizzazioni. Le variabilità

intra-cluster sono state calcolate come media delle distanze di ciascun elemento dal centroide del proprio cluster e come media delle distanze delle coppie degli elementi del cluster.

Da un'analisi dei box plot in figura 4.2, emerge che nella suddivisione a tre cluster, le variabili hanno distribuzioni ben separate tra loro, mentre aumentando il numero di cluster, i box (distribuzioni comprese tra il 25° e il 75°) si sovrappongono. Questo significa che gli elementi nei diversi cluster sono simili.

Da ciò deriva la scelta di proseguire il lavoro con la suddivisione dei pazienti in

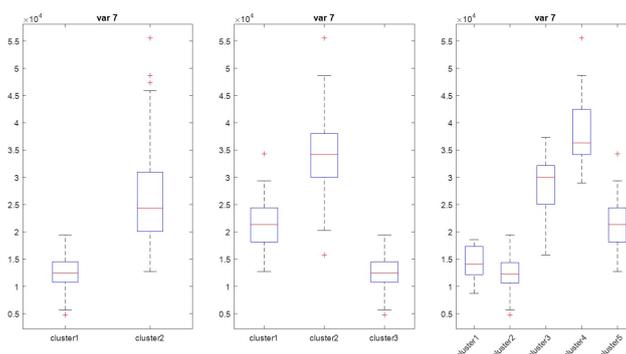


Figura 4.2: Box plot delle distribuzioni del parametro 7 nelle varie clusterizzazioni

tre gruppi.

All'interno di ogni raggruppamento, i pazienti sono divisi in due set, uno di training e uno di validation. Quest'ultimo è composto da 5 pazienti estratti randomicamente.

4.2 Associazione Paziente-Cluster

Scelto il numero di cluster, per validare la suddivisione e per associare un nuovo paziente al cluster che meglio lo rappresenta, sono stati valutati due metodi:

1. Associazione paziente-cluster con distanza paziente-centroide

Sono state valutate quattro combinazioni centroide-distanza. Per ognuno dei tre cluster è stato calcolato il centroide come media e mediana dei valori delle variabili di ciascun paziente, mentre le distanze sono state calcolate utilizzando la distanza euclidea e la distanza "L1 norm" (cityblock). Per ogni paziente del dataset e per ogni combinazione è stata calcolata la sua distanza dai centroidi dei tre cluster e il paziente è stato associato al gruppo la cui distanza paziente-centroide è minima. Per valutare la bontà dell'associazione, per ogni combinazione, è stata costruita una confusion matrix mettendo a confronto il cluster reale del paziente con quello associato. Dai risultati ottenuti, si evince

che per le combinazioni mediana-euclidea, mediana-L1, media-euclidea, media-L1, le percentuali di corretta classificazione sono rispettivamente dell'87,4%, dell'85,4%, dell'86,8% e dell'85,9%. Nell'allegato B sono riportate le confusion matrix.

2. Associazione paziente-cluster con algoritmo k-nn

L'associazione paziente-cluster è stata ottenuta utilizzando un classificatore k-Nearest Neighbor. È stato applicato il "leave one out" su tutto il dataset di pazienti per determinare, di volta in volta, il set di training e il set di test. Il classificatore, infatti, è di volta in volta allenato con un training set formato da tutti i pazienti tranne il soggetto che costituisce il test set, di cui si vuole conoscere il cluster predetto. La prima prova è stata eseguita imponendo k uguale a 14 (circa la radice quadrata del numero totale di pazienti), successivamente si è scelto di valutare i risultati ottenuti variando k da 5 a 13 a passi di 1.

Anche in questo caso, il confronto tra i risultati è stato possibile grazie alla costruzione delle confusion matrix. Valutando le percentuali di corretta classificazione delle varie prove, il classificatore con k pari ad 11 è quello che permette di ottenere la più alta percentuale di corretti classificati (92,4% e nessun elemento non classificato (Tabella 4.4)). Nell'appendice B sono riportate le confusion matrix.

Confrontando i due metodi di associazione, si può affermare che l'associazione paziente-cluster attraverso l'utilizzo dell'algoritmo k-nn con k=11 è il metodo che permette di associare, con la più alta percentuale di corretta classificazione, nuovi pazienti ai cluster corrispondenti.

Per validare il metodo di associazione con k=11, si è allenato il classificatore con

Tabella 4.4: *Confusion Matrix k=11*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	83	5	5
	Cluster 2	2	43	0
	Cluster 3	1	2	58
	NC	0	0	0

un set di training costituito dall'insieme di tutti i pazienti che formano i rispettivi training set di ciascun cluster. In un primo caso il set di test coincide con il training set, mentre, nel secondo caso, è costituito dall'insieme di pazienti che formano i set di validation.

Nelle tabelle 4.5 e 4.6 sono riportate le confusion matrix. Le percentuali di corretta classificazione sono del 93,4% e dell'80% rispettivamente.

Tabella 4.5: *Confusion Matrix k=11 Test set=Training Set*
Cluster Dendrogramma

		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	79	4	5
	Cluster 2	2	40	0
	Cluster 3	0	1	53
	NC	0	0	0

Tabella 4.6: *Confusion Matrix k=11 Test set=Validation Set*
Cluster Dendrogramma

		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	4	1	0
	Cluster 2	0	3	0
	Cluster 3	1	1	5
	NC	0	0	0

Capitolo 5

Ottimizzazione dei piani di radioterapia a partire dal clustering dei pazienti

5.1 Rete RBFNN

A partire dalla clusterizzazione scelta, si prosegue analizzando un cluster alla volta. La rete RBFNN è implementata nella funzione di fitness dell'algoritmo genetico. L'allenamento della rete viene eseguito sui singoli pazienti che compongono il cluster, con l'intento di capire se è possibile determinare un set di parametri della rete che possa permettere di ottenere performance soddisfacenti per tutti i pazienti del cluster stesso.

5.1.1 Estrazione dei dati di input della rete tramite Raystation

Il training della rete è effettuato a partire da un set di dati estratti dal software TPS "Raystation". Per ogni paziente di ciascun cluster sono estratti 25 piani di radioterapia in cui sono imposte diverse combinazioni di MaxEUD alla vescica e al retto (si fanno variare dosi da 1000 a 5000 cGy a passi di 1000 cGy). La prescrizione al target è di 2.7 Gy, somministrati in 26 frazioni. Si è scelta la tecnica VMAT e l'acceleratore lineare LIN 4.

Il software restituisce, per ciascun piano, un valore di fitness che ne definisce la bontà. Minore è il valore, maggiore sarà la qualità del piano.

Nell'estrazione dei dati, il paziente 42 (Cluster 3) e il paziente 228 (Cluster 1) sono stati esclusi per la mancanza di informazioni necessarie alla corretta generazione dei piani radioterapeutici.

5.1.2 Allenamento della rete

Il set delle 25 coppie di dose (input) e relativa fitness (target) è suddiviso in:

- Training set (20 coppie) per l'allenamento della rete
- Validation set (5 coppie) per valutare quando interrompere l'addestramento.

L'algoritmo di apprendimento si basa sul metodo del Gradiente Discendente e le formule implementate per l'aggiornamento dei pesi, dei centri e delle deviazioni [12] sono rispettivamente:

$$w' = w + \frac{2}{M} \cdot \eta_w \cdot \epsilon \cdot \Phi(x, \mu, \sigma) \quad (5.1)$$

$$\mu' = \mu + \frac{2}{M} \cdot \eta_w \cdot \epsilon \cdot \Phi(x, \mu, \sigma) \cdot \frac{x - \mu}{\sigma} \quad (5.2)$$

$$\sigma' = \sigma + \frac{2}{M} \cdot \eta_w \cdot \epsilon \cdot \Phi(x, \mu, \sigma) \cdot \frac{\|x - \mu\|_2^2}{\sigma^3} \quad (5.3)$$

dove M è il numero di input, Φ è la RBF gaussiana e η_σ , η_c e η_w sono i learning rate.

L'inizializzazione dei centri è ottenuta con il metodo del k-means, con k pari al numero di neuroni imposto.

5.1.3 Tuning dei parametri

Si è scelto di effettuare l'allenamento della rete facendo variare il numero di neuroni e i learning rates. In particolare:

- *Neuroni*: da 8 a 20 a passi di 2
- η_w : da 0.1 a 0.5 a passi di 0.1
- η_c : da 0.1 a 0.5 a passi di 0.1
- η_σ : da 0.1 a 0.5 a passi di 0.1

5.1.4 Risultati Tuning RBFNN

Per ogni paziente di ciascun cluster, l'allenamento è stato ripetuto 10 volte. Dopo aver eseguito il tuning dei parametri, sono stati vagliati due metodi di selezione del set migliore. Il primo prevede un'analisi della frequenza dei singoli parametri, mentre il secondo consiste in un'analisi della distribuzione dell'errore quadratico medio in ciascuna delle 875 combinazioni ottenute.

5.1.5 Analisi dei singoli parametri

Per ciascun paziente, in ognuna delle 10 ripetizioni si ottengono 875 combinazioni, delle quali è selezionato il set di parametri che permette di avere il valore di mse più basso sul validation set.

Dunque si ottengono 10 combinazioni migliori, una per ogni ripetizione, ma tra queste, il set scelto per il paziente è quello che presenta l'mse più piccolo sempre sul validation set.

Sono stati messi a confronto il target, sia del training che del validation set, e i valori di fitness restituiti dalla rete in ciascuna delle 10 ripetizioni. Il grafico in figura 5.1 ne è un esempio. I restanti grafici sono riportati nell'appendice C. La

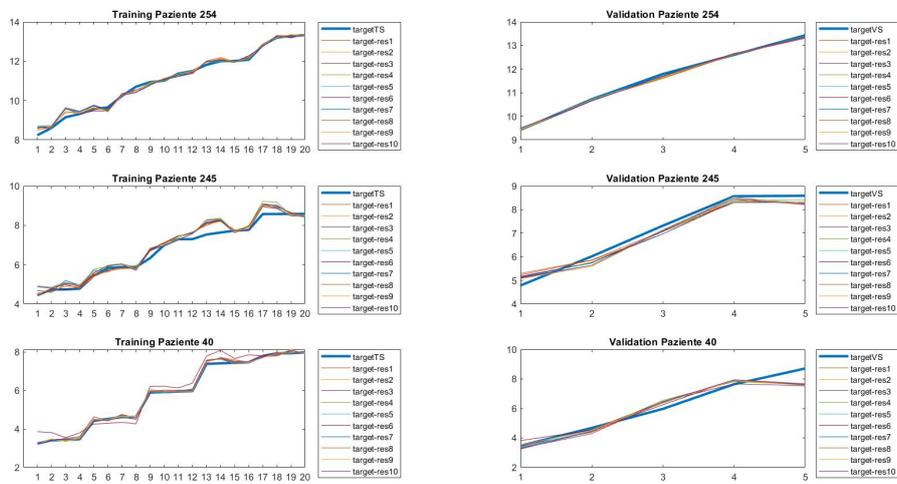


Figura 5.1: Pazienti 254-245-40

combinazione migliore è quella che mostra, nel grafico relativo al validation set di ciascun paziente, l'andamento più simile a quello del target.

Nell'appendice C, per ciascun cluster, si ha una tabella che, per ogni paziente, riporta la combinazione migliore.

Determinato il set migliore per ognuno dei pazienti del cluster, si è valutato, per ogni parametro, quello o quelli più ricorrenti e, a partire da essi, sono state definite le combinazioni che potrebbero essere applicabili a tutti i pazienti dello specifico cluster. I bar diagrams in figura 5.2, 5.3 e 5.4 rappresentano i risultati ottenuti rispettivamente per il cluster 1, 2 e 3. Nelle tabelle 5.1, 5.2 e 5.3, invece, sono riportate le combinazioni selezionate rispettivamente per il cluster 1, 2 e 3.

Cluster 1

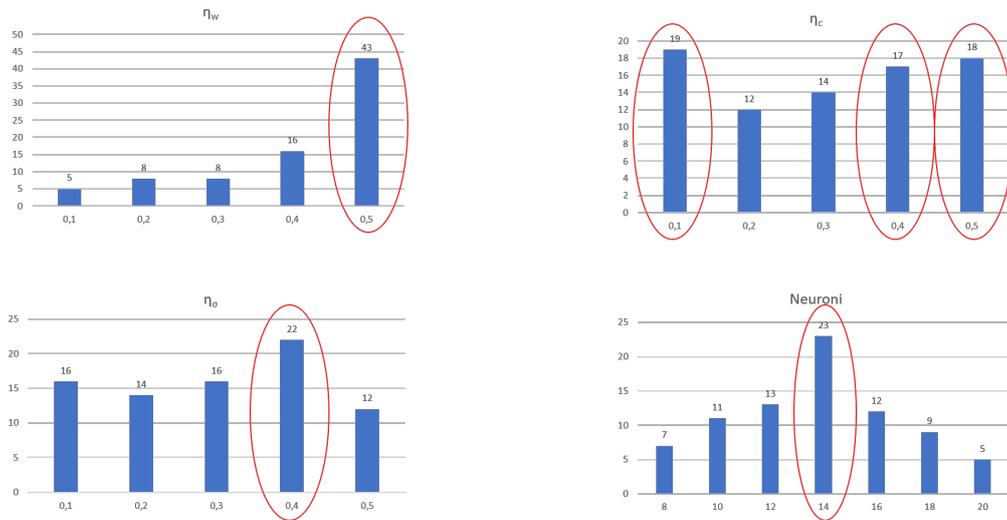


Figura 5.2: Bar Diagram dei parametri

Tabella 5.1: Combinazioni selezionate

Combinazione	Neuroni	η_w	η_c	η_σ
1	14	0,5	0,1	0,4
2	14	0,5	0,4	0,4
3	14	0,5	0,5	0,4

Cluster 2

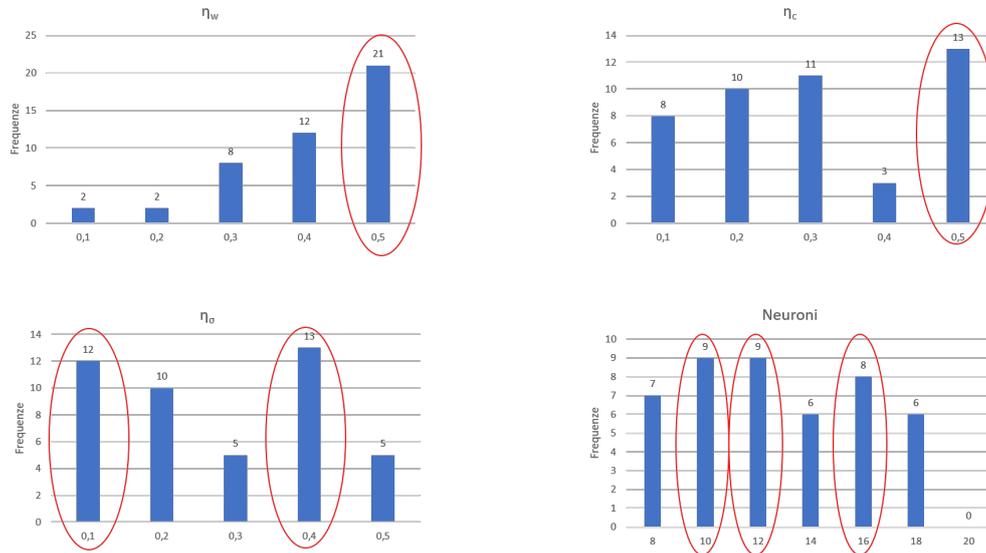


Figura 5.3: Bar Diagram dei parametri

Tabella 5.2: Combinazioni selezionate

Combinazione	Neuroni	η_w	η_c	η_σ
1	10	0,5	0,5	0,1
2	10	0,5	0,5	0,4
3	12	0,5	0,5	0,1
4	12	0,5	0,5	0,4
5	16	0,5	0,5	0,1
6	16	0,5	0,5	0,4

Cluster 3

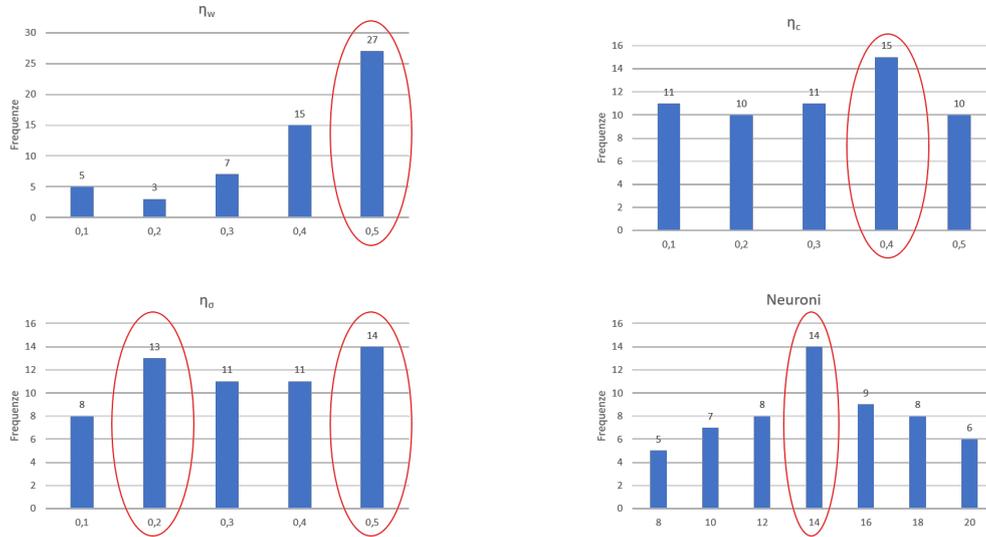


Figura 5.4: Bar Diagram dei parametri

Tabella 5.3: Combinazioni selezionate

Combinazione	Neuroni	η_w	η_c	η_σ
1	14	0,5	0,4	0,2
2	14	0,5	0,4	0,5

Per ciascun cluster, la bontà delle combinazioni scelte è stata valutata confrontando le distribuzioni dell'mse ottenute allenando le reti dei pazienti di ciascun cluster con i set di parametri scelti per lo specifico cluster, con le combinazioni scelte per gli altri due e con il set (Tabella 5.4) ricavato nello studio precedente [10].

Tabella 5.4: Combinazione scelta in [10]

Neuroni	η_w	η_c	η_σ
14	0,3	0,5	0,4

I box plot riportati in figura 5.5, 5.6 e 5.7 non mostrano performance migliori rispetto a quelle che si otterrebbero selezionando il set ricavato precedentemente. Inoltre, in tutti i cluster, non è evidente una differenza sostanziale tra le distribuzioni ottenute con il set specifico e quelle con le altre combinazioni. Nel cluster 2, in particolare, con alcune combinazioni scelte si ottengono risultati peggiori rispetto a quelli ricavati con i set degli altri cluster.

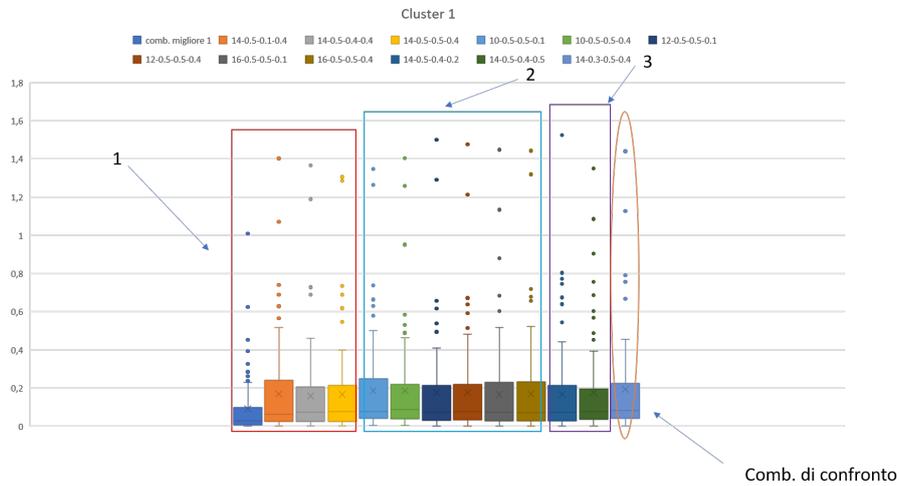


Figura 5.5: Box-plot Cluster 1

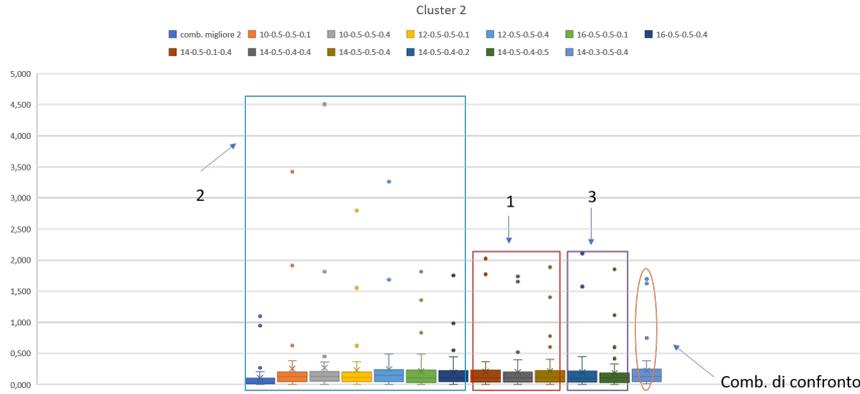


Figura 5.6: Box-plot Cluster 2

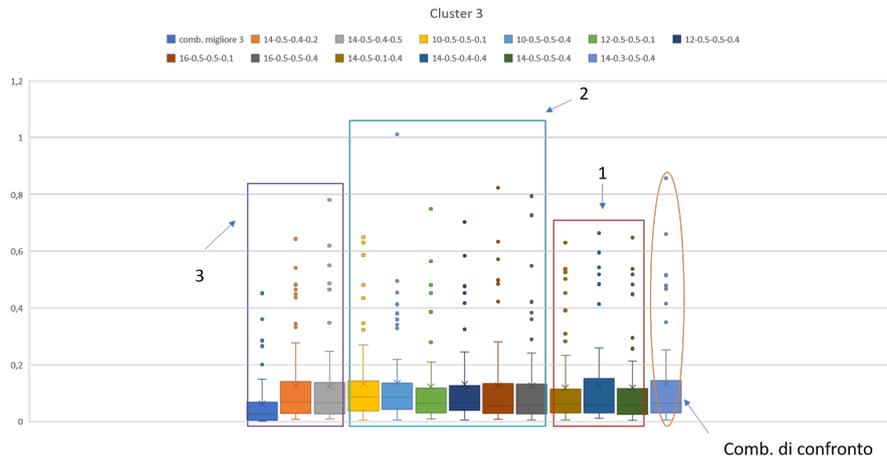


Figura 5.7: Box-plot Cluster 3

5.1.6 Analisi delle combinazioni

Il secondo metodo di selezione delle combinazioni di parametri per la rete RBF è basato sull'analisi della distribuzione dell'errore quadratico medio (mse) in ciascuna delle 875 combinazioni a disposizione per ogni paziente dei 3 cluster.

Per ciascun paziente, delle 10 ripetizioni, è stata selezionata quella che per la specifica combinazione presenta il valore di mse più basso.

Per ogni cluster, si è ottenuta una matrice $N \times 875$ (dove N è il numero di pazienti del gruppo) contenente i valori di mse estrapolati.

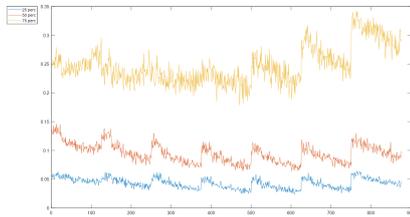
Successivamente, per ogni combinazione, sono stati calcolati:

- 25°, 50° e 75° percentile della distribuzione dell'errore

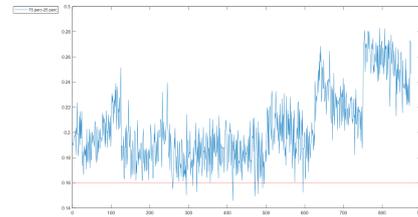
- differenza tra il 75° e il 25° percentile

Sono state selezionate le combinazioni aventi una differenza di percentili minore di una soglia, scelta analizzando l'andamento crescente delle differenze e prendendo il valore di differenza che rappresenta il punto in cui tale curva ha un cambio di pendenza. Nei grafici riportati in figura 5.8, 5.9 e 5.10 sono rappresentati, per ognuna delle 875 combinazioni, il 25°, 50° e 75° percentile della distribuzione dell'errore, la differenza tra il 75° e il 25° percentile e la differenza in ordine crescente, rispettivamente per il cluster 1, 2 e 3.

