

POLITECNICO DI MILANO

Facoltà di Ingegneria dei Sistemi

Corso di Laurea Specialistica in Ingegneria Biomedica



TESI

**Sviluppo e validazione di metodi per la compensazione in
linea dei movimenti d'organo intra-frazione in adroterapia**

Relatore: Prof. Guido BARONI

Correlatori: Ing. Andrea PELLA

Dott. Ing. Marco RIBOLDI

Autore:

Matteo SEREGNI

Matr. n. 734930

Anno Accademico 2009 - 2010

INDICE

1. SOMMARIO	4
1. SUMMARY	11
2. INTRODUZIONE	17
2.1. INTRODUZIONE ALLA RADIOTERAPIA A FASCI ESTERNI	17
2.2. INTRODUZIONE ALL'ADROTERAPIA	20
2.2.1. CENNI STORICI	20
2.2.2. PRINCIPI FISICI	22
2.3. GENERAZIONE DEL FASCIO RADIANTE	28
2.4. STRATEGIE DI DEPOSITO DELLA DOSE	34
2.4.1. TRATTAMENTI NON CONVENZIONALI NELLA RADIOTERAPIA CON FOTONI	34
2.4.2. SISTEMI PER IL RILASCIO DELLA DOSE IN ADROTERAPIA	36
2.5. CONFRONTO TRA RADIOTERAPIA CON FOTONI E CON ADRONI	42
2.5.1. EFFETTI BIOLOGICI	42
2.5.2. DEPOSITO DELLA DOSE	52
3. STRATEGIE DI TRATTAMENTO PER LESIONI MOBILI	56
3.1. ORGAN MOTION	57
3.2. RADIOTERAPIA GUIDATA DA IMMAGINI	58
3.3. ADROTERAPIA GUIDATA DA IMMAGINI	61
3.3.1. SOLUZIONI TECNOLOGICHE	62
3.3.2. IMPLEMENTAZIONE DI SISTEMI DI <i>IMAGE GUIDANCE</i> PRESSO CNAO	67
3.4. LOCALIZZAZIONE DINAMICA DEL BERSAGLIO	71
3.4.1. IL SISTEMA <i>SYNCHRONY™</i> PER <i>CYBERKNIFE</i> .	72
3.4.2. SISTEMI DI <i>TRACKING</i> IN ADROTERAPIA	74
3.5. SCOPO DEL PROGETTO	76
4. STRUMENTI UTILIZZATI	79
4.1. RETI NEURALI	79
4.1.1. INTRODUZIONE	79
4.1.2. PROCEDURE DI ADDESTRAMENTO	84
4.2. MODELLI DI STATO	90
4.2.1. INTRODUZIONE	90
4.2.2. STIMA DEI PARAMETRI DEL MODELLO	92
4.3. DATABASE UTILIZZATO	93
5. SVILUPPO DEL PROGETTO	94
5.1. STUDIO PRELIMINARE	97
5.1.1. RETE NEURALE	99
5.1.2. MODELLO A VARIABILI DI STATO	104
5.1.3. INTERFACCIA GRAFICA	110

5.1.4. RISULTATI	114
5.2. REALIZZAZIONE DEL SOFTWARE DI CORRELAZIONE	117
5.2.1. ALGORITMO DI ADDESTRAMENTO DELLE RETI NEURALI	118
5.2.2. VALUTAZIONE DELL'ERRORE DI STIMA	126
5.2.3. ACQUISIZIONE DEI DATI E CALCOLO DELLA STIMA <i>ONLINE</i>	132
5.3. FANTOCCIO PER LA SIMULAZIONE DEL MOVIMENTO RESPIRATORIO	138
6. VALIDAZIONE	144
6.1. CONFRONTO CON <i>CYBERKNIFE</i>	144
6.2. ACQUISIZIONI SU FANTOCCIO PER LA SIMULAZIONE DEL MOVIMENTO RESPIRATORIO	150
7. DISCUSSIONE DEI RISULTATI E SVILUPPI FUTURI	162
7.1. CONFRONTO CON IL SISTEMA <i>CYBERKNIFE</i>	164
7.2. ATTIVITÀ SPERIMENTALE SU FANTOCCIO	167
7.3. OSSERVAZIONI CONCLUSIVE	172
8. BIBLIOGRAFIA	173
INDICE DELLE FIGURE	178
INDICE DELLE TABELLE	181

1. Sommario

La radioterapia a fasci esterni prevede il trattamento di lesioni neoplastiche attraverso l'irradiazione con onde elettromagnetiche ad alta energia, quali i raggi X. L'energia trasferita al tessuto irradiato comporta il danneggiamento del materiale genetico cellulare, impedendo la riproduzione delle cellule colpite. Poiché tale danneggiamento interessa sia il tessuto neoplastico sia i tessuti sani, il principio fondamentale della radioterapia consiste nel causare il maggior danno possibile alla lesione, risparmiando, nello stesso tempo, i tessuti sani circostanti. A tal fine, ogni trattamento radioterapico deve essere accuratamente pianificato sulla base di dettagliate informazioni riguardanti la posizione e l'estensione del volume bersaglio.

Una recente tecnica di radioterapia, chiamata adroterapia, consiste nell'irradiazione di lesioni tumorali utilizzando fasci di particelle cariche accelerate, tipicamente protoni o ioni carbonio. L'interazione tra le particelle cariche e la materia, in termini di deposizione di dose, è caratterizzata dalla cosiddetta "curva (o picco) di Bragg" [Figura 1.1].

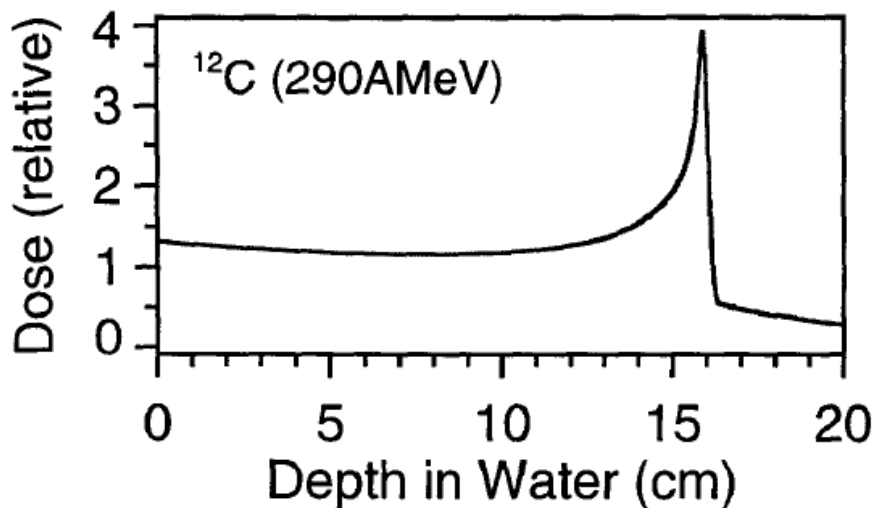


Figura 1.1: Curva di Bragg per gli ioni carbonio accelerati a 290 MeV/u.

Tale curva esprime la dose depositata in funzione della profondità di penetrazione e presenta un plateau iniziale, in cui la dose ceduta al mezzo è molto ridotta, e un picco prossimo alla profondità corrispondente al *range* energetico della particella. Oltre tale profondità di penetrazione, l'energia ceduta risulta limitata ed è dovuta ad eventuali frammenti leggeri. Inoltre, la profondità alla quale si colloca il picco di Bragg può essere modulata variando l'energia del fascio, in accordo con la profondità alla quale si colloca il *target*.

Rispetto alla radioterapia con fotoni, l'adroterapia presenta due vantaggi fondamentali. Il primo, conseguente all'andamento della curva di Bragg, consiste nella capacità di depositare

dosi elevate anche in bersagli profondi, riducendo nel contempo il danneggiamento dei tessuti sani collocati lungo la traiettoria del fascio, sia anteriormente che posteriormente al *target*. La principale conseguenza di tale potenzialità consiste nella possibilità di conformare accuratamente la deposizione di dose sul volume bersaglio, consentendo di eseguire trattamenti caratterizzati da un'elevata selettività geometrica.

Il secondo vantaggio, invece, consiste nella maggiore efficacia biologica relativa (RBE) che i fasci di particelle cariche (in particolare di ioni carbonio) mostrano rispetto ai fotoni. Questo implica che, a parità di dose depositata, un fascio di particelle causa danni al tessuto biologico maggiori di quelli causati da un fascio di raggi X, consentendo un più efficace controllo della malattia.