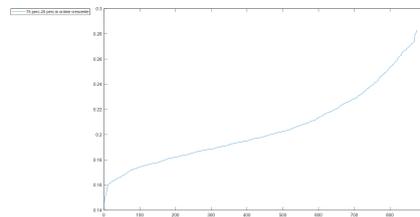
Cluster 1



(a) 25°-50°-75° percentile



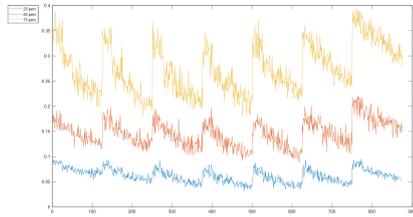
(b) Differenza tra 75° e 25°



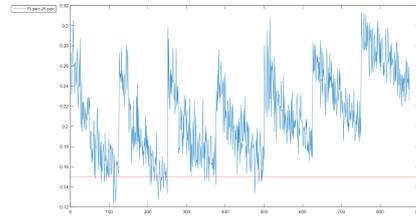
(c) Differenza tra 75° e 25° in ordine crescente

Figura 5.8: Analisi percentili

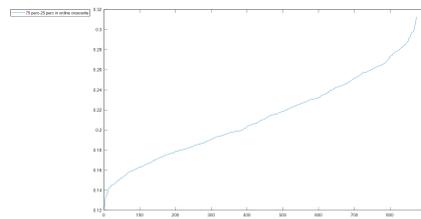
Cluster 2



(a) 25°-50°-75° percentile



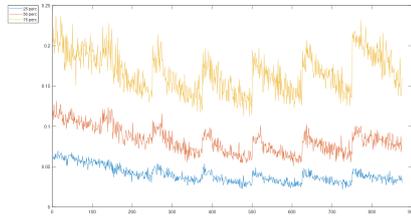
(b) Differenza tra 75° e 25°



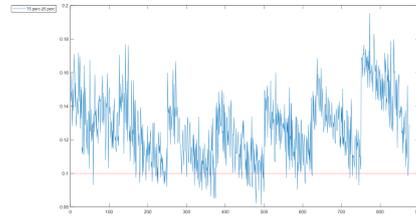
(c) Differenza tra 75° e 25° in ordine crescente

Figura 5.9: Analisi percentili

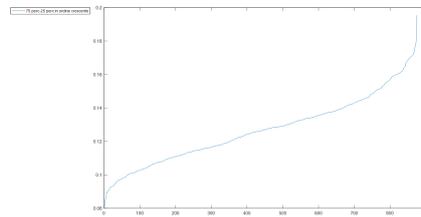
Cluster 3



(a) 25°-50°-75° percentile



(b) Differenza tra 75° e 25°

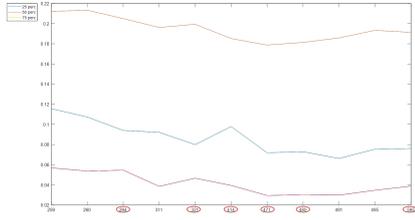


(c) Differenza tra 75° e 25° in ordine crescente

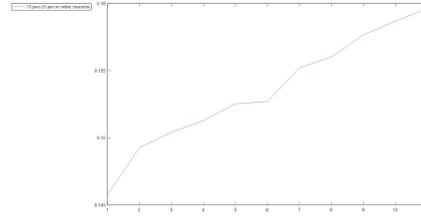
Figura 5.10: Analisi percentili

Delle n combinazioni individuate, riportate nell'appendice D, ne sono state selezionate m , (6, 5 e 10 per il cluster 1, 2 e 3 rispettivamente) allo stesso modo analizzando l'andamento crescente delle differenze delle n combinazioni e prendendo il punto in cui tale curva presenta un cambio di pendenza. In figura 5.11, 5.12 e 5.13 sono riportati i grafici relativi alle differenze in ordine crescente delle m combinazioni e i valori di 25°, 50° e 75° percentile della distribuzione dell'errore per ognuna di esse.

Cluster 1



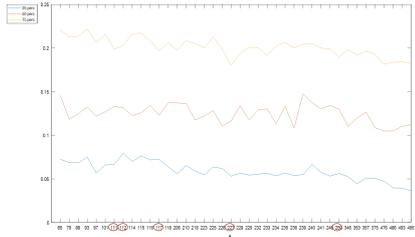
(a) Combinazioni con differenza minore di 0.16



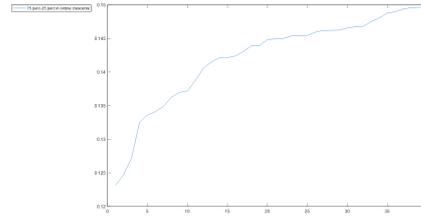
(b) Combinazioni ordinate per differenze crescenti

Figura 5.11: Analisi delle combinazioni

Cluster 2



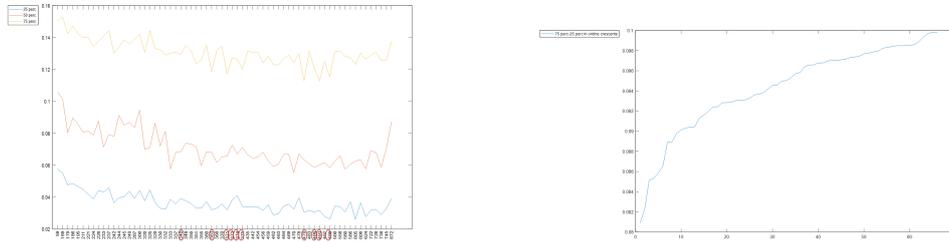
(a) Combinazioni con differenza minore di 0.15



(b) Combinazioni ordinate per differenze crescenti

Figura 5.12: Analisi delle combinazioni

Cluster 3



(a) *Combinazioni con differenza minore di 0.1* (b) *Combinazioni ordinate per differenze crescenti*

Figura 5.13: Analisi delle combinazioni

Per ciascun cluster, le distribuzioni dell'errore, ottenute allenando le reti dei pazienti con le m combinazioni, sono state confrontate con la distribuzione ottenuta imponendo la combinazione scelta nello studio precedente. Da un'analisi dei box-plot, riportati in figura 5.14, è evidente che le combinazioni individuate per i singoli cluster rappresentano una miglior soluzione in termini di accuratezza dei risultati rispetto all'utilizzo del set individuato in [10] (ultimo box-plot nei grafici in figura 5.14).

Tra queste, per ogni cluster, è stata scelta la combinazione con valore di mediana più piccolo. Le combinazioni, rispettivamente per il cluster 1, 2 e 3 sono riportate nella tabella 5.5).

Tabella 5.5: Combinazioni scelte per ciascun cluster

Cluster	Combinazione	Neuroni	η_w	η_c	η_σ
1	471	14	0,4	0,5	0,1
2	227	10	0,5	0,1	0,2
3	499	14	0,5	0,5	0,4

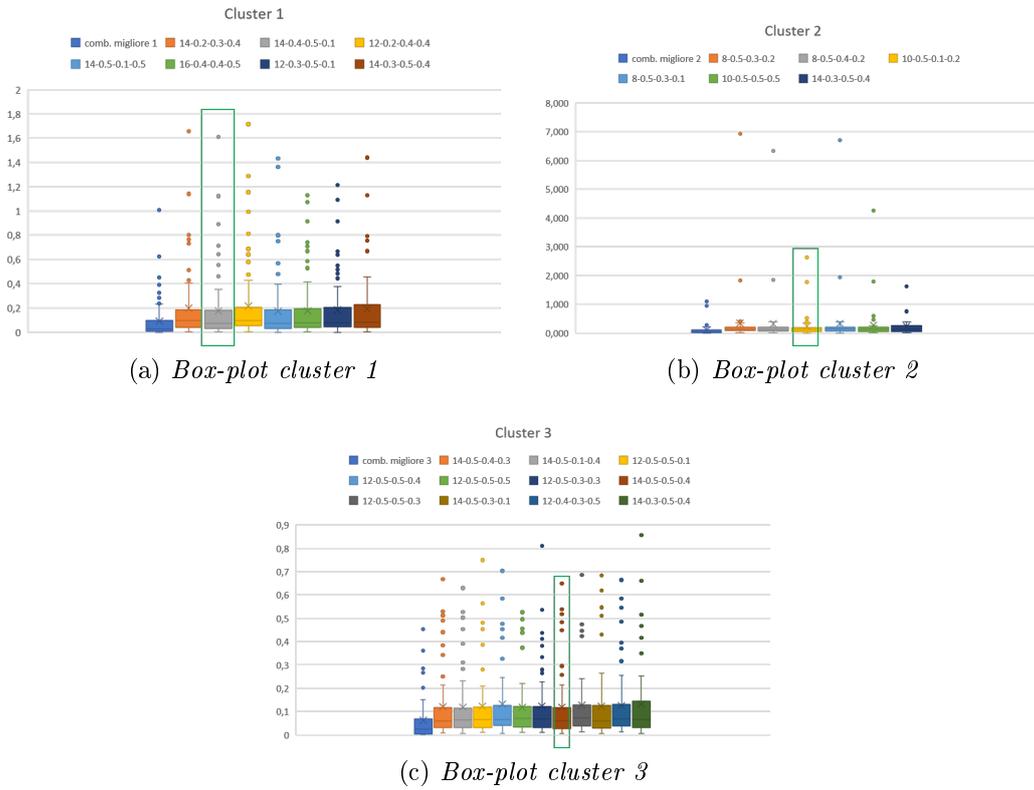


Figura 5.14: Confronto tra le distribuzioni di mse delle combinazioni ottenute e le distribuzioni di mse relative alla combinazione ricavata in [10]

Un'ulteriore valutazione della bontà delle combinazioni scelte deriva dall'analisi delle distribuzioni degli errori ottenute allenando la rete RBF dei pazienti di un cluster con le combinazioni selezionate per gli altri sottogruppi.

I grafici in figura 5.15 mostrano che, con la combinazione scelta per il cluster, si ottengono reti con performance migliori. Le distribuzioni, infatti, hanno valori di errore compresi tra il 25° e il 75° percentile, molto bassi, nel range di 0-0.2. Inoltre, le distribuzioni sono più strette intorno al valore di mediana rispetto a quelle ottenute con le altre combinazioni.

Per validare la scelta delle combinazioni (Tabella 5.5), l'allenamento della rete

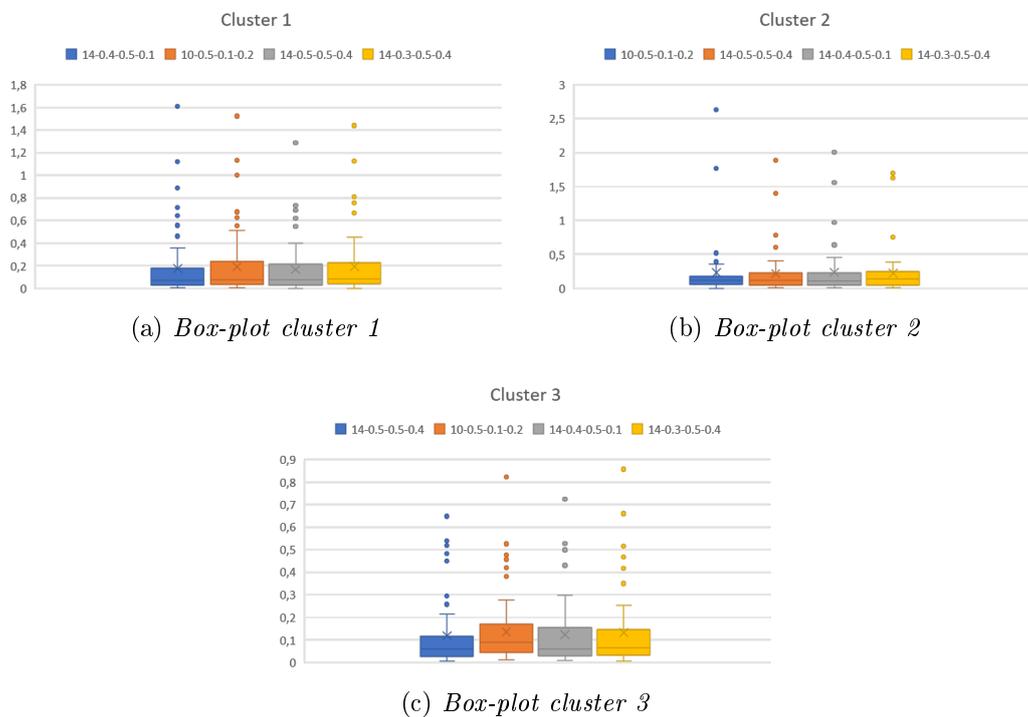


Figura 5.15: Confronto tra le distribuzioni di mse delle combinazioni scelte per il cluster e le distribuzioni di mse relative alle combinazioni degli altri cluster

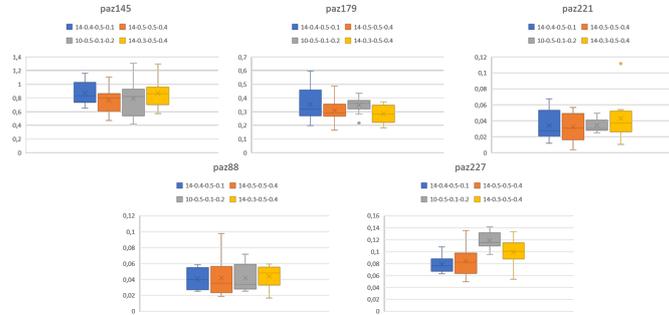
RBF è stato effettuato anche per i pazienti del validation set di ciascun cluster. Per tutti i pazienti, sono state eseguite 10 ripetizioni di addestramento settando il set di parametri scelto per lo specifico cluster, i set degli altri sottogruppi e il set individuato in [10].

Le distribuzioni di mse sono riportate nei grafici in figura 5.16, in cui il primo box-plot rappresenta la distribuzione di mse relativa alla combinazione dello specifico sottogruppo.

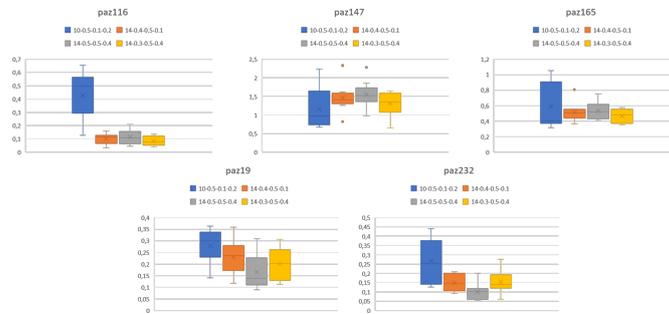
I box-plot permettono di validare la scelta per il primo e il terzo cluster, mentre nel secondo cluster, l'utilizzo della combinazione scelta non permette di ottenere i

risultati migliori in termini di confronto di mse ma i valori di errore sono, comunque, molto bassi.

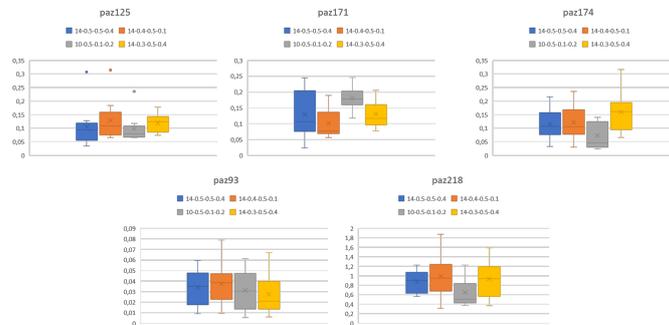
Tra i due metodi di selezione, dunque, quest'ultimo risulta essere il più affidabile.



(a) Box-plot cluster 1



(b) Box-plot cluster 2



(c) Box-plot cluster 3

Figura 5.16: Confronto tra le distribuzioni di mse dei pazienti del validation set

5.2 GA

L'algoritmo genetico utilizzato in questa analisi è stato implementato nel precedente lavoro di tesi [10]. L'algoritmo prevede la codifica Codified Variables per le

soluzioni, in particolare il vettore della soluzione è costituito da 12 bit, i primi 6 codificano il valore di MaxEUD per la vescica, i restanti per il retto. È stato scelto di assegnare un codice binario da 0 a 40 per rappresentare tutti i possibili valori di dose (range 1000-5000 a passi di 100). La decodifica è eseguita convertendo il valore da binario a decimale, aggiungendo ad esso 10 e moltiplicandolo per 100. La funzione di fitness è rappresentata dalla rete RBF, la quale è stata allenata per associare alla soluzione il suo valore di fitness. Per tutti i pazienti dello specifico cluster, la rete utilizzata è quella avente i parametri selezionati nello step precedente, riportati nella tabella 5.5.

I parametri fissi sono:

- Numero di individui=100
- Numero di iterazioni=200
- Numero di genitori=80% del numero di individui
- Selezione dei genitori con Roulette Wheel
- Selezione della nuova popolazione con Random Selection
- Mutazione random
- Crossover a due tagli

Essendo un algoritmo stocastico, si è scelto di valutare la stabilità su 50 ripetizioni. Un numero elevato di soluzioni uguali o simili o aventi la stessa fitness è indice della stabilità dell'algoritmo poiché, con quei parametri e partendo dalla stessa popolazione iniziale, tende a convergere alla stessa soluzione finale.

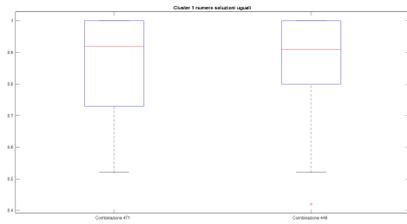
Non è stato eseguito il tuning di pc e pm ma, in questa analisi, sono stati scelti i due parametri $pc=0.85$ e $pm=0.1$ con i quali, in [10], le prestazioni dell'algoritmo si sono dimostrate stabili.

I criteri implementati per valutare la stabilità dell'algoritmo sono il numero di soluzioni uguali e il numero di fitness aventi lo stesso valore. Essi sono normalizzati rispetto al numero massimo di soluzioni o fitness uguali.

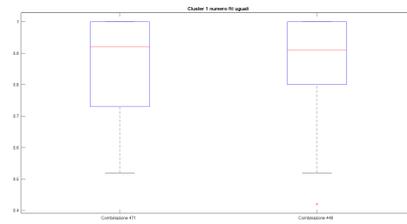
5.2.1 Risultati dell'algoritmo genetico

L'analisi dei box plot riportati in figura 5.17, 5.18, 5.19, 5.20, 5.21 e 5.22 permette di definire l'algoritmo stabile perché, in ogni cluster, i due criteri hanno una distribuzione, la cui mediana è maggiore di 0.85 per i pazienti del training set e maggiore di 0.75 per i pazienti del validation set.

La distribuzione dei valori dei due criteri, inoltre, è confrontata con quella che si ha imponendo i parametri generali della RBF, ricavati in [10].

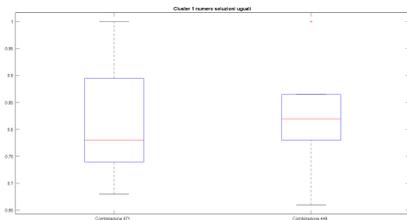


(a) *Box plot Numero soluzioni uguali*

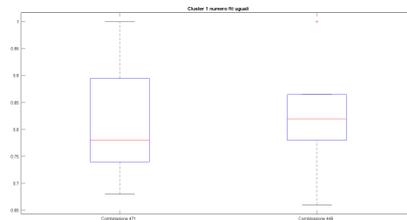


(b) *Box plot Numero fitness uguali*

Figura 5.17: Cluster 1 - Training set - Combinazione 471-449

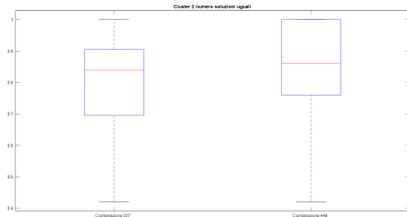


(a) *Box plot Numero soluzioni uguali*

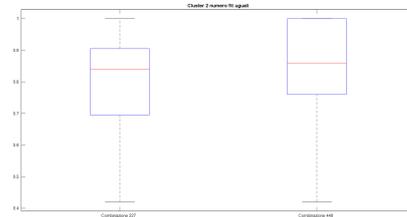


(b) *Box plot Numero fitness uguali*

Figura 5.18: Cluster 1 - Validation set - Combinazione 471-449

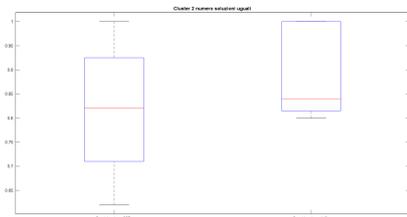


(a) *Box plot Numero soluzioni uguali*

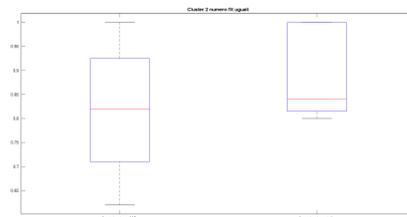


(b) *Box plot Numero fitness uguali*

Figura 5.19: Cluster 2 - Training set - Combinazione 227-449

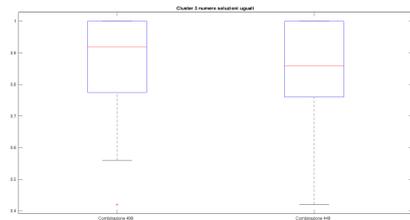


(a) *Box plot Numero soluzioni uguali*

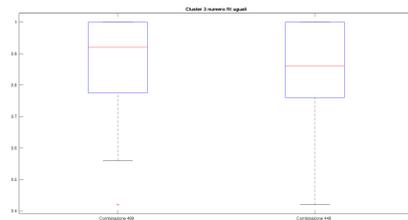


(b) *Box plot Numero fitness uguali*

Figura 5.20: Cluster 2 - Validation set - Combinazione 227-449

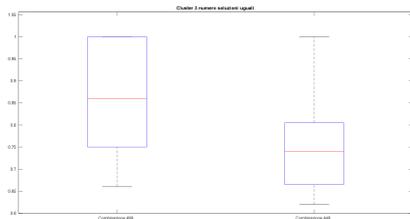


(a) *Box plot Numero soluzioni uguali*

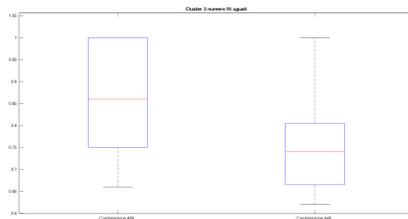


(b) *Box plot Numero fitness uguali*

Figura 5.21: Cluster 3 - Training set - Combinazione 499-449



(a) *Box plot Numero soluzioni uguali*



(b) *Box plot Numero fitness uguali*

Figura 5.22: Cluster 3 - Validation set - Combinazione 499-449

Per ciascun paziente di ogni cluster, le due soluzioni migliori, aventi valore di fitness più basso tra le 50 ripetizioni del GA, (la prima ottenuta con i parametri della RBF specifici per il cluster mentre la seconda con i parametri scelti in [10]), sono state decodificate per ottenere le dosi ottimizzate alla vescica e al retto. Nell'appendice E sono riportati i valori di MaxEUD ottimizzati. Con le due coppie di dosi ricavate sono stati generati, in Raystation, due piani di trattamento, mentre un terzo è stato creato con un algoritmo di pianificazione automatica, sviluppato all'interno dell'azienda "Tecnologie Avanzate".

5.3 Valutazione dei piani radioterapeutici tramite software

I piani sono stati valutati con un algoritmo implementato in Raystation, con il quale si associa un punteggio in funzione di quanto le dosi al target e agli OAR dello specifico piano sono conformi ai range imposti nei Clinical Goals. Nell'appendice F sono riportati gli scores percentuali.

La figura 5.23 mostra le distribuzioni degli scores per i piani dei pazienti che compongono il training set, ottenuti con l'algoritmo di autoplanning (GPS) e con l'algoritmo RBFNN+GA. Da un'analisi dei box plot è possibile dedurre che, in tutti

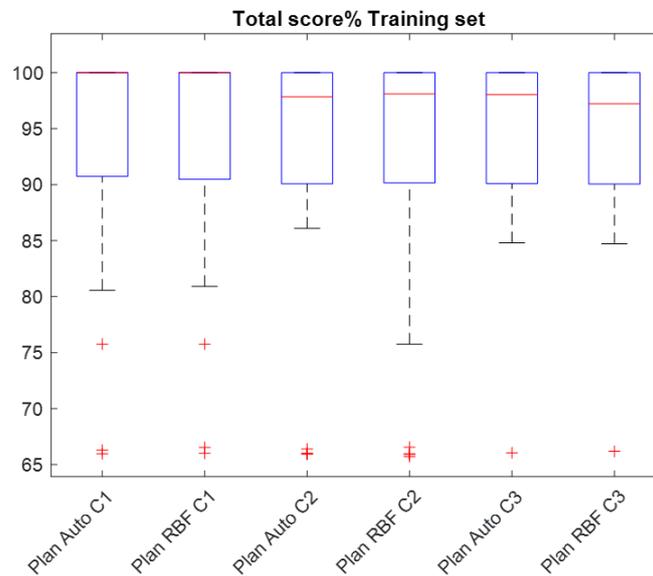


Figura 5.23: Distribuzioni degli scores GPS vs RBFNN+GA

i cluster, i due metodi di ottimizzazione (GPS e RBFNN+GA) sono equivalenti. Ciò è confermato dai risultati ottenuti dal test di Student accoppiato a due code. I

p-value ricavati sono 0.48, 0.87 e 0.48 rispettivamente per il cluster 1, 2 e 3. Sono valori maggiori di 0.05 (livello di significatività del test), dunque le due distribuzioni sono statisticamente equivalenti.

Anche i piani di trattamento ricavati per i pazienti del validation set, sono equivalenti, come dimostrato nella figura 5.24.

Il confronto tra i piani ottenuti con le due varianti dell' algoritmo di ottimizzazione

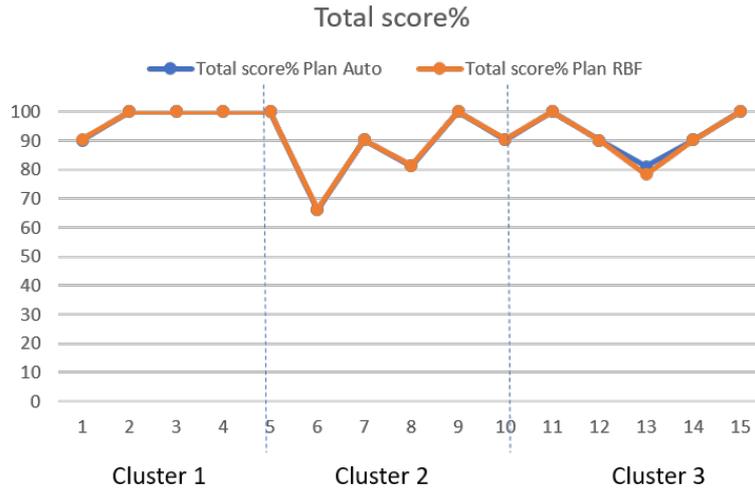


Figura 5.24: Distribuzioni degli scores GPS vs RBFNN+GA (Validation Set)

(RBFNN+GA in [10] e RBFNN+GA, sviluppato nel presente lavoro di tesi, permette di constatare che in entrambi i casi si ottengono buone valutazioni (Figura 5.25).

Gli scores ottenuti dai 3 piani sono, in generale, molto simili, fatta eccezione per

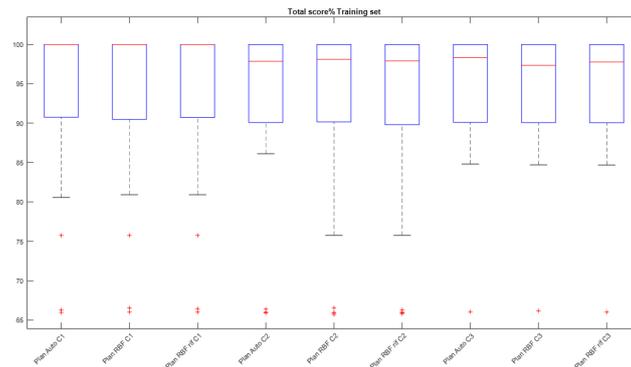


Figura 5.25: Distribuzioni degli scores GPS vs RBFNN+GA vs RBFNN+GA[10]

alcuni pazienti in cui i punteggi si differenziano di oltre 8 punti percentuali.

In alcuni casi, infatti, un metodo di ottimizzazione risulta più performante rispetto agli altri due.

5.3.1 Confronto GPS e RBFNN+GA

Nella tabella 5.6 sono riportati i pazienti per i quali il GPS permette di ottenere piani con punteggi più elevati rispetto all’algoritmo RBFNN+GA.

Tabella 5.6: Confronto scores dei piani di trattamento

Total Score %					
ID	Cluster	T/V	GPS	RBFNN+GA	Differenza score
6	1	T	90,58	81,82	-8,76
126	1	T	98,09	84,71	-13,38
58	2	T	98,37	89,62	-8,75
246	2	T	86,11	77,73	-8,39
247	2	T	90,51	75,76	-14,75
113	3	T	97,25	89,72	-7,54
180	1	T	98,33	90,06	-8,27
23	3	T	98,34	90,10	-8,24

Nella tabella 5.7, invece, sono riportati i pazienti per cui gli scores maggiori sono stati ottenuti dai piani derivanti dall’ottimizzazione delle dosi agli OAR dell’algoritmo implementato in questo lavoro di tesi.

Tabella 5.7: Confronto scores dei piani di trattamento

Total Score %					
ID	Cluster	T/V	GPS	RBFNN+GA	Differenza score
128	1	T	89,85	98,94	9,10
57	2	T	90,10	98,70	8,60
211	2	T	87,20	98,18	10,98
194	2	T	89,54	98,74	9,20
238	2	T	90,00	98,64	8,64
159	3	T	89,32	97,34	8,02

5.3.2 Confronto GPS e RBFNN+GA (versione precedente[10])

Nella tabella 5.8 sono riportati i pazienti per cui l’algoritmo GPS è più performante.

Tabella 5.8: Confronto scores dei piani di trattamento

Total Score %					
ID	Cluster	T/V	GPS	RBFNN+GA[10]	Differenza score
6	1	T	90,58	81,82	8,76
126	1	T	98,09	84,75	13,35
180	1	T	98,33	90,03	8,30
58	2	T	98,37	89,71	8,65
214	2	T	89,83	81,06	8,77
247	2	T	90,51	75,76	14,75
23	3	T	98,34	90,00	8,34
113	3	T	97,25	89,72	7,54
152	2	T	100,00	83,39	16,61

Nella tabella 5.9, invece, sono riportati i pazienti per cui l'algoritmo implementato in [10] dà risultati migliori.

Tabella 5.9: Confronto scores dei piani di trattamento

Total Score %					
ID	Cluster	T/V	GPS	RBFNN+GA[10]	Differenza score
128	1	T	89,85	98,94	-9,10
136	1	T	90,09	99,01	-8,92
57	2	T	90,10	98,70	-8,60
238	2	T	90,00	98,62	-8,62
159	3	T	89,32	96,99	-7,67
211	2	T	87,20	98,18	-10,98

5.3.3 Confronto RBFNN+GA e RBFNN+GA (versione precedente[10])

Nella tabella 5.10, invece, sono riportati i pazienti per cui l'algoritmo implementato in questo lavoro di tesi permette di conferire ai piani un punteggio maggiore.

Tabella 5.10: Confronto scores dei piani di trattamento

Total Score %					
ID	Cluster	T/V	RBFNN+GA	RBFNN+GA[10]	Differenza score
194	2	T	98,74	89,91	8,83
214	2	T	90,11	81,06	9,05
152	2	T	97,73	83,39	14,34

Nella tabella 5.11, invece, sono riportati i pazienti per cui l'algoritmo implementato in [10] dà risultati migliori.

Tabella 5.11: Confronto scores dei piani di trattamento

Total Score %					
ID	Cluster	T/V	RBFNN+GA	RBFNN+GA[10]	Differenza score
246	2	T	77,73	85,85	-8,12
136	1	T	90,02	99,01	-8,99

Dalle tabelle riportate, è interessante osservare come i pazienti che mostrano differenze elevate di scores appartengano, per la maggioranza, al cluster 2. Questo a conferma del fatto che, dei tre cluster individuati, esso ha mostrato criticità in quasi tutti gli step di analisi, dunque i pazienti appartenenti a questo sottogruppo, dovranno essere sottoposti ad ulteriori indagini.

Capitolo 6

Conclusioni

Con il seguente progetto di tesi ci si è posti l'obiettivo di capire se le caratteristiche geometriche della ghiandola prostatica possano contribuire alla definizione dei piani radioterapeutici.

Attraverso un metodo di clustering basato sulla caratterizzazione del volume prostatico è stato possibile suddividere il dataset a disposizione in tre sottogruppi. Inoltre, è stato individuato un set di parametri della rete RBF, implementata nella funzione di fitness dell'algoritmo genetico, valido per tutti i pazienti dello specifico cluster. Allenare le reti impostando, per ciascun cluster, il set individuato ha prodotto un miglioramento delle performance delle reti stesse e di conseguenza, ha permesso di associare con maggior precisione, il valore di fitness alle soluzioni del GA.

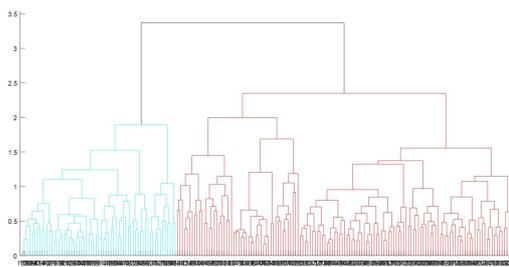
I piani radioterapeutici, generati a partire dai valori di MaxEUD alla vescica e al retto (soluzioni del GA) sono stati valutati con un algoritmo implementato in Raystation ed hanno ottenuto punteggi percentuali molto elevati in tutti i cluster (mediana della distribuzione degli scores sempre maggiore del 90%).

Inoltre, i piani sono risultati equivalenti a quelli ottenuti con un algoritmo attualmente in sviluppo in "Tecnologie Avanzate".

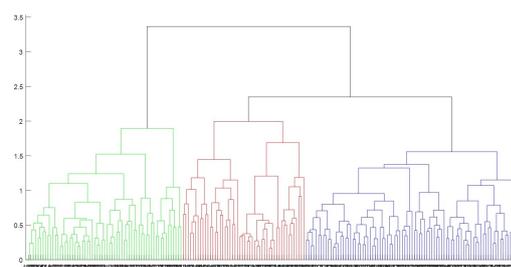
Appendice A

Grafici relativi al clustering

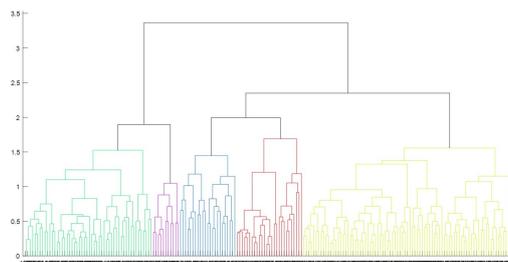
In questa sezione sono riportati i dendrogrammi relativi al clustering dei pazienti, il confronto delle variabilità intra-cluster e la rappresentazione, tramite box-plot, delle distribuzioni dei parametri all'interno di ciascun cluster, per ogni clusterizzazione.



(a) *Dendrogramma con taglio a 2 cluster*

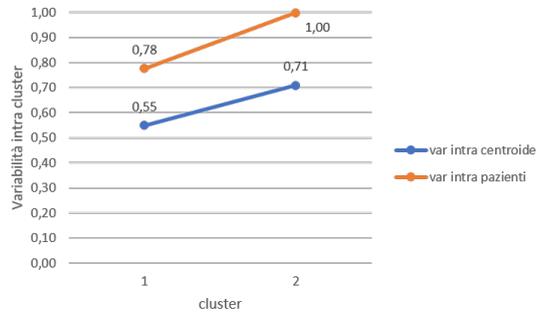


(b) *Dendrogramma con taglio a 3 cluster*

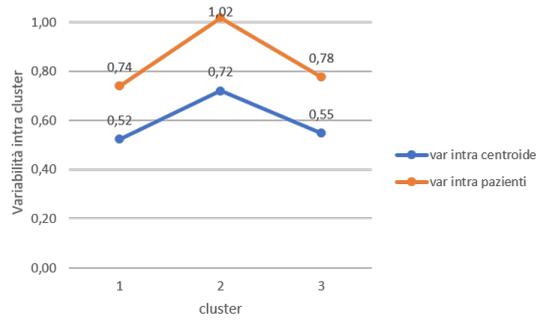


(c) *Dendrogramma con taglio a 5 cluster*

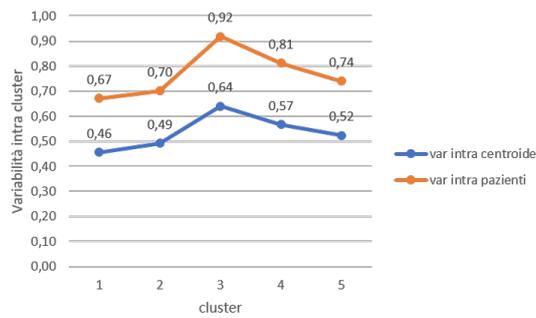
Figura A.1: Tre clusterizzazioni del dataset



(a) Variabilità intra-cluster con taglio a 2 cluster



(b) Variabilità intra-cluster con taglio a 3 cluster



(c) Variabilità intra-cluster con taglio a 5 cluster

Figura A.2: Confronto variabilità intra-cluster

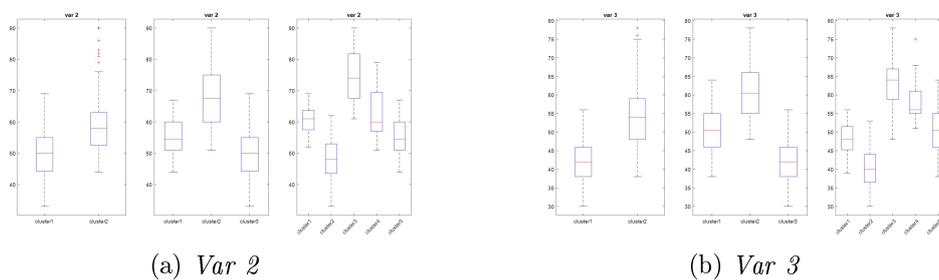


Figura A.3: Box-plot

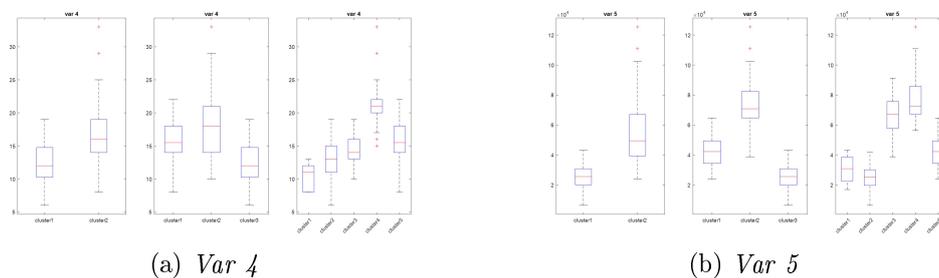


Figura A.4: Box-plot

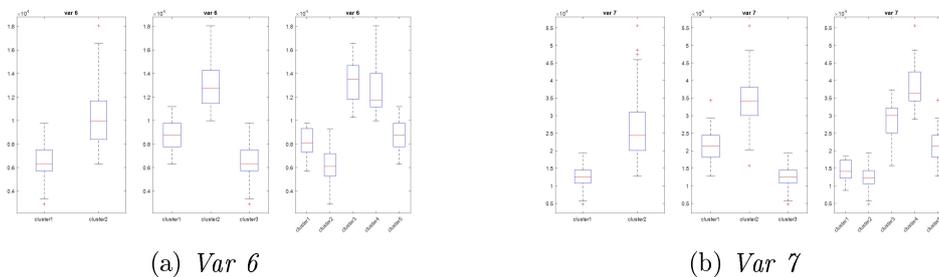


Figura A.5: Box-plot

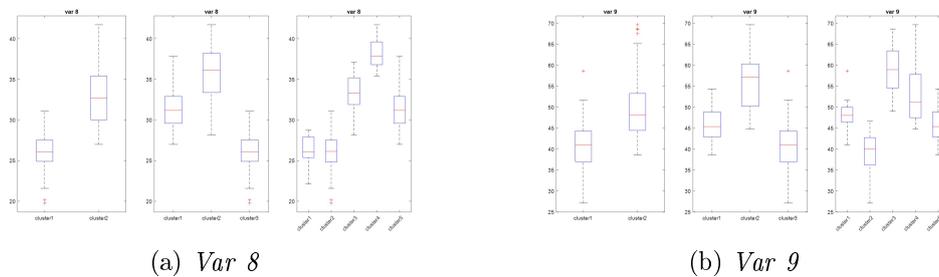


Figura A.6: Box-plot

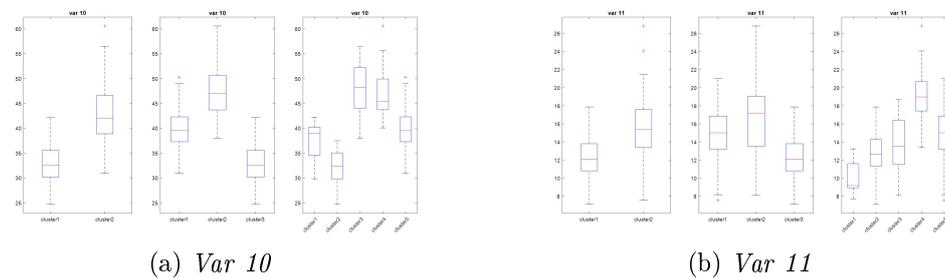


Figura A.7: Box-plot

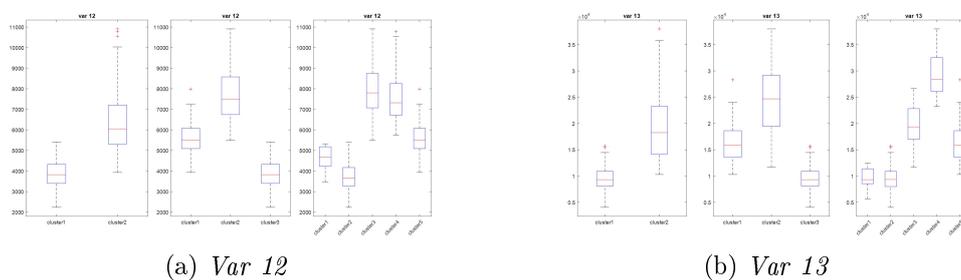


Figura A.8: Box-plot

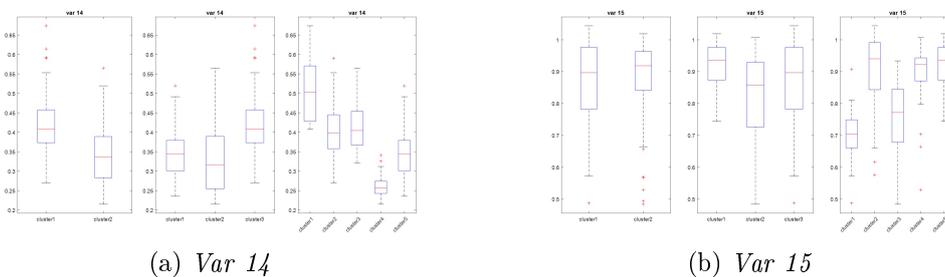


Figura A.9: Box-plot

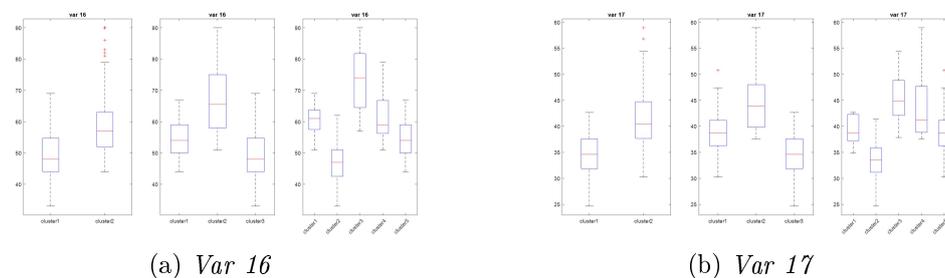


Figura A.10: Box-plot

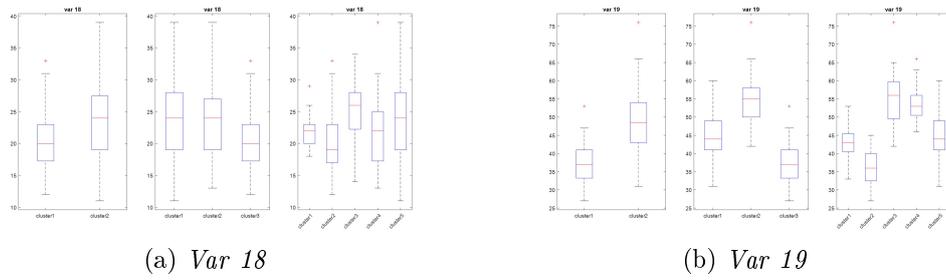


Figura A.11: Box-plot

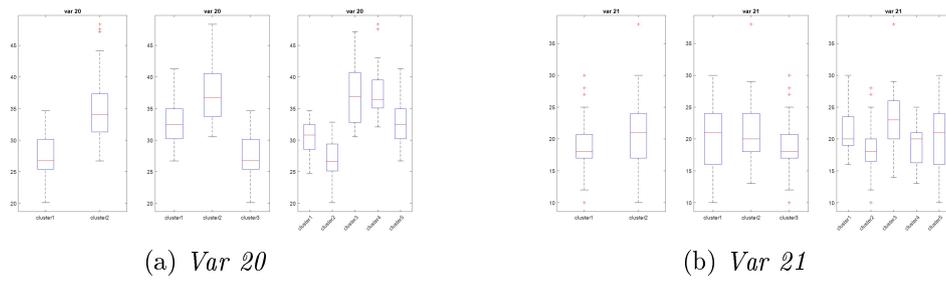


Figura A.12: Box-plot

Appendice B

Risultati associazione Paziente-Cluster

B.1 Distanza Paziente-Centroide

Tabella B.1: *Confusion Matrix mediana-euclidea*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	72	4	7
	Cluster 2	8	46	0
	Cluster 3	6	0	56

Tabella B.2: *Confusion Matrix mediana-L1*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	70	6	7
	Cluster 2	8	44	0
	Cluster 3	8	0	56

Tabella B.3: *Confusion Matrix media-euclidea*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	72	5	7
	Cluster 2	6	45	0
	Cluster 3	8	0	56

Tabella B.4: *Confusion Matrix media-L1*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	70	6	6
	Cluster 2	5	44	0
	Cluster 3	11	0	57

B.2 Associazione k-nn

Tabella B.5: *Confusion Matrix k=5*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	82	2	5
	Cluster 2	1	45	0
	Cluster 3	3	1	58
	NC	0	2	0

Tabella B.6: *Confusion Matrix k=6*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	79	2	3
	Cluster 2	1	45	0
	Cluster 3	0	1	55
	NC	6	2	5

Tabella B.7: *Confusion Matrix k=7*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	84	3	6
	Cluster 2	2	44	0
	Cluster 3	0	1	57
	NC	0	2	0

Tabella B.8: *Confusion Matrix k=8*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	82	3	4
	Cluster 2	2	42	0
	Cluster 3	0	2	55
	NC	2	3	4

Tabella B.9: *Confusion Matrix k=9*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	84	4	6
	Cluster 2	2	43	0
	Cluster 3	0	2	57
	NC	0	1	0

Tabella B.10: *Confusion Matrix k=10*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	79	4	5
	Cluster 2	1	42	0
	Cluster 3	0	2	56
	NC	6	2	2

Tabella B.11: *Confusion Matrix k=11*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	83	5	5
	Cluster 2	2	43	0
	Cluster 3	1	2	58
	NC	0	0	0

Tabella B.12: *Confusion Matrix k=12*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	82	4	3
	Cluster 2	1	42	0
	Cluster 3	1	2	53
	NC	2	2	7

Tabella B.13: *Confusion Matrix k=13*

		Cluster Dendrogramma		
		Cluster 1	Cluster 2	Cluster 3
Cluster Associato	Cluster 1	83	4	8
	Cluster 2	2	44	0
	Cluster 3	1	2	55
	NC	0	0	0

Appendice C

Risultati tuning parametri della rete RBFNN-Analisi dei singoli parametri

Cluster 1

Tabella C.1: Set parametri migliori

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz2	16	0,5	0,5	0,2	0,003
paz5	18	0,5	0,2	0,4	0,033
paz6	16	0,4	0,3	0,4	0,624
paz7	18	0,4	0,4	0,4	0,001
paz9	12	0,4	0,5	0,4	0,031
paz12	12	0,4	0,3	0,2	0,003
paz13	16	0,4	0,1	0,2	0,091
paz17	16	0,5	0,2	0,3	0,392
paz26	12	0,4	0,5	0,4	0,088
paz27	16	0,5	0,1	0,4	0,010
paz31	12	0,5	0,4	0,5	0,005
paz47	12	0,2	0,4	0,5	0,025
paz235	14	0,2	0,2	0,5	0,122
paz236	14	0,5	0,2	0,2	0,012
paz244	16	0,3	0,5	0,5	0,0001
paz251	18	0,2	0,1	0,5	0,004
paz253	14	0,5	0,4	0,4	0,172