È chiaro, però, che tali vantaggi vengono pienamente sfruttati solo nel caso in cui il trattamento venga somministrato con estrema accuratezza, ovvero garantendo che, in ogni istante, il fascio radiante incida sul volume bersaglio. Se ciò non dovesse avvenire, il trattamento causerebbe gravi danni collaterali oltre a un inefficiente controllo della malattia.

Assicurare l'accuratezza richiesta risulta particolarmente critico nel caso di trattamenti a lesioni soggette ad *organ motion*, ovvero al movimento a cui sono sottoposti gli organi interni per effetto di processi fisiologici, quali la respirazione e il battito cardiaco.

I sistemi di *image guidance* sono stati sviluppati sia per la radioterapia (IGRT, Image Guided RadioTherapy) che per l'adroterapia (IGHT, Image Guided HadronTherapy) al fine di limitare le incertezze nella localizzazione del bersaglio, sia dovute al posizionamento del paziente che dovute all'*organ motion*. Tali sistemi consentono l'acquisizione di immagini del paziente direttamente in sala di trattamento ed immediatamente prima dell'irradiazione. In tal modo è possibile localizzare accuratamente il bersaglio prima di irradiarlo.

Le tecniche di *image guidance*, però, non sono in grado di compensare le inaccuranze introdotte da movimenti del bersaglio rapidi e ciclici (*organ motion* intra-frazione), come ad esempio quelli dovuti alla respirazione. Una possibile tecnica di compensazione di tali movimenti consiste, invece, nel *tracking* della lesione, ovvero nella localizzazione in tempo reale del *target* e nell'inseguimento di quest'ultimo controllando la direzione del fascio radiante. Al tal fine è necessario stimare la posizione interna del bersaglio ricorrendo a segnali surrogati esterni tramite un modello di corrispondenza interno/esterno. Non è infatti possibile rilevare direttamente la posizione del bersaglio con acquisizioni radiografiche molto frequenti o attraverso fluoroscopia, in quanto, così facendo, si somministrerebbe al paziente una dose indesiderata inaccettabile.

Ad oggi, il *tracking* è implementato a livello clinico solo nell'ambito della radioterapia con fotoni, per la quale è disponibile un sistema di trattamento robotizzato chiamato *CyberKnife*, nel quale un surrogato esterno, costituito dalla posizione di tre marcatori collocati sulla superficie corporea del paziente, viene utilizzato per stimare la posizione interna del bersaglio da irradiare (segnale interno). Al fine di costruire il modello di correlazione interno/esterno e di valutare l'accuratezza delle stime effettuate, il sistema acquisisce periodicamente delle immagini radiografiche stereoscopiche in cui viene osservata l'effettiva posizione della lesione. Nell'ambito dell'adroterapia non è disponibile, ad oggi, nessun sistema di *tracking*. Nel caso di trattamenti con particelle, infatti, tale sistema richiede una complessità maggiore in quanto la compensazione del movimento non può avvenire, come in radioterapia con fotoni, soltanto sul piano perpendicolare al fascio ma deve avvenire anche in direzione longitudinale, tramite la modulazione della sua energia. Una volta implementati con successo, i sistemi di *tracking* permetterebbero, anche in adroterapia, un efficace trattamento delle lesioni mobili.

Tra i centri che potrebbero usufruire di un tale sistema vi è il CNAO (Centro Nazionale di Adroterapia Oncologica) di Pavia, dove, a partire da 2012, verranno effettuati trattamenti di adroterapia sia con fasci di protoni che di ioni carbonio.

Di conseguenza alle necessità cliniche appena esposte, il progetto proposto nel presente elaborato consiste nello sviluppo di metodi per il *tracking* in adroterapia. In particolare ci si è proposti di realizzare un prototipo che implementi un algoritmo in grado di apprendere la correlazione tra la posizione interna della lesione e la posizione esterna di marcatori localizzati da un sistema ottico. La correlazione appresa deve essere poi utilizzata per la stima *online* del segnale interno in funzione del segnale esterno. In aggiunta, lo strumento proposto deve anche essere in grado di aggiornare la correlazione qualora l'errore di stima, rilevato in appositi punti di controllo, risulti superiore ad una data soglia.

Si è deciso di affrontare il problema con un approccio a scatola nera, in quanto non vi sono modelli fisici in grado di interpretare, in modo accurato e in tempo reale, la correlazione tra il movimento esterno della superficie corporea e quello interno della lesione da irradiare. Si è quindi considerato l'impiego di due strumenti computazionali in grado di affrontare in modo efficace i problemi di tipo *black box*: le reti neurali artificiali e i modelli a variabili di stato.