Tabella C.1: continua

Tabella C.1: continua

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz32	10	0,5	0,4	0,4	0,008
paz49	14	0,5	0,5	0,4	0,041
paz237	18	0,5	0,5	0,4	0,199
paz29	8	0,3	0,2	0,2	0,024
paz36	16	0,2	0,5	0,4	0,0002
paz39	10	0,5	0,4	0,1	0,238
paz89	14	0,3	0,4	0,2	0,023
paz220	18	0,5	0,1	0,2	0,005
paz222	10	0,5	0,3	0,5	0,054
paz224	14	0,5	0,3	0,2	0,030
paz225	16	0,1	0,4	0,2	0,007
paz233	10	0,5	0,2	0,4	0,012
paz234	8	0,4	0,5	0,4	1,008
paz53	14	0,5	0,5	0,3	0,046
paz91	14	0,5	0,1	0,3	0,175
paz45	10	0,5	0,3	0,3	0,188
paz63	14	0,5	0,5	0,4	0,326
paz68	14	0,5	0,5	0,1	0,089
paz95	10	0,4	0,4	0,3	0,091
paz202	20	0,4	0,3	0,2	0,0001
paz203	20	0,4	0,4	0,1	0,0004
paz205	10	0,4	0,3	0,1	0,452
paz207	14	0,5	0,1	0,4	0,006
paz208	20	0,2	0,4	0,1	0,00001
paz210	14	0,3	0,4	0,3	0,006
paz212	16	0,5	0,3	0,3	0,230
paz213	14	0,2	0,5	0,1	0,085
paz215	18	0,5	0,4	0,3	0,011
paz43	16	0,3	0,1	0,5	0,025
paz219	12	0,5	0,1	0,4	0,043
paz41	12	0,5	0,3	0,4	0,035
paz52	14	0,4	0,5	0,5	0,008
paz87	20	0,3	0,1	0,5	0,017
paz109	14	0,3	0,1	0,1	0,007
paz120	14	0,5	0,3	0,4	0,262
paz121	12	0,2	0,1	0,3	0,049
paz122	20	0,5	0,4	0,2	0,025
paz124	10	0,5	0,5	0,5	0,120

Tabella C.1: continua

Tabella C.1: continua

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz126	12	0,4	0,4	0,1	0,202
paz127	16	0,2	0,2	0,1	0,006
paz128	10	0,4	0,4	0,4	0,177
paz136	10	0,3	0,1	0,3	0,054
paz138	18	0,5	0,3	0,2	0,056
paz151	8	0,5	0,4	0,2	0,008
paz180	14	0,5	0,1	0,4	0,001
paz181	8	0,4	0,5	0,1	0,158
paz183	18	0,5	0,2	0,4	0,019
paz184	12	0,1	0,3	0,3	0,009
paz188	8	0,5	0,5	0,3	0,047
paz189	14	0,4	0,5	0,1	0,010
paz190	14	0,5	0,1	0,3	0,001
paz195	14	0,5	0,3	0,3	0,016
paz140	14	0,1	0,2	0,4	0,020
paz146	16	0,1	0,2	0,5	0,004
paz150	12	0,5	0,1	0,1	0,011
paz154	12	0,5	0,1	0,3	0,040
paz156	14	0,5	0,5	0,5	0,043
paz160	18	0,5	0,3	0,1	0,086
paz161	10	0,1	0,2	0,1	0,003
paz167	14	0,5	0,2	0,2	0,010
paz172	8	0,5	0,1	0,1	0,337
paz176	8	0,5	0,1	0,1	0,284
paz166	12	0,5	0,1	0,3	0,100

Cluster 2

Tabella C.2: Set parametri migliori

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz10	8	0,3	0,5	0,2	0,048
paz112	12	0,3	0,5	0,2	0,158
paz115	8	0,5	0,4	0,4	0,078
paz118	16	0,4	0,1	0,1	0,152
paz132	16	0,5	0,5	0,1	0,018
paz137	16	0,5	0,1	0,2	0,945
paz14	12	0,3	0,3	0,4	0,027

Tabella C.2: continua

Tabella C.2: continua

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz15	18	0,2	0,3	0,1	0,002
paz152	10	0,5	0,1	0,5	0,109
paz153	12	0,3	0,5	0,5	0,009
paz169	12	0,4	0,2	0,5	0,009
paz170	18	0,2	0,4	0,3	0,022
paz173	10	0,3	0,1	0,4	0,080
paz175	12	0,5	0,1	0,5	0,043
paz182	14	0,5	0,3	0,1	0,042
paz185	10	0,5	0,2	0,4	0,210
paz193	16	0,5	0,5	0,1	0,025
paz194	16	0,3	0,5	0,5	0,016
paz196	8	0,5	0,2	0,2	0,096
paz201	10	0,3	0,3	0,1	0,037
paz209	8	0,5	0,3	0,4	0,170
paz21	8	0,1	0,2	0,1	0,011
paz211	12	0,4	0,2	0,2	0,062
paz214	16	0,4	0,1	0,4	0,021
paz217	18	0,4	0,1	0,1	0,004
paz226	18	0,5	0,3	0,4	0,051
paz231	18	0,1	0,3	0,4	0,090
paz238	10	0,5	0,3	0,4	0,004
paz243	14	0,4	0,5	0,2	0,005
paz246	12	0,4	0,5	0,4	1,101
paz247	14	0,5	0,2	0,3	0,083
paz248	10	0,5	0,5	0,1	0,182
paz249	8	0,4	0,5	0,2	0,045
paz25	10	0,5	0,1	0,1	0,127
paz252	10	0,5	0,3	0,2	0,155
paz30	12	0,5	0,2	0,1	0,095
paz33	10	0,5	0,2	0,3	0,010
paz57	8	0,3	0,4	0,3	0,089
paz58	18	0,5	0,2	0,4	0,0005
paz60	14	0,4	0,5	0,2	0,001
paz61	16	0,5	0,2	0,4	0,029
paz64	14	0,5	0,3	0,3	0,027
paz65	12	0,4	0,5	0,4	0,020
paz75	14	0,4	0,3	0,2	0,270
paz92	16	0,4	0,5	0,1	0,006

Cluster 3

Tabella C.3: Set parametri migliori

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz4	14	0,3	0,4	0,3	0,057
paz11	18	0,5	0,3	0,2	0,091
paz16	12	0,5	0,5	0,1	0,068
paz18	20	0,2	0,3	0,3	0,003
paz20	12	0,5	0,5	0,5	0,047
paz22	20	0,5	0,4	0,2	0,003
paz34	16	0,3	0,4	0,5	0,291
paz48	10	0,4	0,4	0,1	0,029
paz110	10	0,4	0,4	0,3	0,011
paz3	14	0,5	0,1	0,4	0,063
paz24	12	0,1	0,2	0,3	0,0004
paz62	14	0,5	0,4	0,5	0,011
paz66	18	0,5	0,3	0,5	0,027
paz111	20	0,5	0,1	0,1	0,004
paz113	12	0,1	0,4	0,1	0,005
paz114	8	0,5	0,4	0,5	0,361
paz117	12	0,5	0,5	0,3	0,027
paz129	14	0,4	0,3	0,2	0,004
paz130	20	0,5	0,2	0,5	0,026
paz131	16	0,1	0,2	0,5	0,004
paz139	14	0,4	0,5	0,2	0,025
paz142	16	0,5	0,3	0,4	0,008
paz162	16	0,4	0,2	0,2	0,002
paz163	8	0,5	0,1	0,5	0,008
paz168	18	0,1	0,3	0,3	0,007
paz73	18	0,5	0,1	0,4	0,050
paz8	14	0,3	0,1	0,2	0,023
paz94	14	0,4	0,5	0,4	0,012
paz1	14	0,4	0,3	0,5	0,044
paz119	16	0,4	0,5	0,1	0,002
paz123	8	0,5	0,1	0,5	0,072
paz143	10	0,5	0,2	0,1	0,069
paz155	16	0,5	0,4	0,2	0,002
paz158	18	0,3	0,4	0,2	0,030
paz159	16	0,5	0,1	0,5	0,142
paz96	10	0,5	0,2	0,1	0,117

Tabella C.3: continua

Tabella C.3: continua

ID	Neuroni	η_w	η_c	η_{sigma}	mse
paz23	10	0,5	0,5	0,4	0,138
paz164	18	0,5	0,3	0,5	0,005
paz177	14	0,2	0,4	0,3	0,017
paz178	12	0,1	0,4	0,3	0,0004
paz187	14	0,3	0,4	0,4	0,011
paz191	12	0,5	0,5	0,4	0,116
paz192	14	0,4	0,2	0,4	0,004
paz197	14	0,3	0,2	0,5	0,001
paz198	8	0,4	0,3	0,4	0,202
paz199	20	0,3	0,1	0,2	0,002
paz200	8	0,4	0,4	0,2	0,266
paz229	16	0,5	0,3	0,4	0,150
paz90	20	0,2	0,4	0,5	0,014
paz223	18	0,4	0,1	0,2	0,002
paz230	16	0,5	0,3	0,2	0,051
paz240	10	0,5	0,5	0,3	0,031
paz242	18	0,5	0,2	0,4	0,019
paz250	14	0,4	0,2	0,2	0,452
paz254	12	0,4	0,1	0,3	0,003
paz245	10	0,5	0,5	0,3	0,059
paz40	14	0,4	0,1	0,1	0,286