Per quanto riguarda i dati utilizzati per lo sviluppo e la validazione del progetto, questi sono stati estratti da un database relativo ai trattamenti eseguiti tramite *CyberKnife* presso il Georgetown University Medical Center di Washington, DC.

La Figura 1.2 mostra l'iter seguito nello sviluppo del progetto proposto.

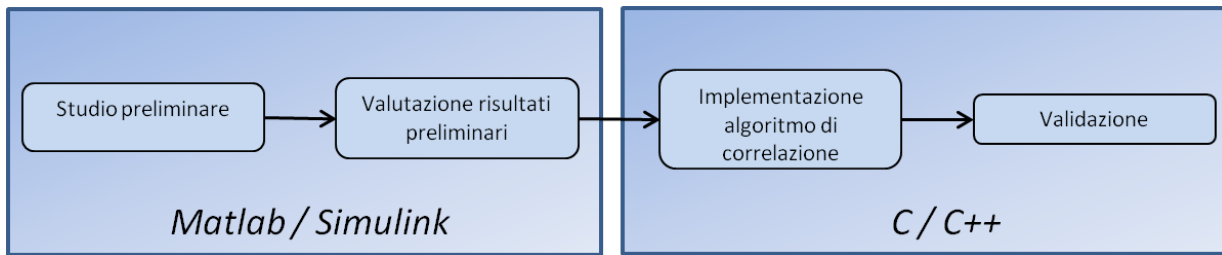


Figura 1.2: Schema a blocchi dell'iter seguito per lo svolgimento del progetto.

La prima fase del progetto proposto consiste in uno studio preliminare svolto in ambiente Matlab/Simulink, allo scopo di stabilire quale strumento computazionale, tra i due proposti, fosse il più adeguato per l'apprendimento della correlazione interno/esterno. Nell'ambito di questo studio preparatorio, il segnale esterno considerato consiste nella posizione spaziale di un singolo marcatore, mentre il segnale interno che si vuole stimare consiste nella stima della posizione interna della lesione eseguita dall'apposito modulo del sistema *CyberKnife*, chiamato *modeler*.

È stato realizzato un unico modello in Simulink, comprendente sia una rete neurale a doppio strato sia un modello a variabili di stato. L'apprendimento della correlazione interno/esterno è stato effettuato, per quanto riguarda le reti neurali, tramite una procedura di addestramento con il metodo di Levenberg-Marquardt, mentre, per il modello di stato, tramite la stima dei suoi parametri attraverso il metodo dei minimi quadrati. In entrambi i casi l'apprendimento è avvenuto considerando 500 esempi (coppie ingresso/uscita) in cui, ad ogni posizione del marcatore esterno registrata, viene associata la stima effettuata dal *modeler* nello stesso istante temporale.

Per condurre tale studio, sono state considerate otto frazioni di trattamento contenute nel database precedentemente citato, estraendo da ognuna 5000 campioni (corrispondenti a circa 200 secondi) del segnale esterno (posizione di un marcatore) ed i corrispondenti campioni del segnale interno (stima eseguita dal *modeler*). Si è proceduto quindi all'addestramento della rete e al calcolo dei parametri del modello di stato utilizzando le prime 500 coppie delle 5000 considerate. Successivamente, si è stimato il segnale interno in funzione del segnale esterno in corrispondenza di ogni campione.

La valutazione della performance dei due strumenti è stata effettuata calcolando il segnale differenza, campione per campione, tra la stima eseguita dallo strumento considerato rispetto a quella eseguita dal *modeler*. I risultati, relativi agli otto casi considerati, sono mostrati in Tabella 1.1.

Media del segnale differenza [mm]			
Strumento/Coordinata	X	Y	Z
Rete neurale	0,4885	0,4166	0,3455
Modello di stato	0,9089	0,6993	0,4281

Tabella 1.1: Valori medi del segnale differenza ottenuti per la rete neurale e per il modello di stato.

Tali risultati mostrano che, relativamente ai casi considerati, la rete neurale è in grado di apprendere la correlazione interno/esterno più efficacemente del modello di stato.

Si è quindi proceduto alla realizzazione di un algoritmo di correlazione basato su reti neurali, operando in ambiente C/C++.