C.1 Confronto target e valore di fitness restituito dalla rete

Cluster 1

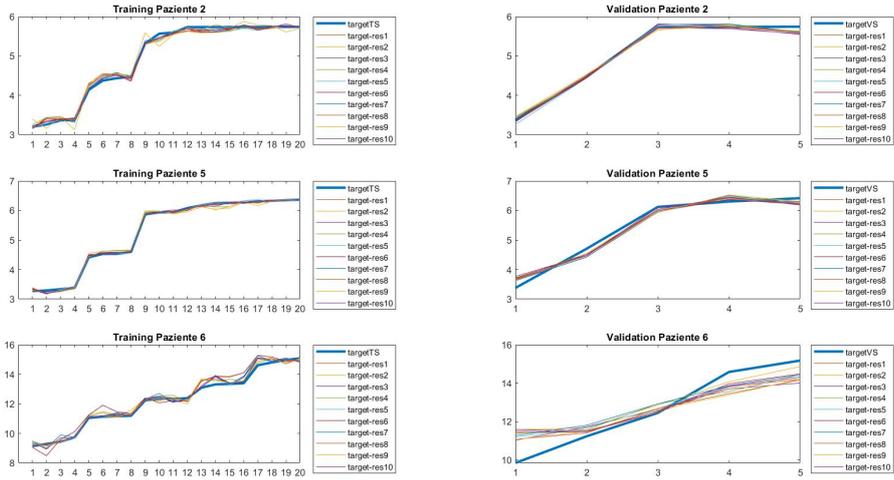


Figura C.1: Pazienti 2-5-6

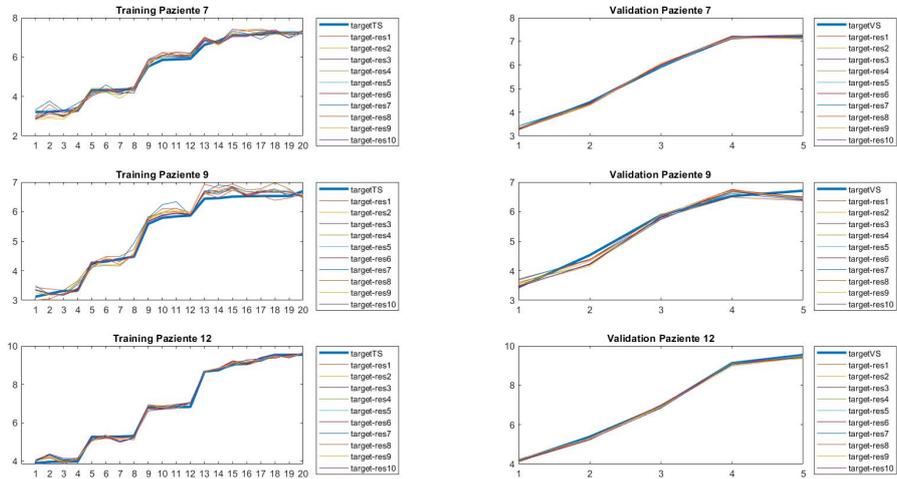


Figura C.2: Pazienti 7-9-12

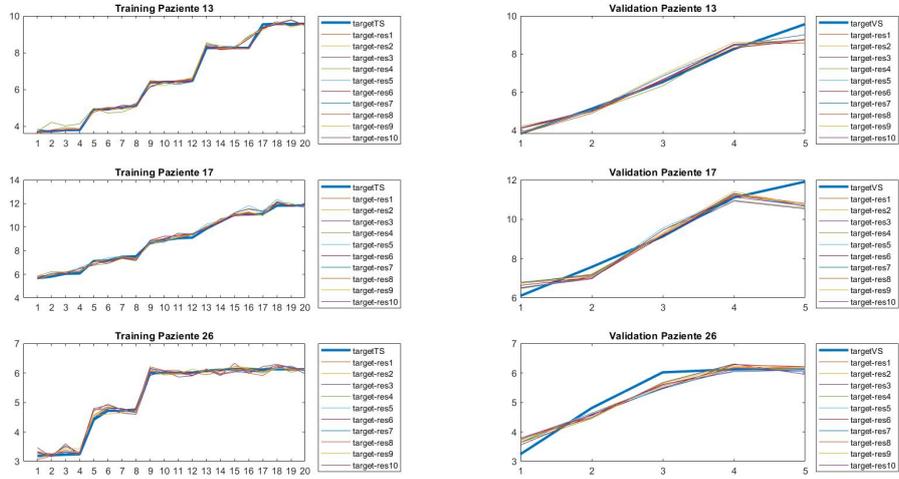


Figura C.3: Pazienti 13-17-26

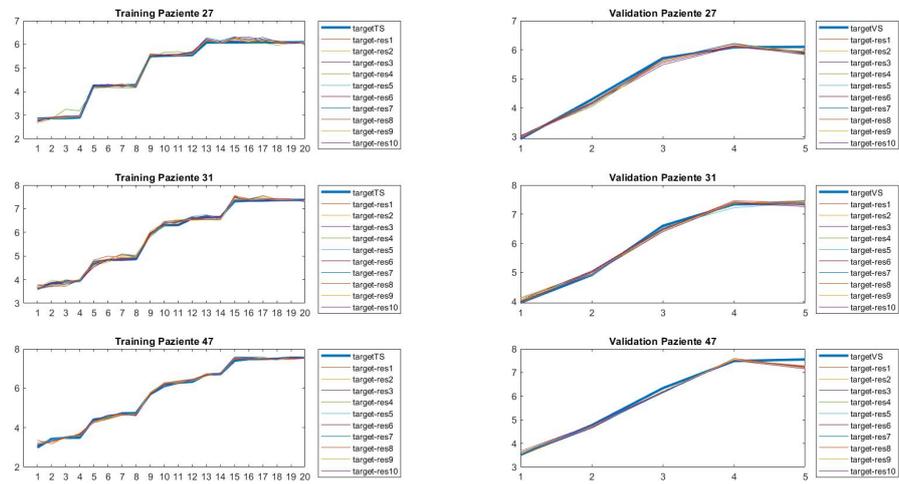


Figura C.4: Pazienti 27-31-47

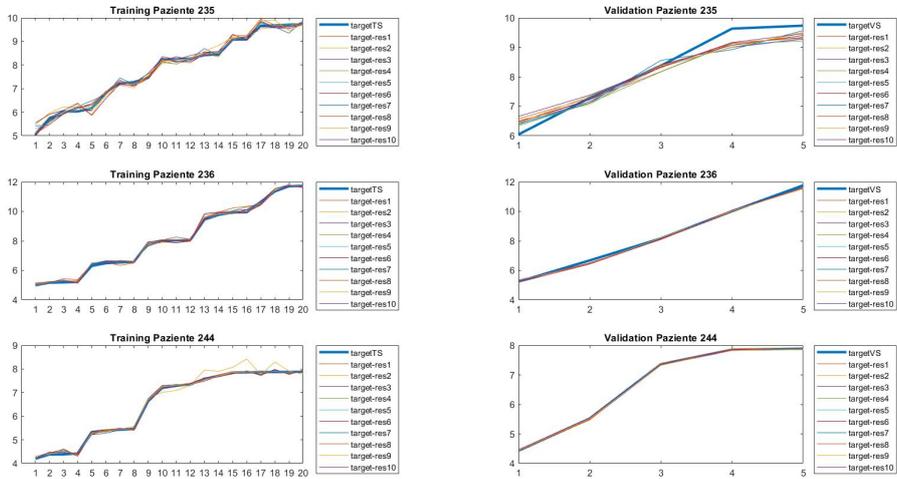


Figura C.5: Pazienti 235-236-244

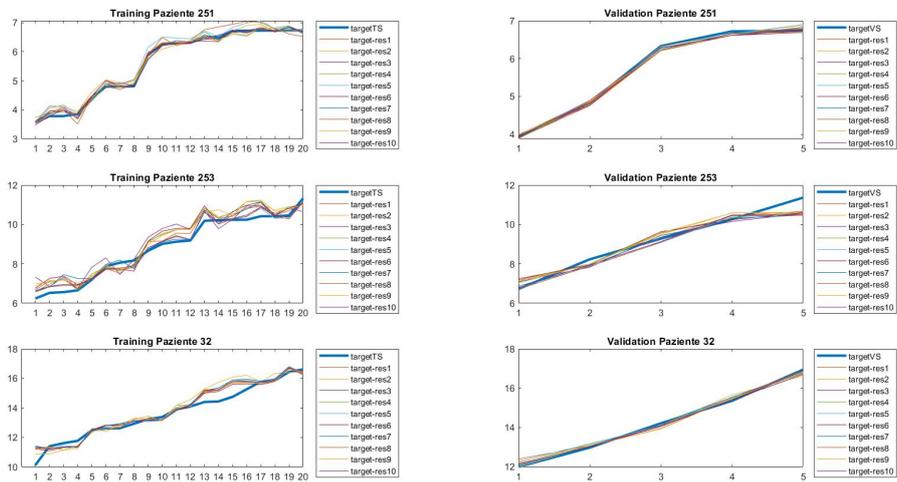


Figura C.6: Pazienti 251-253-32

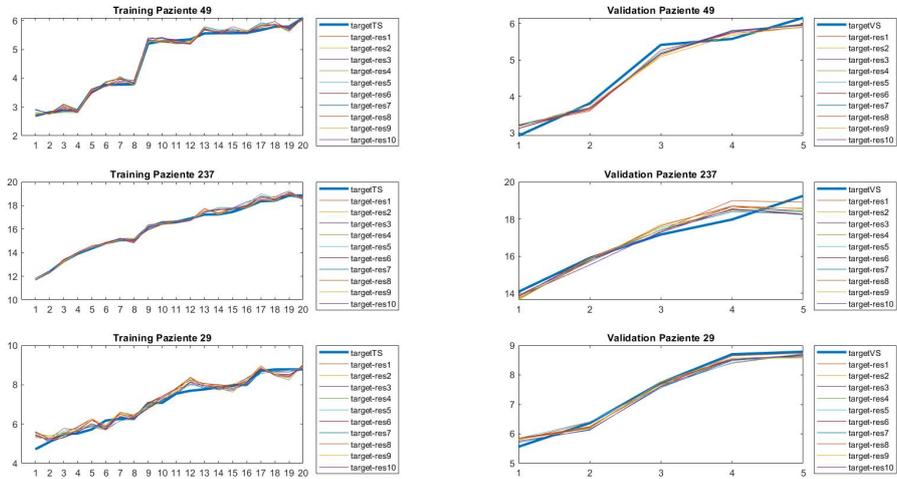


Figura C.7: Pazienti 49-237-29

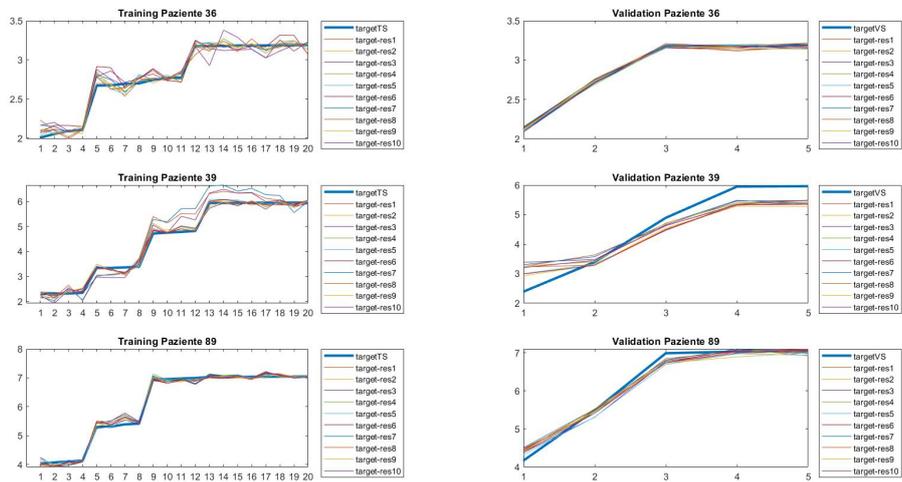


Figura C.8: Pazienti 36-89-39

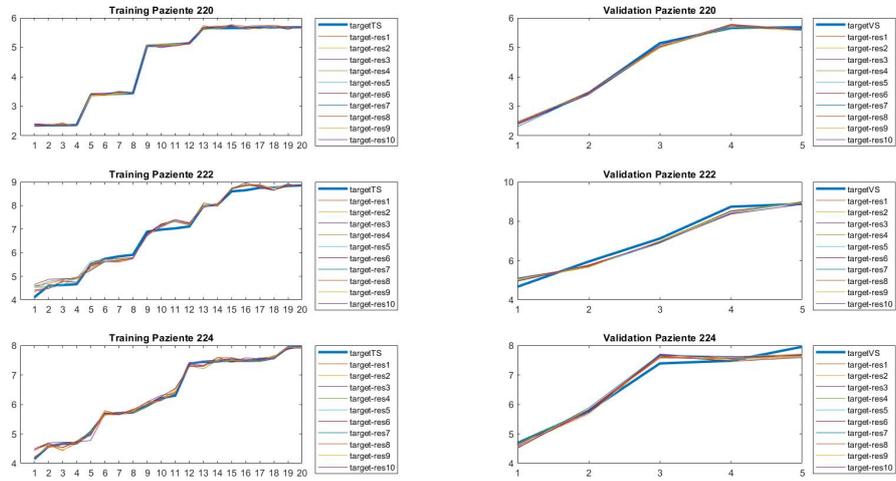


Figura C.9: Pazienti 220-222-224

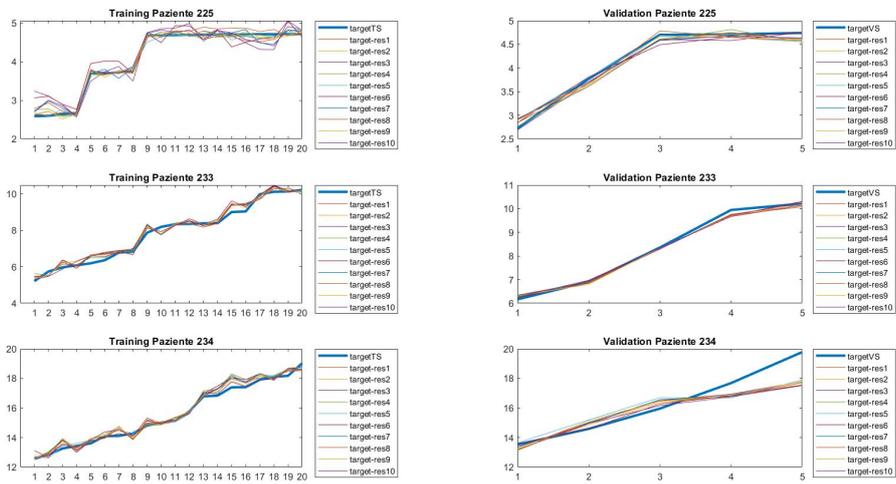


Figura C.10: Pazienti 225-233-234

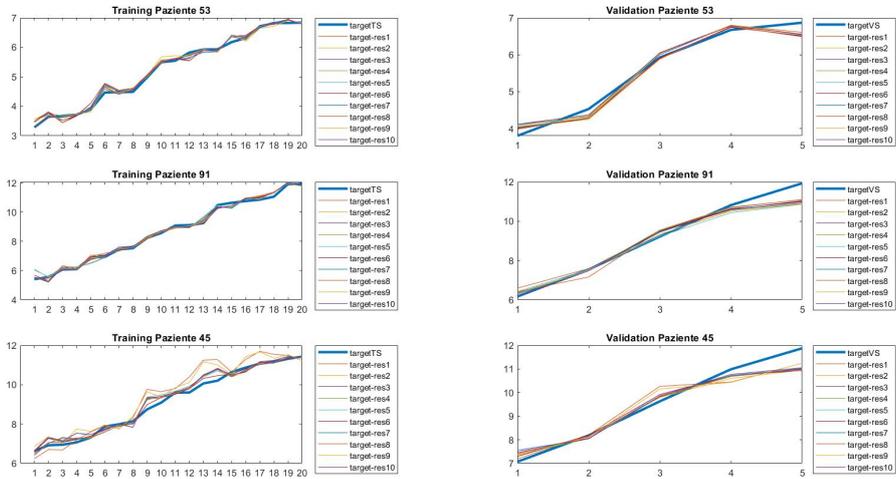


Figura C.11: Pazienti 53-91-45

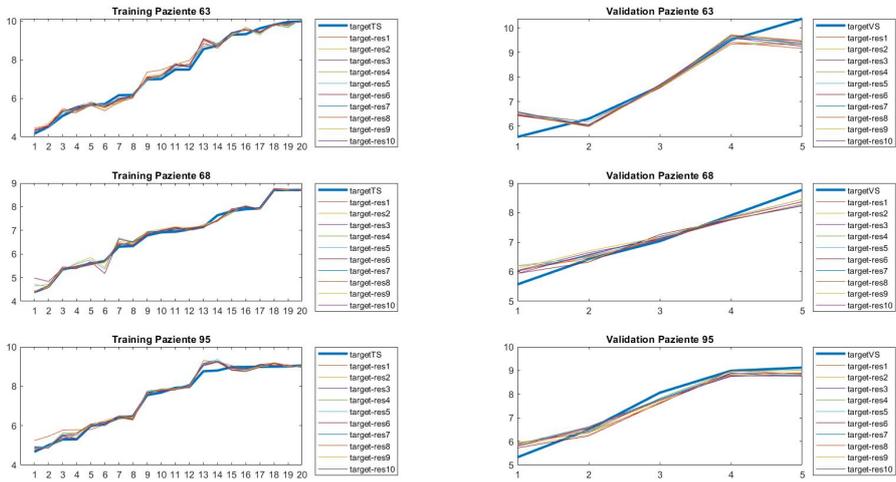


Figura C.12: Pazienti 63-68-95

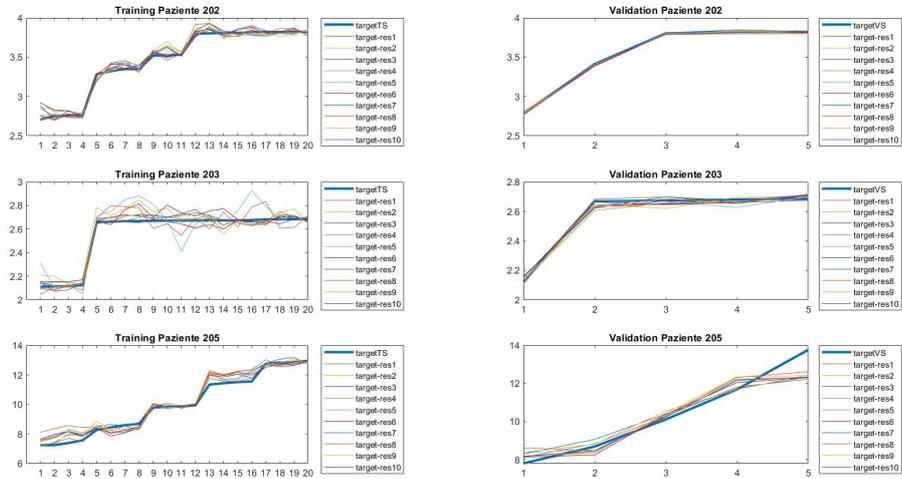


Figura C.13: Pazienti 202-203-205

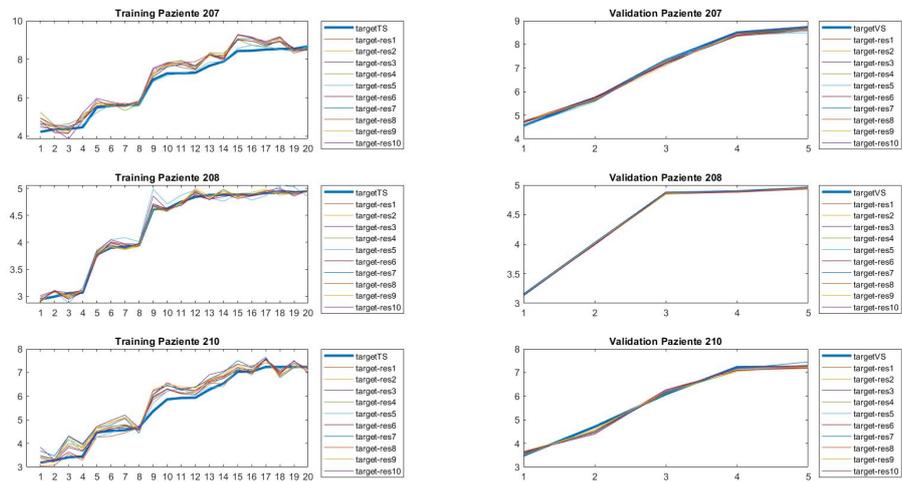


Figura C.14: Pazienti 207-208-210

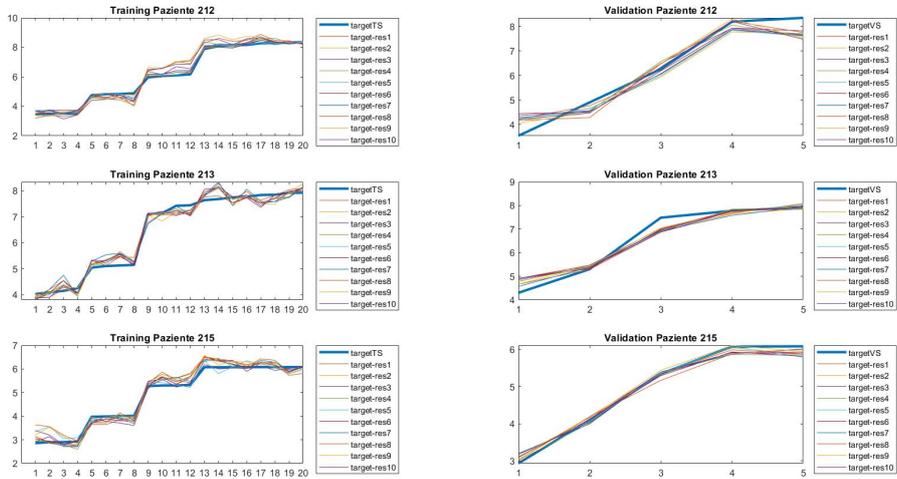


Figura C.15: Pazienti 212-213-215

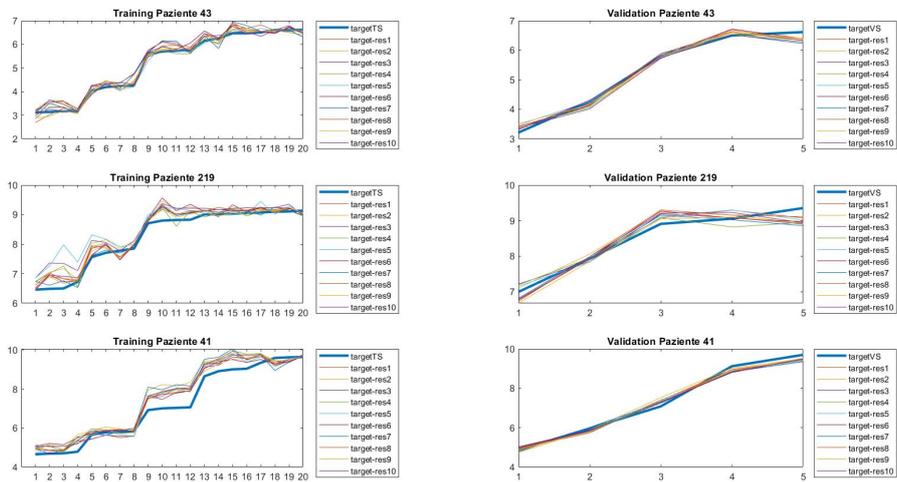


Figura C.16: Pazienti 43-219-41

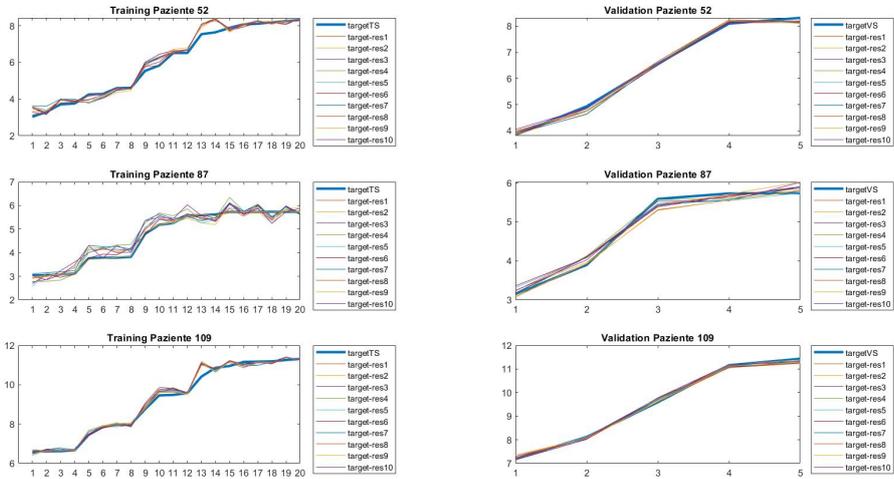


Figura C.17: Pazienti 52-87-109

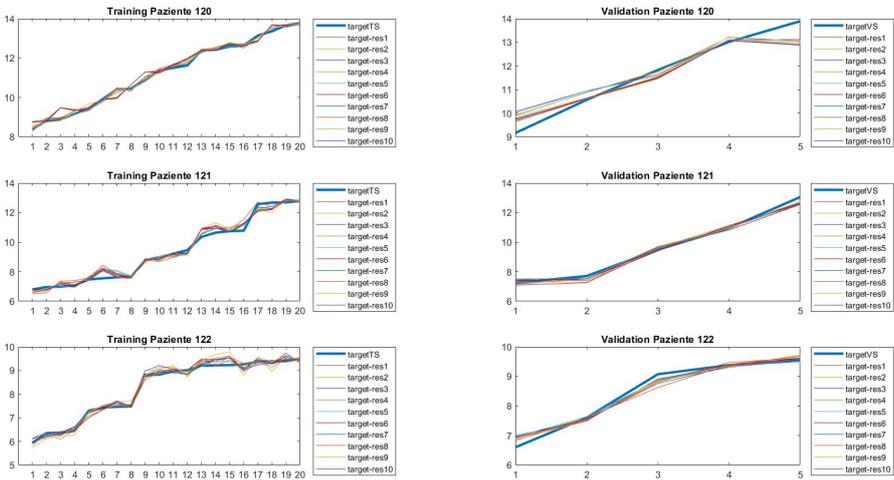


Figura C.18: Pazienti 120-121-122

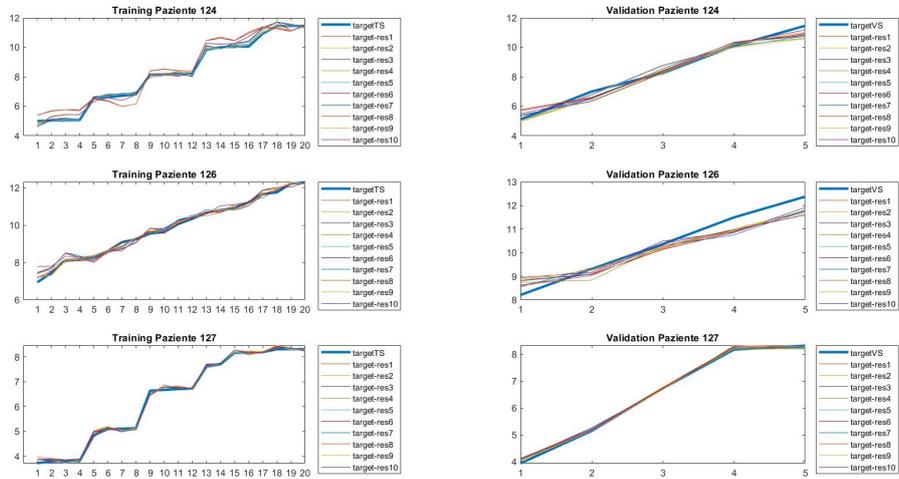


Figura C.19: Pazienti 124-126-127

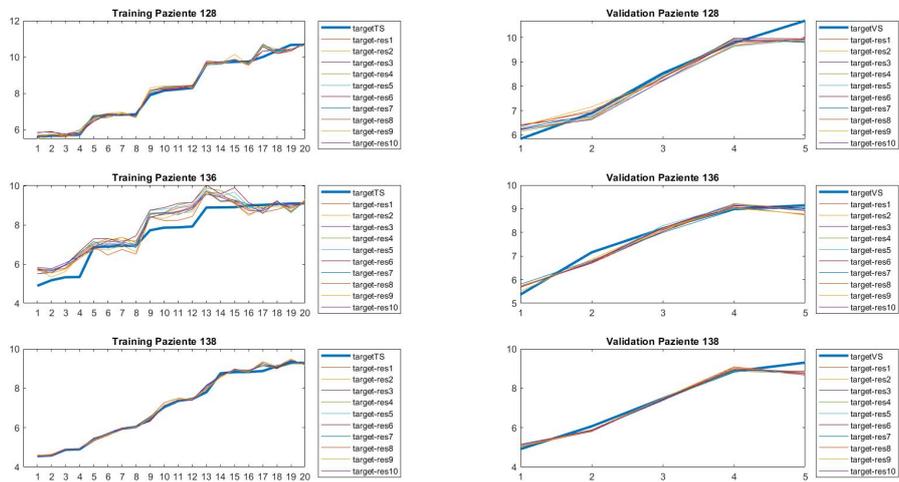


Figura C.20: Pazienti 128-136-138

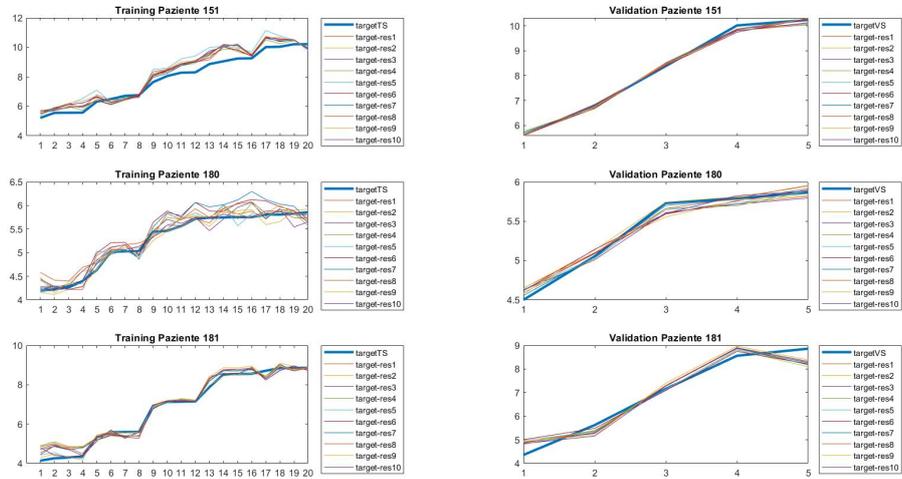


Figura C.21: Pazienti 151-180-181

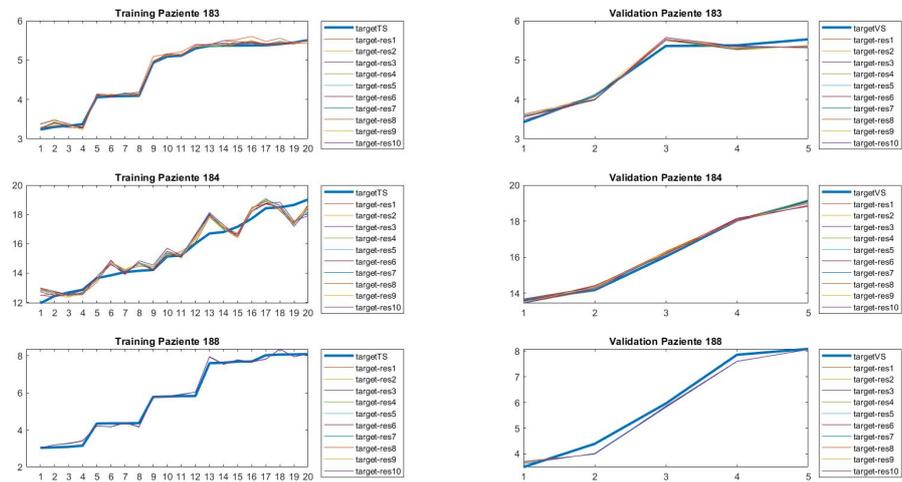


Figura C.22: Pazienti 183-184-188

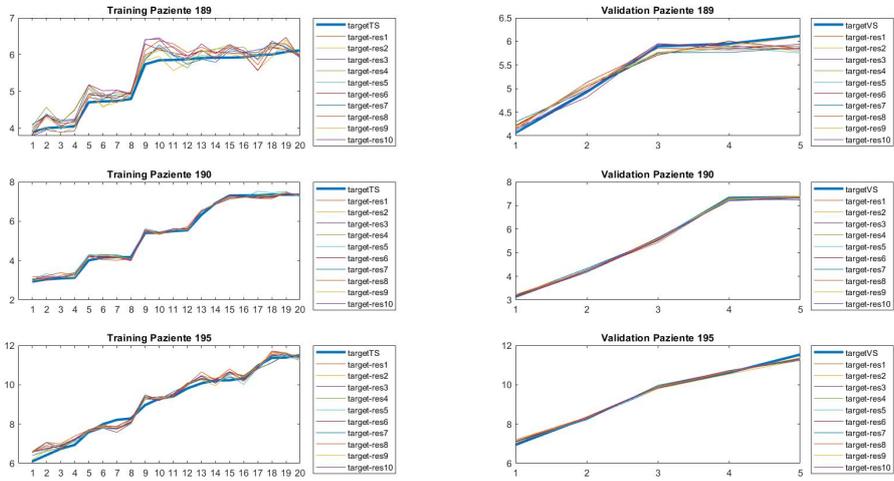


Figura C.23: Pazienti 189-190-195

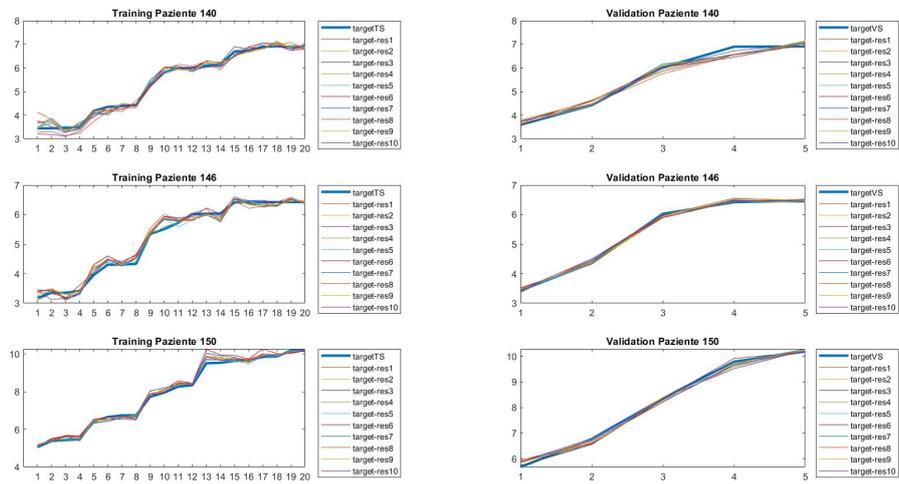


Figura C.24: Pazienti 140-146-150.

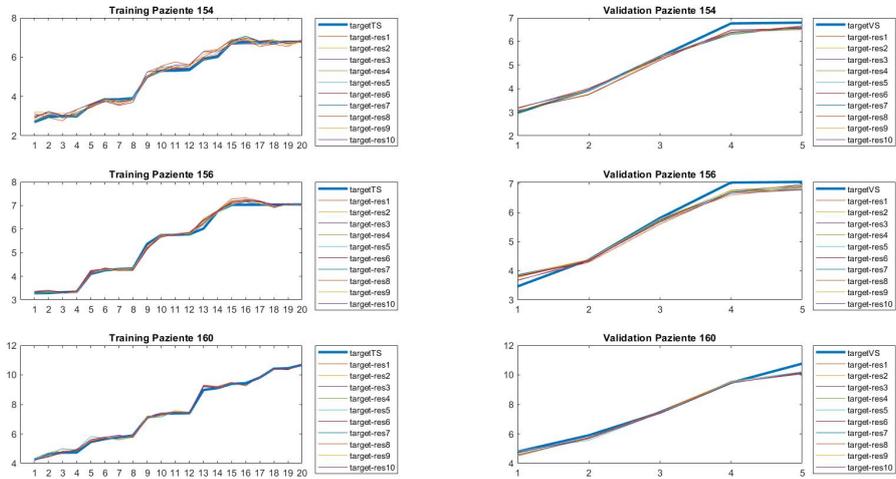


Figura C.25: Pazienti 154-156-160

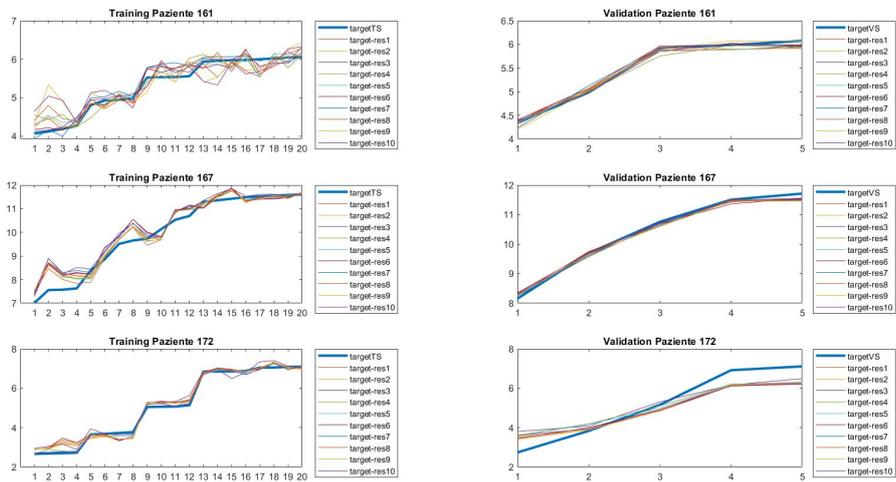


Figura C.26: Pazienti 161-167-172

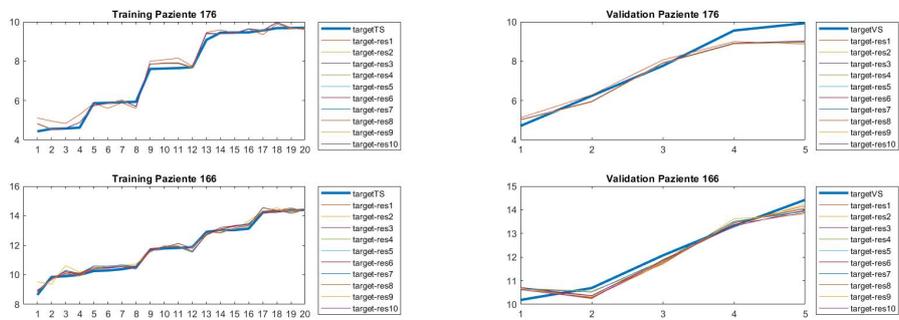


Figura C.27: Pazienti 176-166

Cluster 2

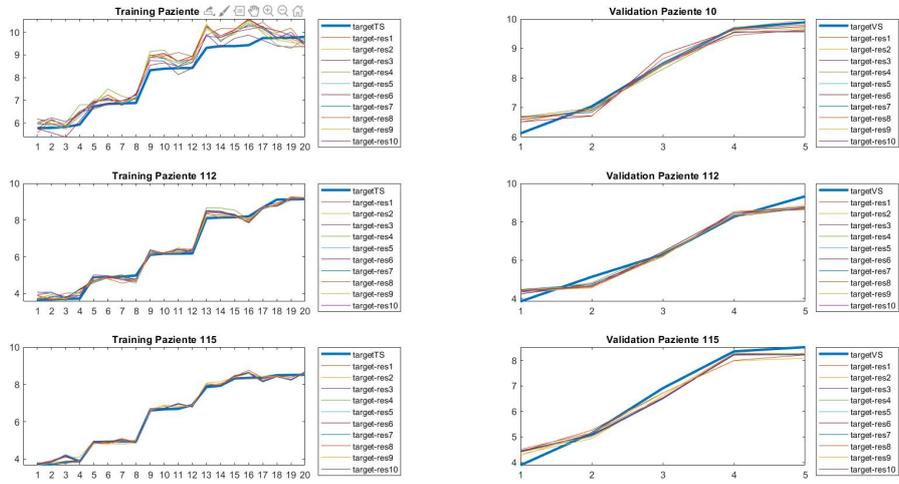


Figura C.28: Pazienti 10-112-115

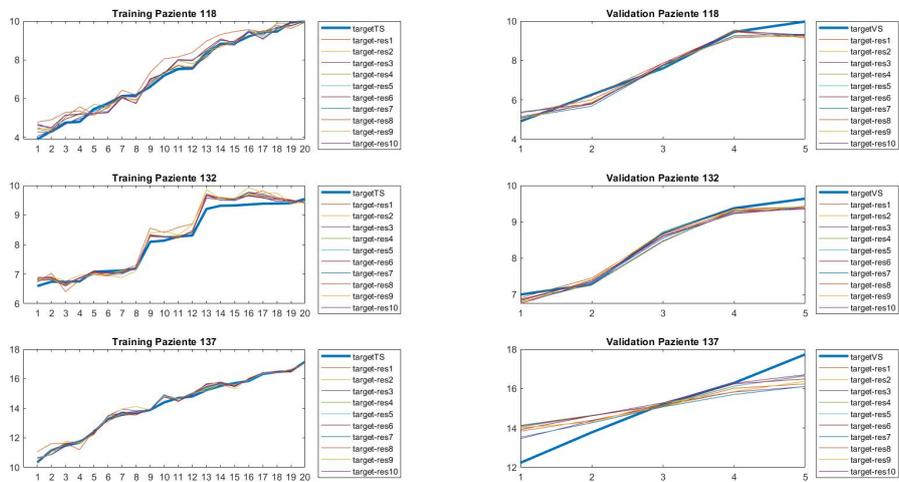


Figura C.29: Pazienti 118-132-137

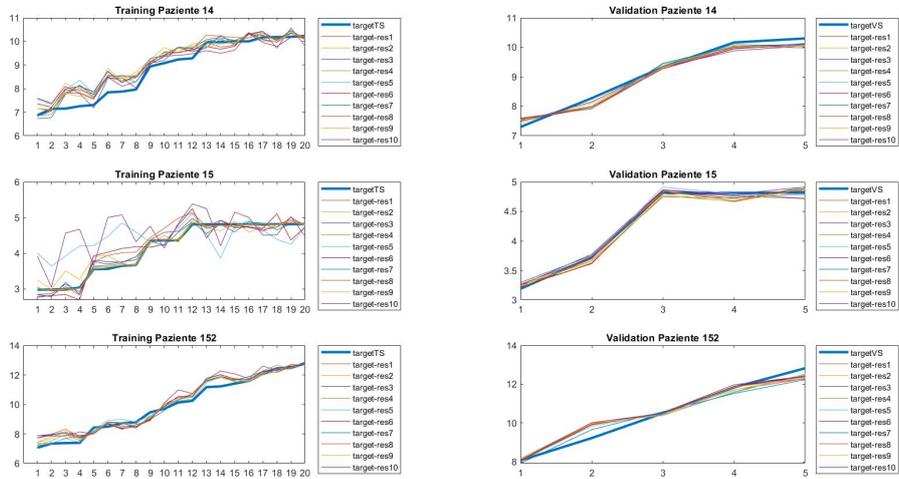


Figura C.30: Pazienti 14-15-152

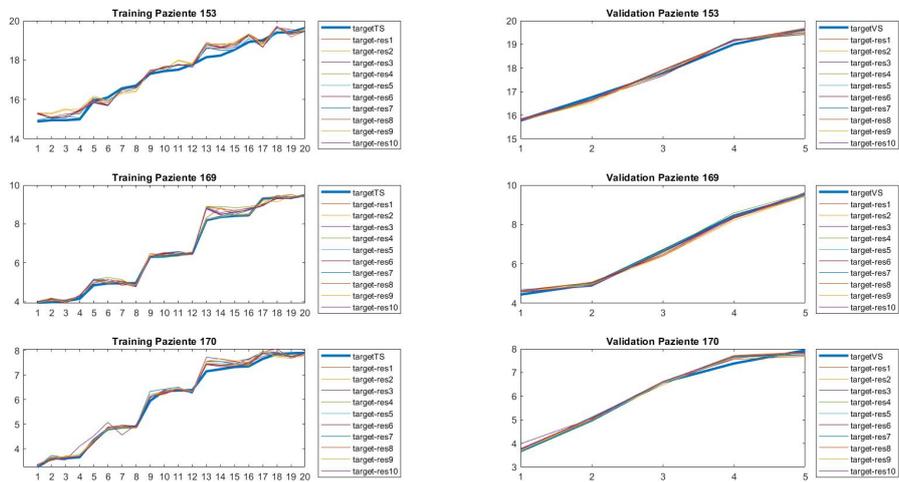


Figura C.31: Pazienti 153-169-170

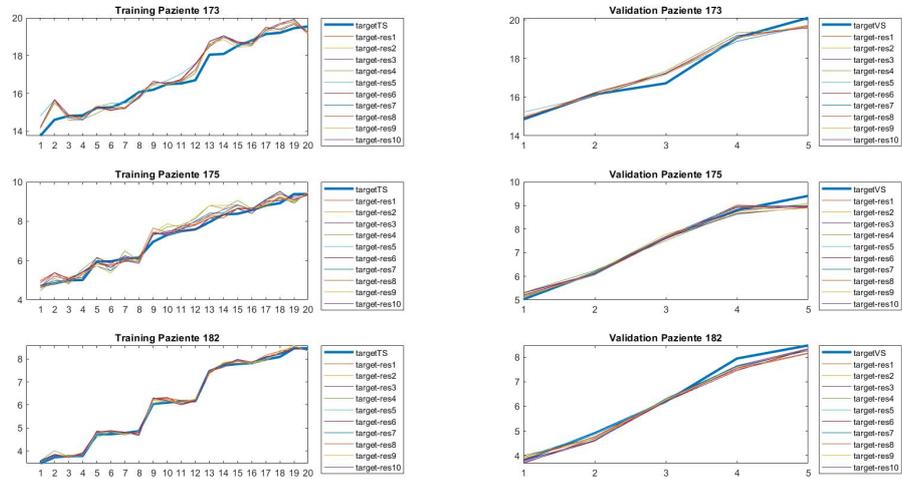


Figura C.32: Pazienti 173-175-182

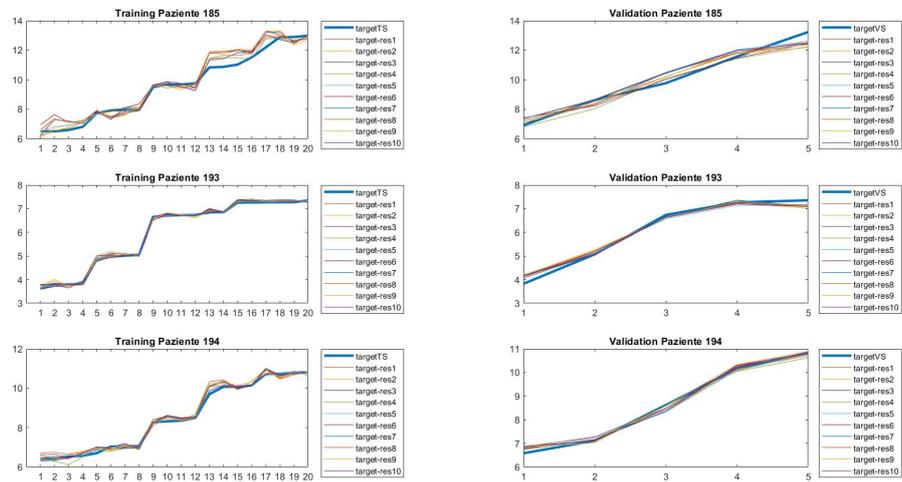


Figura C.33: Pazienti 185-193-194

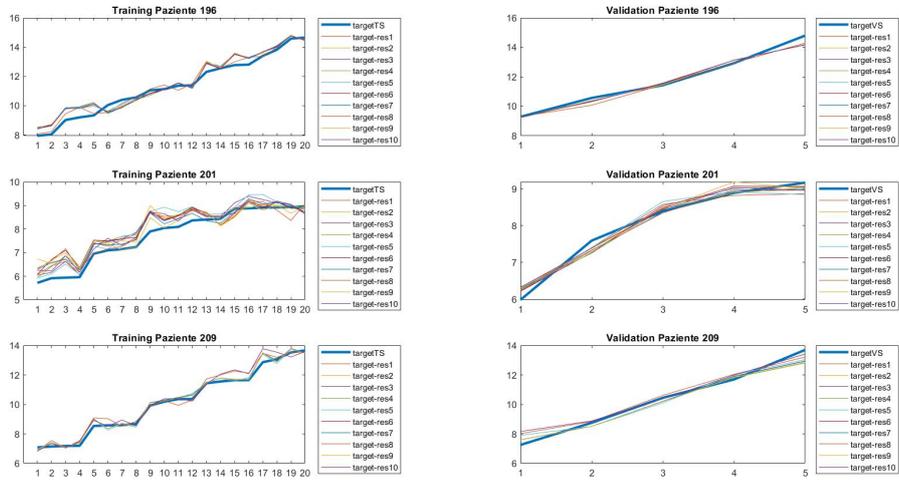


Figura C.34: Pazienti 196-201-209

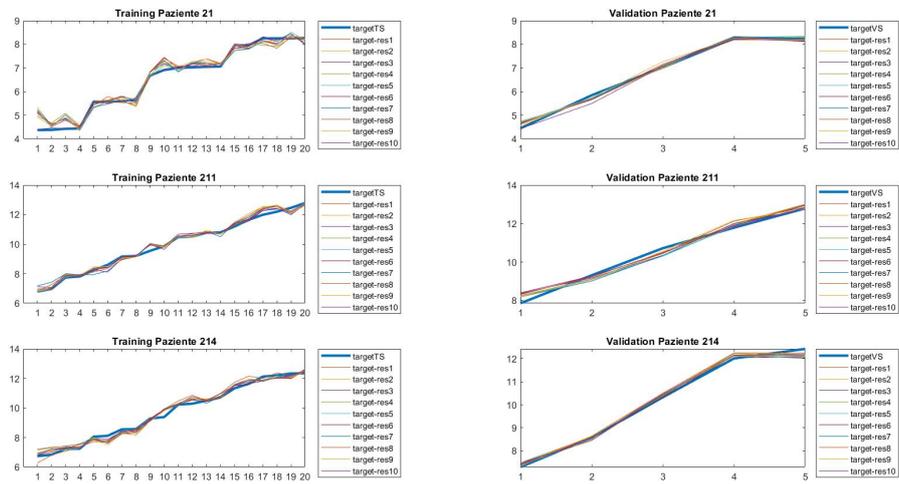


Figura C.35: Pazienti 21-211-214

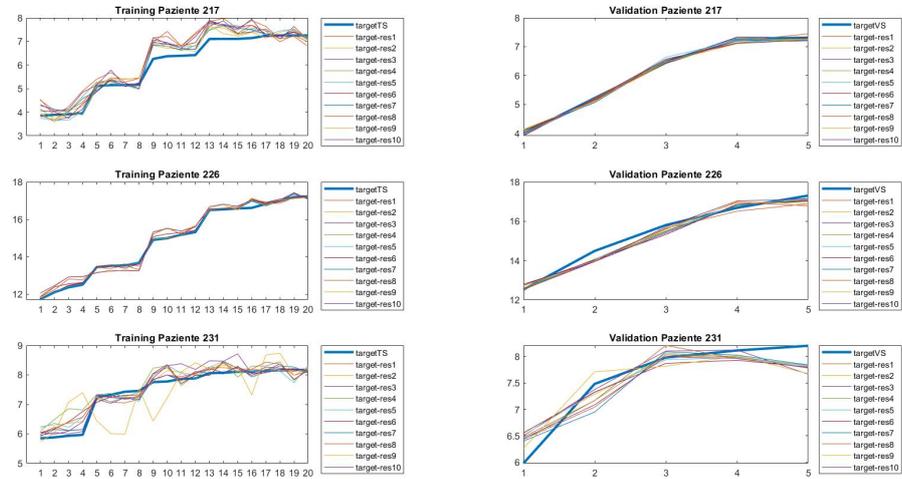


Figura C.36: Pazienti 217-226-231

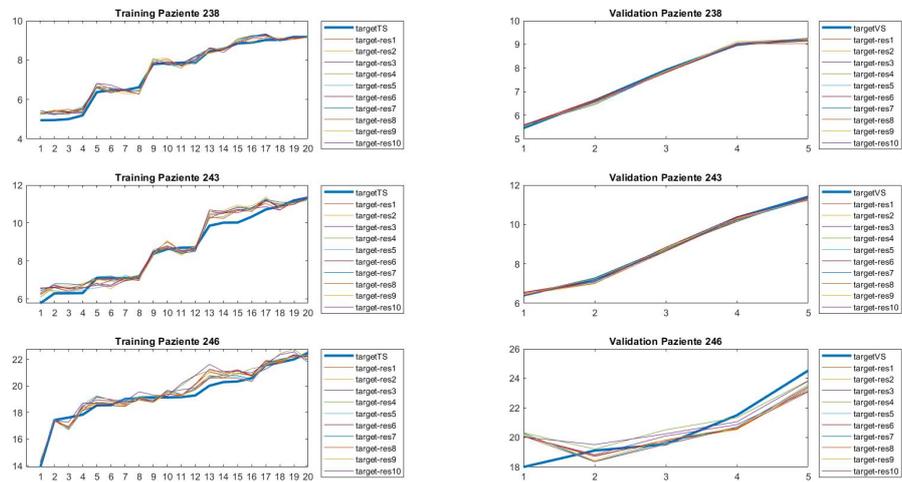


Figura C.37: Pazienti 238-243-246

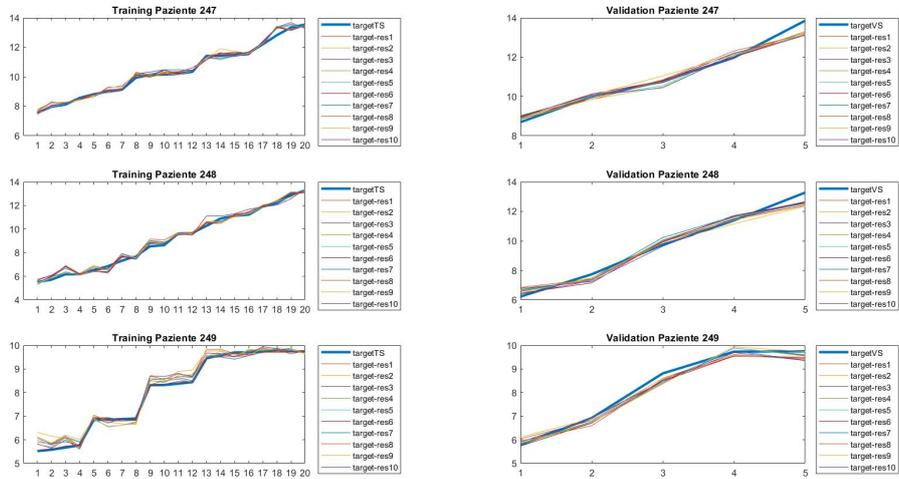


Figura C.38: Pazienti 247-248-249

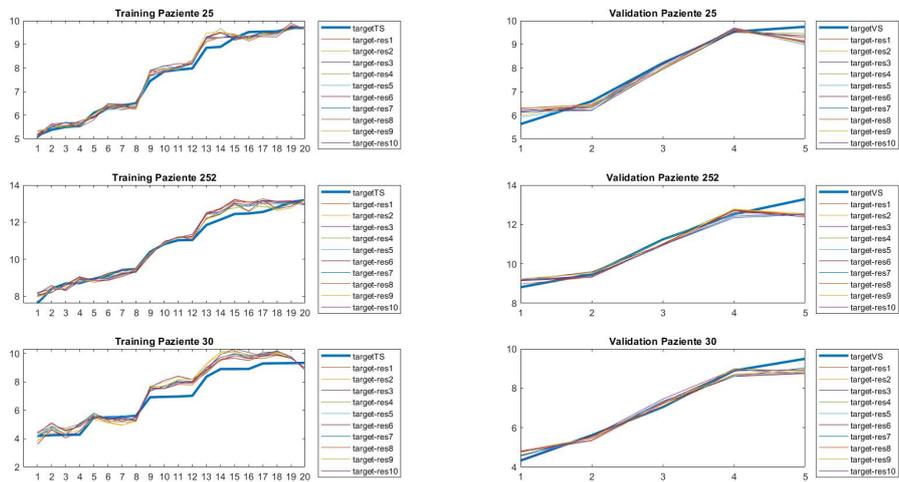


Figura C.39: Pazienti 25-252-30

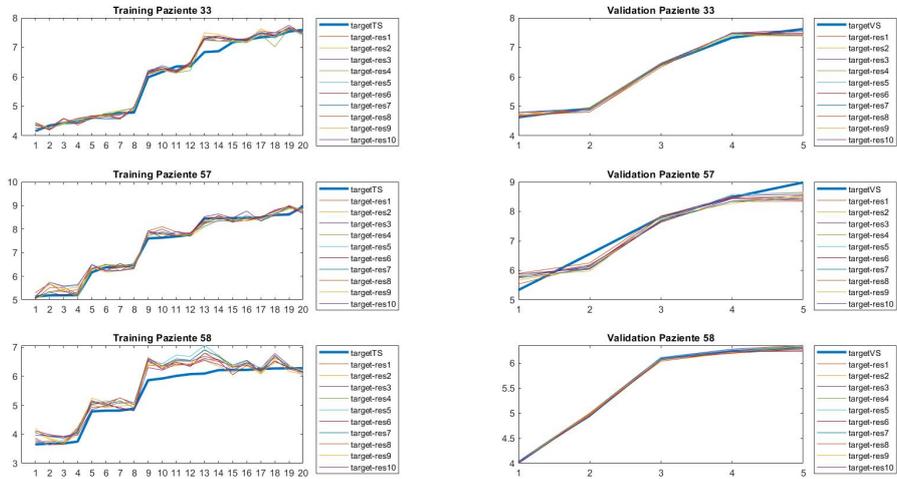


Figura C.40: Pazienti 33-57-58

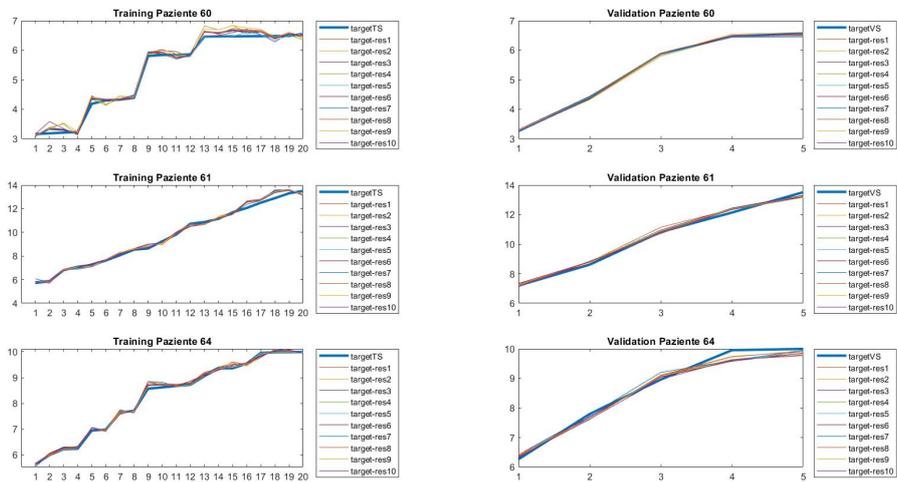


Figura C.41: Pazienti 60-61-64

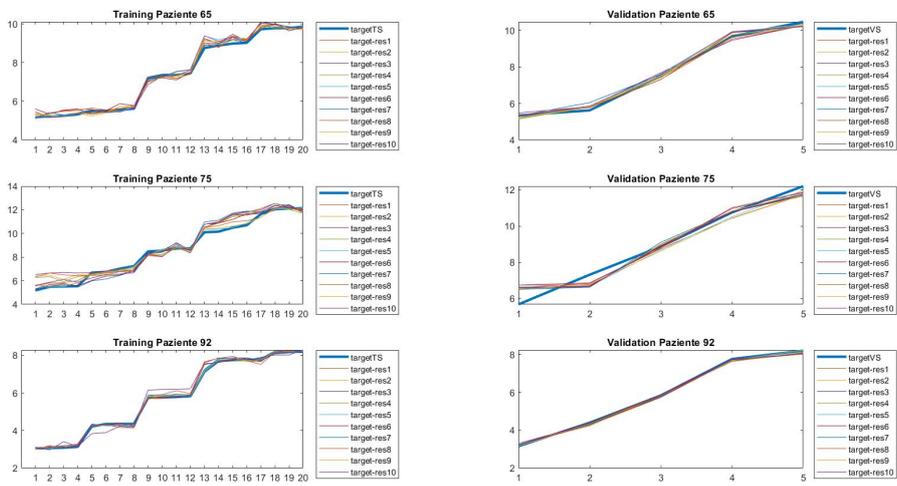


Figura C.42: Pazienti 65-75-92

Cluster 3

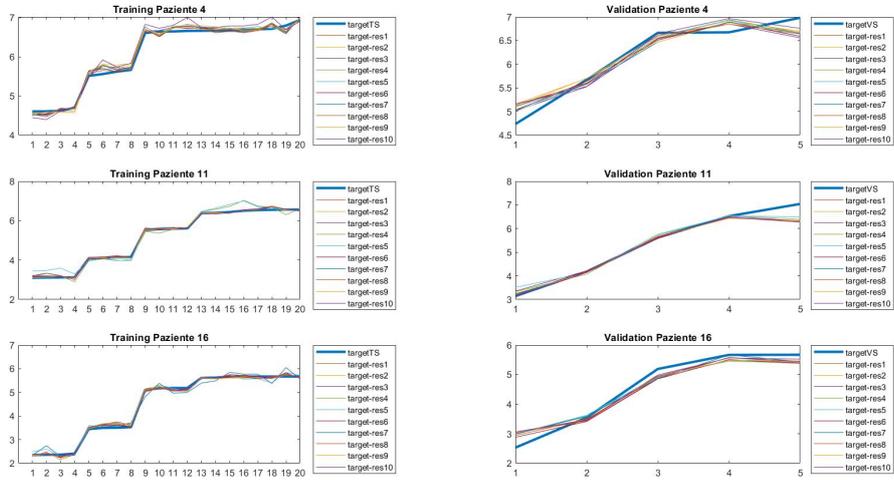


Figura C.43: Pazienti 4-11-16

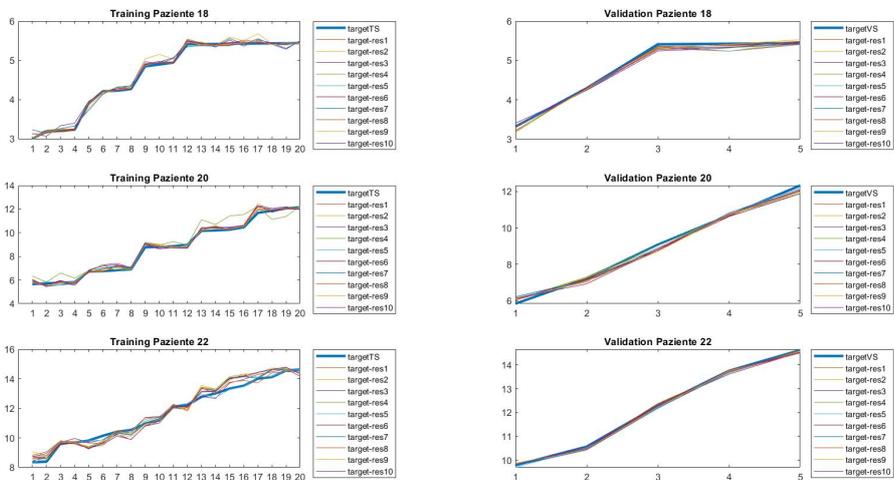


Figura C.44: Pazienti 18-20-22

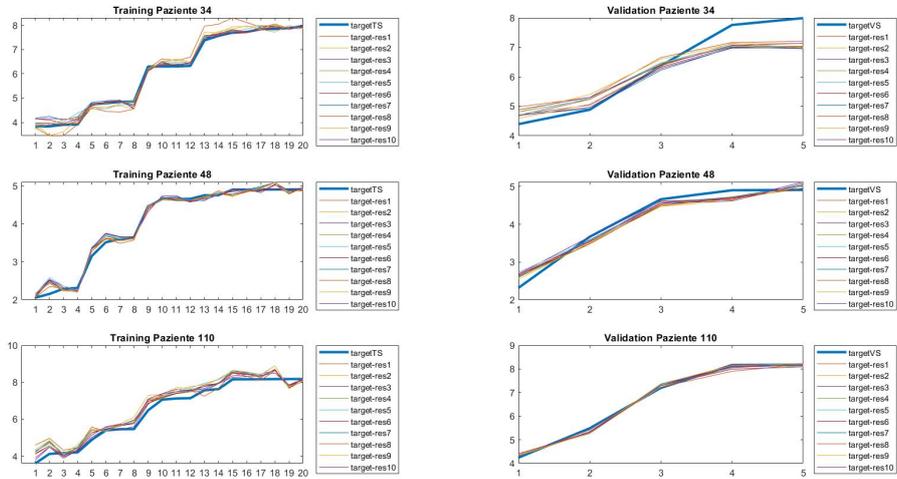


Figura C.45: Pazienti 34-48-110

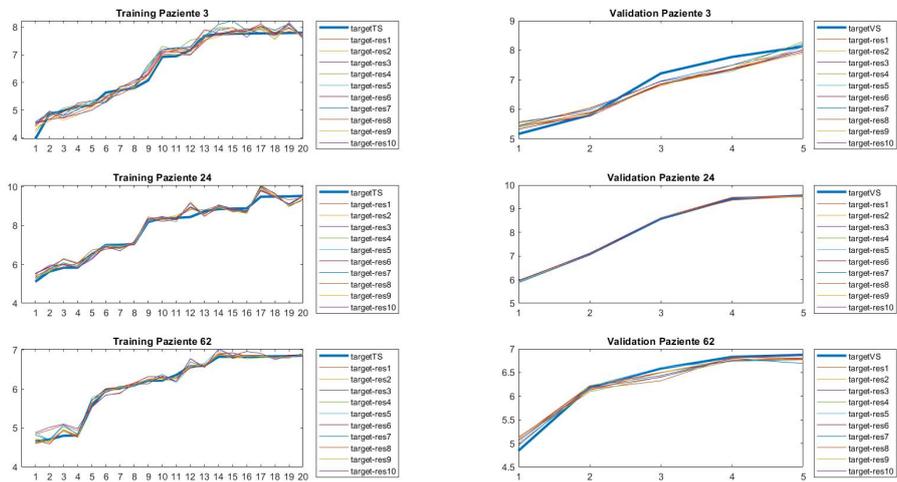


Figura C.46: Pazienti 3-24-62

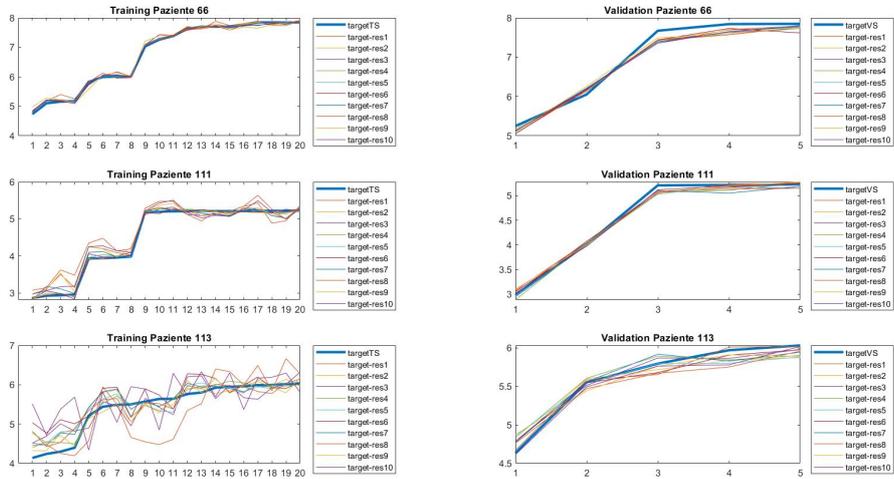


Figura C.47: Pazienti 66-111-113

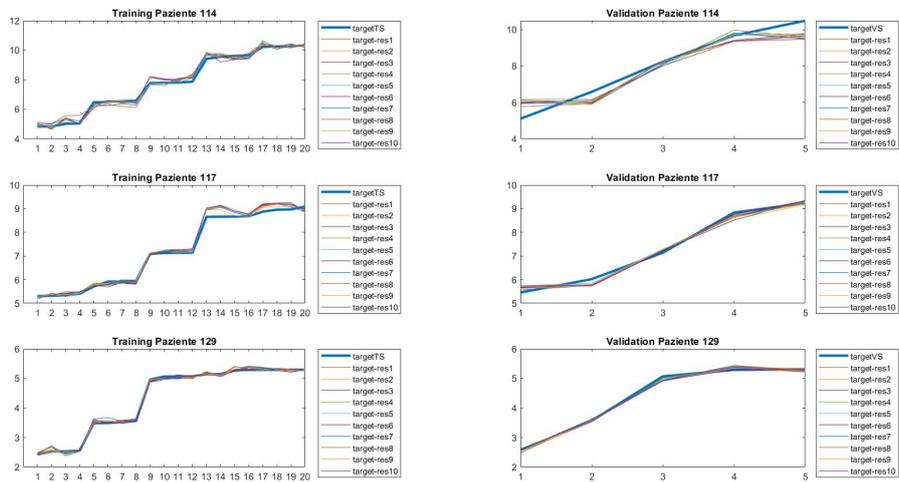


Figura C.48: Pazienti 114-117-129

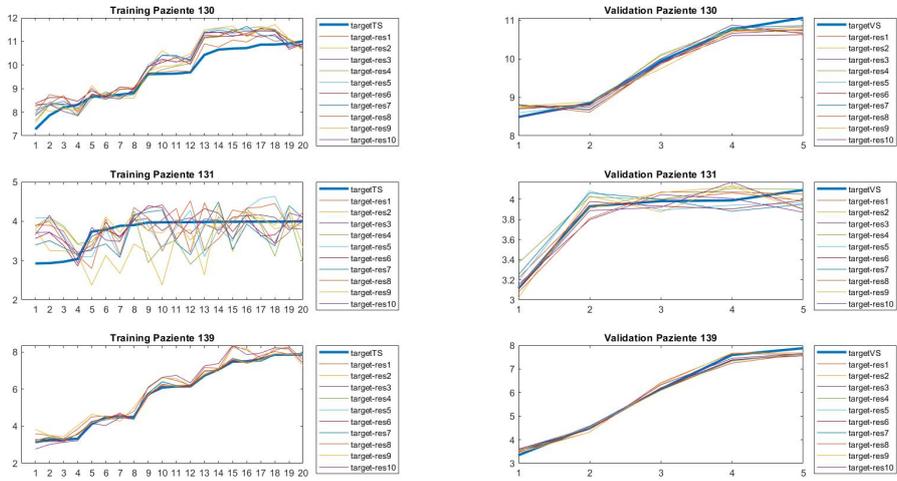


Figura C.49: Pazienti 130-131-139

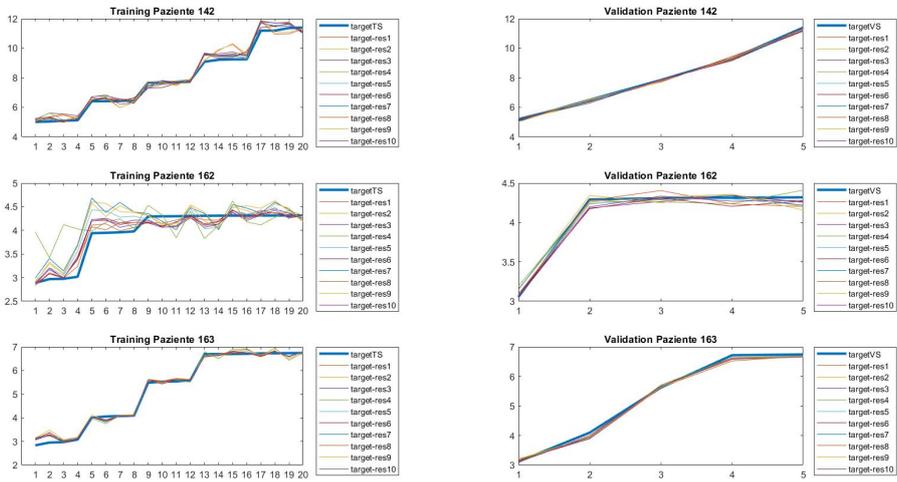


Figura C.50: Pazienti 142-162-163

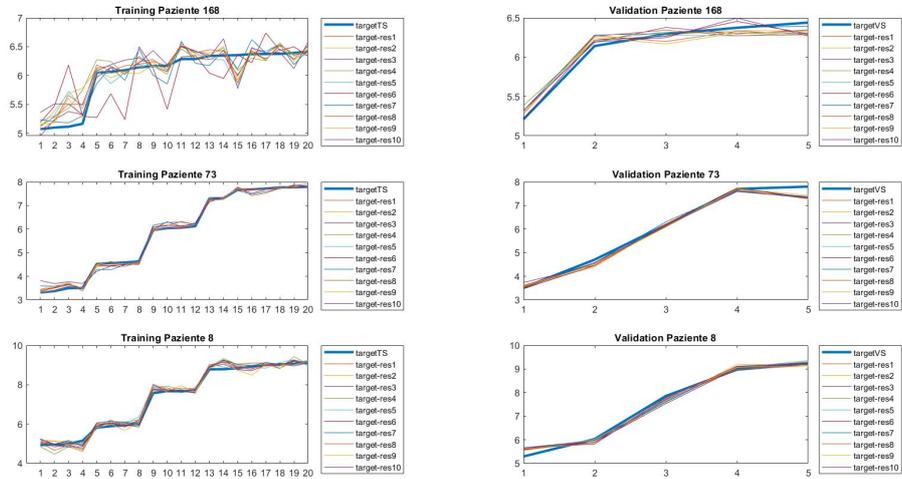


Figura C.51: Pazienti 168-73-8

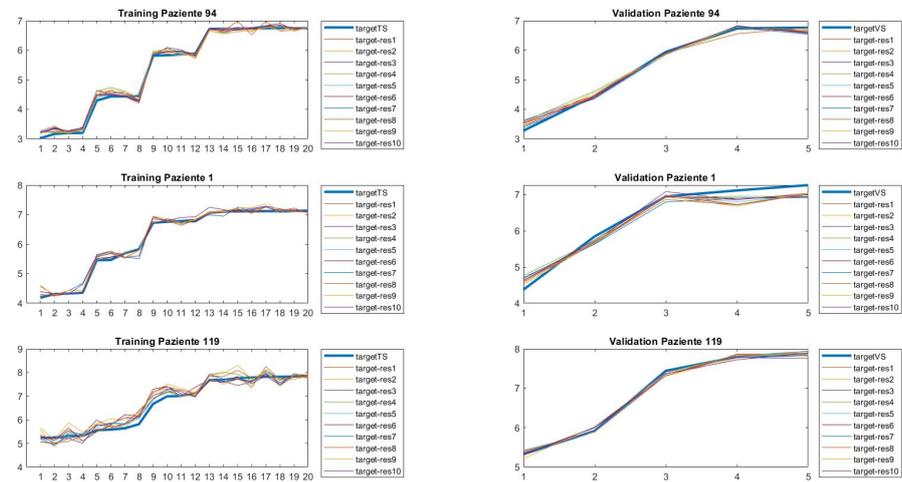


Figura C.52: Pazienti 94-1-119

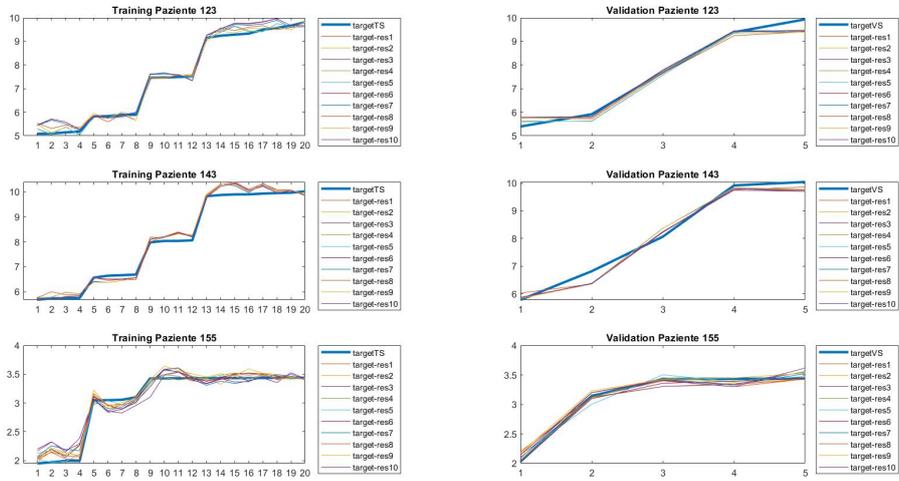


Figura C.53: Pazienti 123-143-155

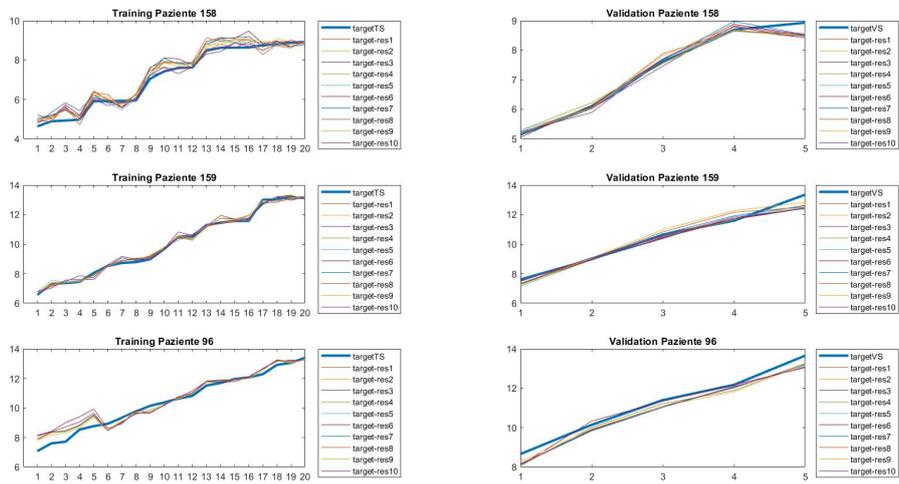


Figura C.54: Pazienti 158-159-96

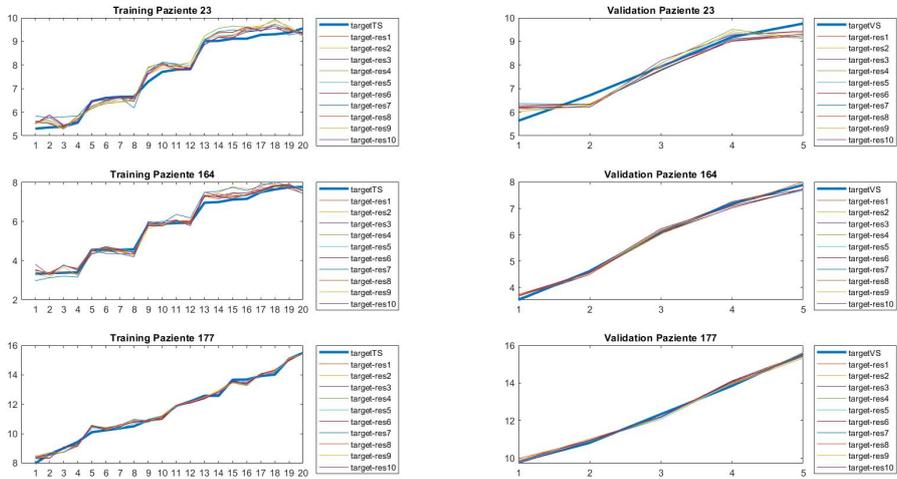


Figura C.55: Pazienti 23-164-177

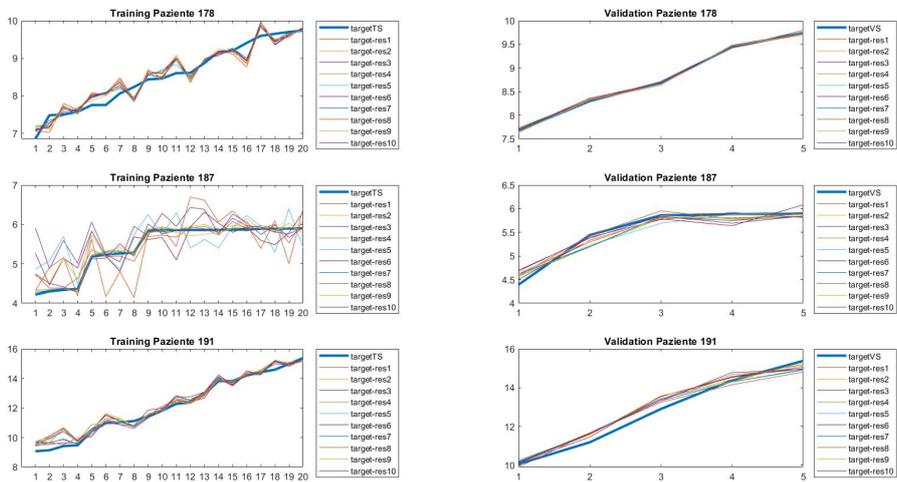


Figura C.56: Pazienti 178-187-191

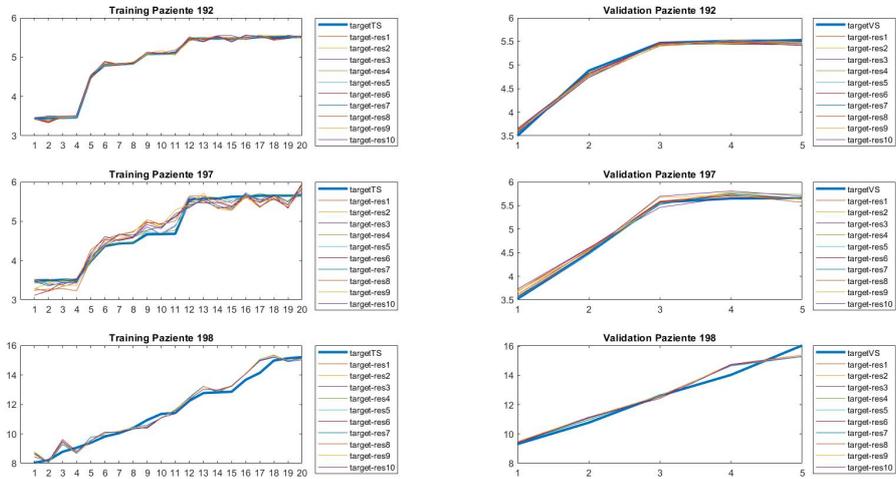


Figura C.57: Pazienti 192-197-198

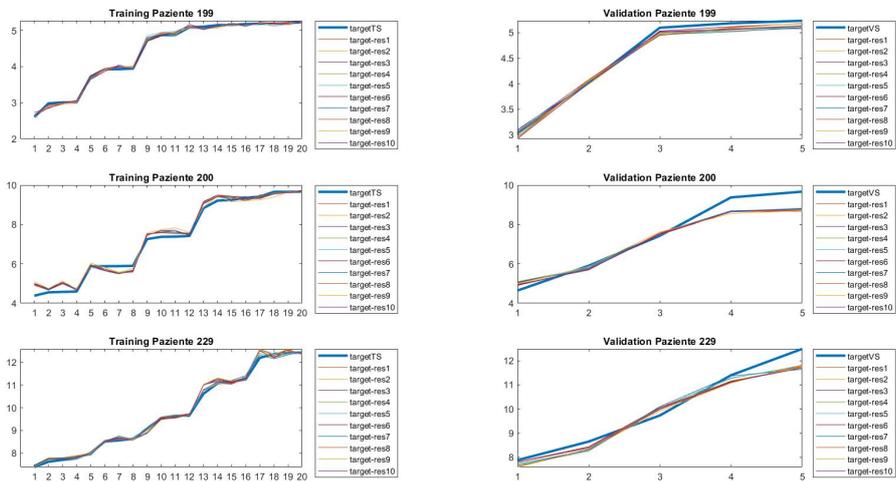


Figura C.58: Pazienti 199-200-229

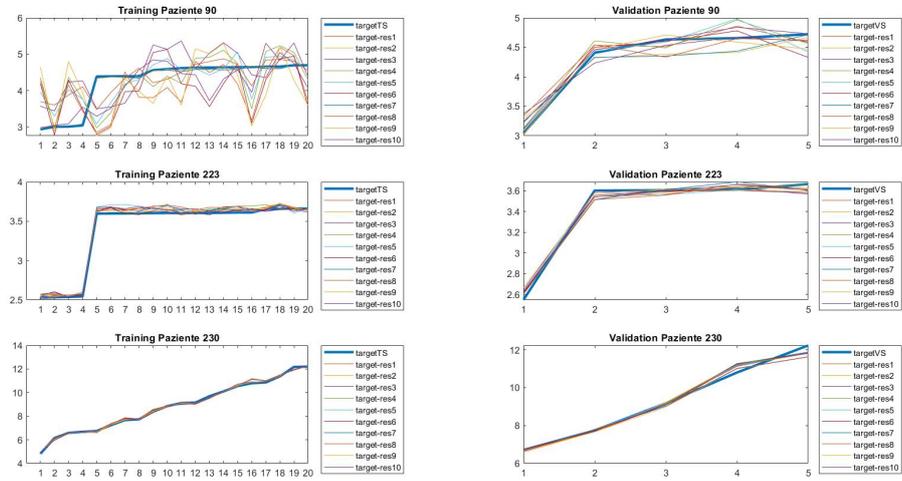


Figura C.59: Pazienti 90-223-230

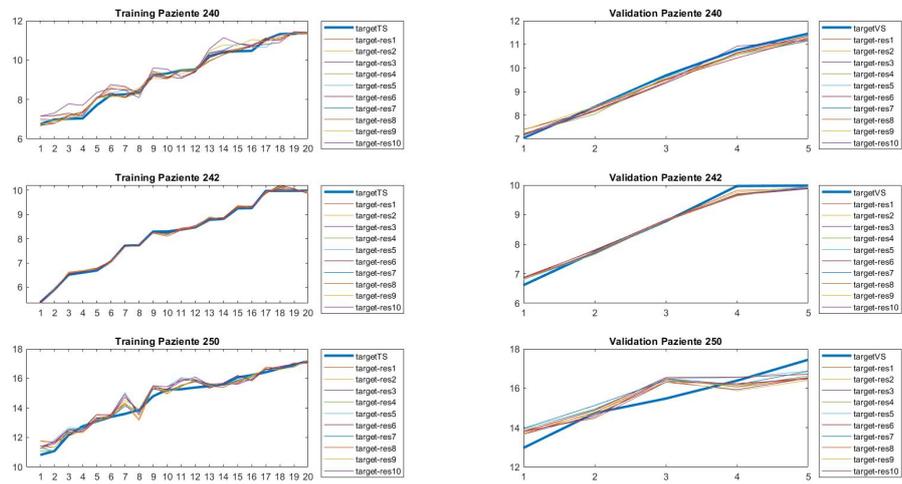


Figura C.60: Pazienti 240-242-250

Appendice D

Risultati tuning parametri della rete RBFNN-Analisi delle combinazioni

Cluster 1

Tabella D.1: Set parametri migliori

Combinazione	Neuroni	η_w	η_c	η_{sigma}	Differenza percentili
414	14	0,2	0,3	0,4	0,1458
471	14	0,4	0,5	0,1	0,1493
294	12	0,2	0,4	0,4	0,1504
480	14	0,5	0,1	0,5	0,1513
595	16	0,4	0,4	0,5	0,1525
321	12	0,3	0,5	0,1	0,1527
259	12	0,1	0,2	0,4	0,1552
491	14	0,5	0,4	0,1	0,1560
311	12	0,3	0,3	0,1	0,1577
495	14	0,5	0,4	0,5	0,1587
260	12	0,1	0,2	0,5	0,1596

Cluster 2

Tabella D.2: Set parametri migliori

Combinazione	Neuroni	η_w	η_c	η_{sigma}	Differenza percentili
112	8	0,5	0,3	0,2	0,1232
117	8	0,5	0,4	0,2	0,1247
227	10	0,5	0,1	0,2	0,1272
111	8	0,5	0,3	0,1	0,1325
250	10	0,5	0,5	0,5	0,1336
476	14	0,5	0,1	0,1	0,1341
233	10	0,5	0,2	0,3	0,1349
116	8	0,5	0,4	0,1	0,1363
226	10	0,5	0,1	0,1	0,1370
228	10	0,5	0,1	0,3	0,1372
240	10	0,5	0,3	0,5	0,1388
115	8	0,5	0,3	0,5	0,1406
206	10	0,4	0,2	0,1	0,1415
375	12	0,5	0,5	0,5	0,1421
118	8	0,5	0,4	0,3	0,1421
241	10	0,5	0,4	0,1	0,1424
210	10	0,4	0,2	0,5	0,1431
79	8	0,4	0,1	0,4	0,1439
486	14	0,5	0,3	0,1	0,1439
232	10	0,5	0,2	0,2	0,1448
493	14	0,5	0,4	0,3	0,1449
88	8	0,4	0,3	0,3	0,1450
346	12	0,4	0,5	0,1	0,1454
357	12	0,5	0,2	0,2	0,1454
223	10	0,4	0,5	0,3	0,1455
246	10	0,5	0,5	0,1	0,1459
218	10	0,4	0,4	0,3	0,1462
495	14	0,5	0,4	0,5	0,1462
236	10	0,5	0,3	0,1	0,1462
114	8	0,5	0,3	0,4	0,1466
229	10	0,5	0,1	0,4	0,1467
93	8	0,4	0,4	0,3	0,1468
353	12	0,5	0,1	0,3	0,1475
65	8	0,3	0,3	0,5	0,1481
234	10	0,5	0,2	0,4	0,1488
225	10	0,4	0,5	0,5	0,1490

Tabella D.2: continua

Tabella D.2: continua

Combinazione	Neuroni	η_w	η_c	η_{sigma}	Differenza percentili
239	10	0,5	0,3	0,4	0,1494
101	8	0,5	0,1	0,1	0,1496
235	10	0,5	0,2	0,5	0,1496
97	8	0,4	0,5	0,2	0,1497

Cluster 3

Tabella D.3: Set parametri migliori

Combinazione	Neuroni	η_w	η_c	η_{sigma}	Differenza percentili
493	14	0,5	0,4	0,3	0,0809
479	14	0,5	0,1	0,4	0,0823
371	12	0,5	0,5	0,1	0,0852
374	12	0,5	0,5	0,4	0,0853
375	12	0,5	0,5	0,5	0,0859
363	12	0,5	0,3	0,3	0,0866
499	14	0,5	0,5	0,4	0,0889
373	12	0,5	0,5	0,3	0,0889
486	14	0,5	0,3	0,1	0,0898
340	12	0,4	0,3	0,5	0,0901
478	14	0,5	0,1	0,3	0,0903
351	12	0,5	0,1	0,1	0,0904
586	16	0,4	0,3	0,1	0,0905
333	12	0,4	0,2	0,3	0,0913
470	14	0,4	0,4	0,5	0,0916
249	10	0,5	0,5	0,4	0,0920
456	14	0,4	0,2	0,1	0,0924
464	14	0,4	0,3	0,4	0,0924
355	12	0,5	0,1	0,5	0,0928
320	12	0,3	0,4	0,5	0,0928
745	18	0,5	0,4	0,5	0,0929
59	8	0,3	0,2	0,4	0,0931
463	14	0,4	0,3	0,3	0,0931
459	14	0,4	0,2	0,4	0,0931
469	14	0,4	0,4	0,4	0,0933
226	10	0,5	0,1	0,1	0,0936
242	10	0,5	0,4	0,2	0,0937
606	16	0,5	0,2	0,1	0,0938

Tabella D.3: continua

Tabella D.3: continua

Combinazione	Neuroni	η_w	η_c	η_{sigma}	Differenza percentili
462	14	0,4	0,3	0,2	0,0942
179	10	0,3	0,1	0,4	0,0946
244	10	0,5	0,4	0,4	0,0946
337	12	0,4	0,3	0,2	0,0950
211	10	0,4	0,3	0,1	0,0950
224	10	0,4	0,5	0,4	0,0953
350	12	0,4	0,5	0,5	0,0957
195	10	0,3	0,4	0,5	0,0958
329	12	0,4	0,1	0,4	0,0964
332	12	0,4	0,2	0,2	0,0966
738	18	0,5	0,3	0,3	0,0966
442	14	0,3	0,4	0,2	0,0968
454	14	0,4	0,1	0,4	0,0968
500	14	0,5	0,5	0,5	0,0970
497	14	0,5	0,5	0,2	0,0970
605	16	0,5	0,1	0,5	0,0970
348	12	0,4	0,5	0,3	0,0971
722	18	0,4	0,5	0,2	0,0972
558	16	0,3	0,2	0,3	0,0973
580	16	0,4	0,1	0,5	0,0974
435	14	0,3	0,2	0,5	0,0975
233	10	0,5	0,2	0,3	0,0977
169	10	0,2	0,4	0,4	0,0977
360	12	0,5	0,2	0,5	0,0979
300	12	0,2	0,5	0,5	0,0979
221	10	0,4	0,5	0,1	0,0982
245	10	0,5	0,4	0,5	0,0983
728	18	0,5	0,1	0,3	0,0984
624	16	0,5	0,5	0,4	0,0985
237	10	0,5	0,3	0,2	0,0985
186	10	0,3	0,3	0,1	0,0985
872	20	0,5	0,5	0,2	0,0985
370	12	0,5	0,4	0,5	0,0986
368	12	0,5	0,4	0,3	0,0989
330	12	0,4	0,1	0,5	0,0994
287	12	0,2	0,3	0,2	0,0997
325	12	0,3	0,5	0,5	0,0998
483	14	0,5	0,2	0,3	0,0998

Appendice E

Risultati GA-Valori delle MaxEUD vescica e retto

Cluster 1

Tabella E.1: Dosi ottimizzate

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
2	1300	1100	1300	1000	830	828	T
5	1700	1000	1400	1000	538	1035	T
6	2800	1000	3900	1000	598	2318	T
7	2000	1000	1700	1000	771	1017	T
9	1400	1000	4400	1000	978	1102	T
12	3800	1000	4100	1000	596	1290	T
13	1500	1000	1800	1000	490	1285	T
17	1600	1000	1500	1000	992	1249	T
26	4900	1000	4400	1000	575	885	T
27	4300	1000	4100	1000	469	850	T
29	1700	1000	1400	1000	1201	966	T
31	1100	1000	1100	1200	1112	954	T
32	1500	1600	1200	1300	1246	1784	T
36	1900	1000	1700	1000	773	547	T
39	4500	1000	4600	1000	670	993	T
41	1200	1000	3800	1100	964	1238	T
43	4000	1000	4400	1000	1076	1012	T
45	2000	1000	1800	1200	1957	1161	T
47	1600	1000	1600	1000	918	851	T

Tabella E.1: continua

Tabella E.1: continua

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
49	1400	1100	1400	1100	1008	752	T
52	1500	1300	1300	1200	1039	949	T
53	1500	1100	1200	1200	1168	947	T
63	1300	1000	1300	1000	653	1294	T
68	1300	1000	1400	1000	1198	874	T
87	3800	1000	3800	1000	724	897	T
89	1500	1000	1400	1000	505	1005	T
91	1500	1000	1600	1000	1625	1138	T
95	1200	1000	1500	1100	1513	1047	T
109	4500	1000	2700	1000	754	1270	T
120	2000	1000	2000	1100	2396	780	T
121	1500	1200	1600	1100	1316	1530	T
122	1500	1000	1500	1000	941	1109	T
124	4100	1000	4100	1000	737	1394	T
126	2100	1000	2200	1000	1582	1109	T
127	1600	1000	1600	1000	824	1177	T
128	4400	1000	4400	1000	969	1326	T
136	2600	1000	1100	1000	907	1528	T
138	1600	1100	1600	1000	1537	1109	T
140	4500	1000	1600	1000	527	920	T
146	1800	1000	1500	1000	760	926	T
150	2000	1000	2200	1000	1376	1081	T
151	1800	1100	3900	1000	1186	1119	T
154	1300	1200	1500	1000	908	947	T
156	2200	1000	2600	1000	832	884	T
160	1000	1000	1200	1000	1079	1230	T
161	1800	1000	3100	1000	598	930	T
166	1300	1000	1200	1000	682	1942	T
167	1000	1000	1000	1000	658	1554	T
172	1700	1000	3600	1000	359	1361	T
176	1300	1100	1400	1000	1041	1285	T
180	3200	1000	3000	1000	557	930	T
181	1100	1000	1400	1000	1063	873	T
183	1600	1000	1700	1000	843	865	T
184	1700	1500	1600	1500	1602	1823	T
188	3100	1000	2900	1000	620	1338	T
189	4000	1000	4300	1000	611	975	T

Tabella E.1: continua

Tabella E.1: continua

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
190	2500	1000	2200	1000	396	1189	T
195	1800	1000	2000	1000	1726	1336	T
202	3000	1000	3200	1000	510	801	T
203	4200	1000	1200	1000	355	443	T
205	1400	1000	1500	1000	867	1663	T
207	1700	1000	1700	1000	855	1113	T
208	1100	1000	1100	1000	685	786	T
210	1700	1000	1500	1000	1839	946	T
212	1900	1000	1900	1000	532	1329	T
213	1700	1000	2800	1000	760	1034	T
215	1600	1000	1700	1100	523	784	T
219	3000	1100	2500	1000	670	1117	T
220	2900	1000	4300	1000	521	829	T
222	1200	1000	1000	1000	1334	1097	T
224	1400	1000	1200	1300	1543	807	T
225	4600	1000	1500	1000	472	730	T
233	1400	1000	1200	1200	1054	1117	T
234	1400	1100	1500	1100	3362	1432	T
235	1000	1000	1000	1400	2031	761	T
236	1300	1000	1500	1000	781	1263	T
237	1300	1100	1300	1100	1300	1884	T
244	4100	1000	1200	1000	1422	1090	T
251	1100	1000	1200	1100	1469	1047	T
253	1700	1100	1000	1000	729	1383	T
145	1500	1000	1500	1100	752	1735	V
179	1300	1000	1500	1100	925	1165	V
221	1800	1300	1700	1300	744	1123	V
227	1700	1000	1700	1000	1616	995	V
88	1700	1000	1600	1000	688	1029	V

Cluster 2

Tabella E.2: Dosi ottimizzate

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
10	2500	1100	4300	1000	750	1334	T
14	1600	1000	1500	1000	691	1243	T
15	2800	1000	4400	1000	467	990	T
21	3800	1000	4700	1000	882	1083	T
25	1600	1000	1700	1100	1241	1138	T
30	3900	1100	3500	1100	823	1225	T
33	1600	1100	1400	1100	1239	1113	T
57	2100	1000	4600	1000	1142	1185	T
58	3700	1000	3600	1000	564	1043	T
60	1700	1000	1900	1000	475	1085	T
61	1600	1300	1500	1200	1731	1300	T
64	1900	1000	1700	1000	1636	986	T
65	4500	1500	4600	1500	226	626	T
75	1900	1200	1400	1400	1512	1485	T
92	1900	1200	1800	1000	517	1069	T
112	2700	1000	2100	1000	545	1531	T
115	4400	1000	4700	1000	927	1200	T
118	1500	1000	1200	1000	1316	1193	T
132	4100	1300	4100	1200	832	1282	T
137	1300	1500	1300	1100	1676	1529	T
152	2500	1400	2700	1000	1755	1106	T
153	1500	1200	1200	1000	2028	1552	T
169	2700	1000	2400	1000	725	1207	T
170	1500	1000	1500	1000	1044	928	T
173	1200	1100	1100	1100	3848	1291	T
175	2100	1100	2100	1000	1501	1112	T
182	1700	1100	1500	1000	800	1113	T
185	1700	1200	1700	1100	958	1395	T
193	1500	1000	1500	1000	686	1135	T
194	1700	1300	1400	1500	573	1431	T
196	1400	1100	1500	1000	817	1563	T
201	1100	1000	1200	1000	1557	1108	T
209	4500	1000	4300	1000	1234	1476	T
211	1400	1000	1400	1000	1814	1129	T
214	1800	1100	1700	1000	1432	1116	T
217	1700	1000	1600	1000	445	1276	T

Tabella E.2: continua

Tabella E.2: continua

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
226	2200	1000	2100	1000	2685	1250	T
231	4000	1000	4000	1000	590	1125	T
238	4400	1000	2300	1000	790	1190	T
243	1700	1500	1200	1000	923	1389	T
246	1000	1200	1000	1000	2450	2006	T
247	1800	1100	1900	1000	1654	1169	T
248	1600	1200	1600	1300	1203	1468	T
249	4600	1100	1700	1000	1104	1215	T
252	1100	1100	1100	1100	852	1439	T
116	1700	1000	1300	1000	1657	1706	V
147	4500	1200	4700	1200	699	1559	V
165	1500	1000	1300	1100	994	1606	V
19	4100	1000	4200	1000	1461	1461	V
232	1800	1000	1700	1000	1118	1842	V

Cluster 3

Tabella E.3: Dosi ottimizzate

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
1	1800	1000	1500	1000	581	1096	T
3	1000	1400	1000	1000	1402	937	T
4	4400	1000	4400	1000	406	895	T
8	4000	1000	4300	1000	702	1172	T
11	1500	1000	1700	1000	923	1066	T
16	4400	1000	3800	1000	356	996	T
18	1300	1000	1100	1000	912	693	T
20	1700	1100	1500	1100	1133	1412	T
22	1600	1100	1300	1000	1449	1667	T
23	4300	1100	4500	1000	507	1300	T
24	1700	1000	1700	1000	1373	757	T
34	4300	1100	2500	1000	930	1074	T
40	1200	1400	1300	1000	919	1092	T
48	1300	1000	1400	1000	706	868	T
62	1500	1000	1700	1000	1256	818	T

Tabella E.3: continua

Tabella E.3: continua

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
66	1100	1100	1100	1200	1146	978	T
73	1300	1100	4000	1000	497	1156	T
90	1000	1000	1500	1000	273	599	T
94	3600	1000	3700	1000	433	1011	T
96	1900	1000	1800	1000	2103	969	T
110	1500	1000	1100	1000	1288	1011	T
111	3000	1000	1200	1000	656	905	T
113	5000	1000	5000	1000	1017	754	T
114	4600	1000	4700	1000	482	1342	T
117	2400	1000	2100	1000	691	1327	T
119	4200	1100	1900	1000	473	1065	T
123	2100	1300	2100	1200	433	1437	T
129	1800	1000	1400	1000	928	937	T
130	1100	1000	1100	1000	908	1406	T
131	4800	1000	1700	1300	471	696	T
139	1400	1000	1500	1000	980	1033	T
142	4300	1000	4200	1000	653	1661	T
143	1600	1000	4100	1200	847	1421	T
155	2400	1000	4000	1000	369	620	T
158	1200	1000	1100	1100	890	1252	T
159	1300	1000	1100	1000	897	1547	T
162	1100	1000	3700	1000	853	594	T
163	2100	1000	2900	1100	292	1047	T
164	4200	1000	4800	1000	860	1110	T
168	2600	1000	2800	1000	633	883	T
177	1300	1200	1300	1100	1598	1549	T
178	1900	1400	1800	1300	1123	1032	T
187	1600	1000	4500	1000	547	916	T
191	2400	1300	2400	1100	1945	1329	T
192	1100	1000	4300	1000	887	813	T
197	3700	1000	1500	1000	1080	802	T
198	1300	1100	1700	1100	1565	1510	T
199	1100	1000	1200	1000	1014	737	T
200	4200	1100	4100	1000	670	1181	T
223	2100	1000	2100	1000	762	573	T
229	1000	1000	1000	1000	960	1310	T
230	1000	1000	1100	1000	1488	1319	T

Tabella E.3: continua

Tabella E.3: continua

ID	MaxEUD RBF		MaxEUD [10]		MaxEUD GPS		T/V
	Vescica	Retto	Vescica	Retto	Vescica	Retto	
240	2400	1000	2200	1000	594	1433	T
242	1400	1000	1200	1000	1797	791	T
245	2100	1100	2200	1000	1382	729	T
250	1000	1700	1000	1300	0	0	T
254	1600	1000	1500	1000	1883	1178	T
125	3900	1000	3600	1000	500	884	V
171	2400	1000	2100	1000	1400	939	V
174	1900	1200	1900	1300	1503	1362	V
218	1400	1000	1400	1000	1069	1498	V
93	1600	1000	1700	1000	612	1050	V

Appendice F

Valutazione dei piani di trattamento

Cluster 1

Tabella F.1: Scores dei piani di trattamento

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
2	100,00	100,00	100,00	T
5	98,87	98,49	98,28	T
6	90,58	81,82	81,82	T
7	100,00	100,00	100,00	T
9	100,00	100,00	100,00	T
12	97,20	100,00	100,00	T
13	100,00	100,00	100,00	T
17	90,18	90,18	90,22	T
26	100,00	100,00	100,00	T
27	100,00	100,00	100,00	T
29	100,00	100,00	100,00	T
31	100,00	100,00	100,00	T
32	81,35	81,29	81,47	T
36	100,00	100,00	100,00	T
39	100,00	100,00	100,00	T
41	100,00	100,00	100,00	T
43	100,00	100,00	100,00	T
45	99,22	96,58	99,04	T
47	100,00	100,00	100,00	T

Tabella F.1: continua

Tabella F.1: continua

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
49	100,00	100,00	100,00	T
52	100,00	100,00	100,00	T
53	100,00	100,00	100,00	T
63	100,00	100,00	100,00	T
68	100,00	100,00	100,00	T
87	100,00	100,00	100,00	T
89	90,02	90,03	90,05	T
91	97,26	100,00	100,00	T
95	100,00	100,00	100,00	T
109	90,33	89,21	89,11	T
120	75,76	75,76	75,76	T
121	90,12	90,07	90,06	T
122	89,68	88,15	88,15	T
124	90,20	90,25	90,25	T
126	98,09	84,71	84,75	T
127	98,44	100,00	100,00	T
128	89,85	98,94	98,94	T
136	90,09	90,02	99,01	T
138	100,00	100,00	100,00	T
140	100,00	100,00	100,00	T
146	100,00	100,00	100,00	T
150	100,00	100,00	100,00	T
151	100,00	100,00	100,00	T
154	100,00	100,00	100,00	T
156	100,00	100,00	100,00	T
160	100,00	100,00	100,00	T
161	97,31	97,46	97,52	T
166	90,46	90,47	90,52	T
167	90,25	90,15	90,15	T
172	100,00	100,00	100,00	T
176	96,76	100,00	100,00	T
180	98,33	90,06	90,03	T
181	100,00	100,00	100,00	T
183	100,00	100,00	100,00	T
184	66,29	66,53	66,44	T
188	100,00	100,00	100,00	T
189	89,92	90,04	90,05	T

Tabella F.1: continua

Tabella F.1: continua

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
190	100,00	100,00	100,00	T
195	97,59	98,46	98,46	T
202	100,00	100,00	100,00	T
203	100,00	100,00	100,00	T
205	90,30	90,48	90,39	T
207	100,00	100,00	100,00	T
208	100,00	100,00	100,00	T
210	100,00	100,00	100,00	T
212	97,10	98,36	98,36	T
213	100,00	100,00	100,00	T
215	100,00	100,00	98,19	T
219	90,35	90,45	90,41	T
220	100,00	100,00	100,00	T
222	100,00	100,00	100,00	T
224	99,22	98,93	98,53	T
225	98,54	98,68	98,69	T
233	98,49	98,19	98,43	T
234	65,97	66,00	66,00	T
235	90,91	90,91	90,91	T
236	98,80	97,98	98,08	T
237	80,56	80,91	80,91	T
244	96,96	96,78	96,82	T
251	100,00	100,00	100,00	T
253	88,27	88,41	88,35	T
145	90,15	90,44	90,41	V
179	100,00	100,00	100,00	V
221	100,00	100,00	100,00	V
227	100,00	100,00	100,00	V
88	100,00	100,00	100,00	V

Cluster 2

Tabella F.2: Scores dei piani di trattamento

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
10	90,12	90,16	90,15	T
14	90,22	90,30	90,33	T
15	100,00	100,00	100,00	T
21	100,00	100,00	100,00	T
25	100,00	100,00	100,00	T
30	98,35	100,00	100,00	T
33	100,00	100,00	100,00	T
57	90,10	98,70	98,70	T
58	98,37	89,62	89,71	T
60	98,45	98,37	98,36	T
61	100,00	100,00	100,00	T
64	90,15	90,17	90,15	T
65	98,69	99,34	99,34	T
75	98,77	100,00	98,34	T
92	100,00	100,00	100,00	T
112	100,00	100,00	100,00	T
115	100,00	100,00	100,00	T
118	100,00	100,00	100,00	T
132	90,29	90,24	90,27	T
137	65,92	65,71	65,81	T
152	100,00	97,73	83,39	T
153	66,38	66,55	66,30	T
169	100,00	100,00	100,00	T
170	100,00	100,00	100,00	T
173	66,00	65,96	66,01	T
175	100,00	100,00	100,00	T
182	100,00	100,00	100,00	T
185	90,14	89,93	89,82	T
193	100,00	100,00	100,00	T
194	89,54	98,74	89,91	T
196	93,39	93,33	92,23	T
201	87,93	87,93	88,05	T
209	88,14	88,14	86,52	T
211	87,20	98,18	98,18	T
214	89,83	90,11	81,06	T
217	98,69	97,46	97,90	T

Tabella F.2: continua

Tabella F.2: continua

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
226	65,94	65,90	65,90	T
231	90,36	90,26	90,26	T
238	90,00	98,64	98,62	T
243	98,45	98,11	98,35	T
246	86,11	77,73	85,85	T
247	90,51	75,76	75,76	T
248	97,84	97,12	97,54	T
249	90,20	90,25	90,27	T
252	90,18	90,19	90,19	T
116	66,05	66,20	66,23	V
147	90,35	90,26	90,33	V
165	81,29	81,38	81,33	V
19	100,00	100,00	100,00	V
232	90,37	90,52	90,60	V

Cluster 3

Tabella F.3: Scores dei piani di trattamento

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
1	98,75	98,94	98,78	T
3	100,00	100,00	100,00	T
4	90,33	90,24	90,24	T
8	89,96	90,04	90,08	T
11	100,00	100,00	100,00	T
16	100,00	100,00	100,00	T
18	100,00	100,00	100,00	T
20	90,01	89,97	89,98	T
22	88,19	88,21	88,33	T
23	98,34	90,10	90,00	T
24	100,00	100,00	100,00	T
34	100,00	100,00	100,00	T
40	100,00	100,00	100,00	T
48	100,00	100,00	100,00	T
62	90,09	90,08	90,05	T

Tabella F.3: continua

Tabella F.3: continua

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
66	90,00	90,05	90,03	T
73	100,00	100,00	100,00	T
90	100,00	98,08	98,17	T
94	89,97	89,96	89,98	T
96	84,81	84,71	84,69	T
110	100,00	100,00	100,00	T
111	100,00	100,00	100,00	T
113	97,25	89,72	89,72	T
114	90,28	90,36	90,29	T
117	90,12	90,44	90,45	T
119	98,75	98,42	97,94	T
123	90,23	90,21	90,26	T
129	100,00	100,00	100,00	T
130	90,22	88,65	88,65	T
131	100,00	100,00	100,00	T
139	100,00	100,00	100,00	T
142	90,17	90,41	90,36	T
143	90,19	90,27	90,33	T
155	100,00	100,00	100,00	T
158	97,70	97,07	97,77	T
159	89,32	97,34	96,99	T
162	100,00	100,00	100,00	T
163	100,00	100,00	100,00	T
164	100,00	100,00	100,00	T
168	90,17	90,01	89,98	T
177	90,04	89,72	89,65	T
178	100,00	100,00	100,00	T
187	89,91	89,80	89,79	T
191	66,05	66,18	66,01	T
192	100,00	96,89	97,13	T
197	100,00	100,00	100,00	T
198	89,68	89,82	89,90	T
199	100,00	100,00	100,00	T
200	90,23	90,22	90,17	T
223	100,00	100,00	100,00	T
229	90,10	89,86	89,86	T
230	97,73	98,69	98,55	T

Tabella F.3: continua

Tabella F.3: continua

Total Score %				
ID	GPS	RBF	RBF [10]	T/V
240	90,50	90,37	90,40	T
242	89,23	90,91	90,91	T
245	100,00	100,00	100,00	T
250	0,00	0,00	0,00	T
254	87,40	89,52	89,47	T
125	100,00	100,00	100,00	V
171	90,20	90,19	90,22	V
174	81,14	78,31	78,54	V
218	90,23	90,24	90,24	V
93	100,00	100,00	100,00	V

Bibliografia

- [1] aimac. *La radioterapia per il cancro della prostata*. 2018. URL: <https://www.aimac.it/libretti-tumore/cancro-prostata/radioterapia-cancro-prostata>.
- [2] Jan L Bruse et al. «A statistical shape modelling framework to extract 3D shape biomarkers from medical imaging data: assessing arch morphology of repaired coarctation of the aorta». In: *BMC medical imaging* 16.1 (2016), p. 40.
- [3] C Aggarwal Charu e K REDDY Chandan. *Data clustering: algorithms and applications*. 2013.
- [4] Nadja Ebert, Falk Tillner e Michael Baumann. «Radiation Oncology». In: (2019).
- [5] Andries P Engelbrecht. *Computational intelligence: an introduction*. John Wiley & Sons, 2007.
- [6] Guojun Gan, Chaoqun Ma e Jianhong Wu. *Data clustering: theory, algorithms, and applications*. SIAM, 2007.
- [7] Humanitas Research Hospital. *Tumore della prostata*. URL: <https://www.humanitas.it/malattie/tumore-della-prostata>.
- [8] RaySearch Laboratories. *Raystation*. URL: <https://www.raysearchlabs.com/raystation/>.
- [9] Toru Matsugasumi et al. «Morphometric analysis of prostate zonal anatomy using magnetic resonance imaging: impact on age-related changes in patients in Japan and the USA». In: *BJU international* 120.4 (2017), pp. 497–504.
- [10] Valentina Perolini. «Utilizzo di algoritmi genetici per l’ottimizzazione di piani per radioterapia= Use of genetic algorithms for the optimization of radiotherapy plans». Tesi di dott. Politecnico di Torino, 2018.
- [11] SN Sivanandam e SN Deepa. «Genetic algorithms». In: *Introduction to genetic algorithms*. Springer, 2008, pp. 15–37.
- [12] Yue Wu et al. «Using radial basis function networks for function approximation and classification». In: *ISRN Applied Mathematics* 2012 (2012).