Anche per lo sviluppo di tale algoritmo sono stati utilizzati i dati relativi ai trattamenti effettuati con *CyberKnife*: in particolare si è considerato un segnale esterno costituito dalla posizione dei tre marcatori posti sulla superficie corporea del paziente, mentre come segnale interno, è stata considerata la posizione del *target* rilevata dalle immagini radiografiche che il sistema *CyberKnife* acquisisce prima dell'irradiazione (per costruire il modello di correlazione) e durante la stessa (per verificare l'accuratezza della stima).

La principale funzionalità dello strumento sviluppato consiste nel costruire ad addestrare tre distinte reti neurali, ognuna delle quali stima la posizione del *target* in funzione della posizione di un marcatore esterno. L'architettura delle reti, a discrezione dell'utente, può comprendere o meno uno strato nascosto formato da un numero arbitrario di neuroni. Lo strato d'uscita, inoltre, può avere una funzione d'attivazione lineare o non lineare (tangente iperbolica).

L'addestramento iniziale delle reti avviene utilizzando 500 esempi ricavati dall'interpolazione delle coppie ingresso/uscita rese disponibili dalle immagini che il sistema *CyberKnife* utilizza per la costruzione del modello di correlazione.

Terminato l'addestramento iniziale, le tre reti stimano la posizione del *target* in corrispondenza di ogni campione del segnale esterno registrato nel trattamento in analisi e le loro uscite vengono mediate per ottenere un'unica stima.

Inoltre, in corrispondenza delle immagini acquisite come controllo, l'algoritmo verifica l'accuratezza delle stime effettuata e calcola lo scostamento tra la posizione stimata del bersaglio e quella effettiva. Se tale differenza risulta superiore ad una soglia imposta dall'utente, le reti neurali vengono riaddestrate aggiornando il *training set* con logica *first in first out*, ovvero escludendo la coppia meno recente ed inserendo quella ottenuta dall'ultima immagine acquisita.

Tutte le stime effettuate vengono quindi registrate in un file formato testo, in modo da poter essere facilmente esportate ed analizzate in ambiente Matlab.

L'accuratezza della predizione effettuata dal modello implementato è stata confrontata con quella del *modeler*, attraverso il calcolo della differenza tridimensionale tra la posizione stimata e quella effettiva del bersaglio, in corrispondenza delle acquisizioni delle immagini di controllo. Per tale confronto sono state considerate 20 frazioni di trattamento estratte dal database raccolto presso il Georgetown University Medical Center, divise in due *subset*: uno contenente i casi in cui il *modeler* mostra il maggiore errore di puntamento medio (*worst*) e uno, chiamato *control*, contenente casi selezionati casualmente nel database, ovviamente ad esclusione di casi *worst*.

I valori medi complessivi dei *targeting error (TE)*, ovvero la distanza 3D tra la posizione stimata del bersaglio e la sua posizione effettiva, calcolati per il *modeler* e per le reti neurali sono riportati nella Tabella 1.2.

<i>Subset</i>	Media TE <i>modeler</i> [mm]	Media TE reti neurali [mm]
<i>control</i>	1,6675	1,3870
<i>worst</i>	6,6086	4,8084

Tabella 1.2: Valori medi del targeting error per modeler e per le reti neurali.

Tali risultati mostrano che, relativamente ai 20 casi considerati e limitatamente ai punti in cui si acquisiscono le immagini di controllo, il modello sviluppato è in grado di fornire stime della posizione del *target* con accuratezza maggiore rispetto al *modeler*. La differenza tra le distribuzioni degli errori di puntamento è risultata statisticamente significativa (test di Wilcoxon) sia per i casi *control* ($p = 2,1066 \times 10^{-5} < 0,01$) che per i casi *worst* ($p = 0,0012 < 0,01$).

Al fine di poter integrare il prototipo all'interno di sistemi commerciali, come il sistema di localizzazione ottica SMART-D (BTS S.p.a.) installato presso CNAO, esso è stato modificato in modo da poter operare *online*. In tal modo, i dati non vengono importati da un database ma vengono forniti in ingresso all'algoritmo nello stesso istante in cui vengono acquisiti.

È stato possibile testare il funzionamento *online* grazie ad un fantoccio per la simulazione del movimento respiratorio, sul quale si è deciso di posizionare quattro marcatori, dei quali tre simulano il segnale esterno, mentre il quarto, sorretto tramite elementi elastici a rigidità variabile, simula il movimento del bersaglio da inseguire. Tale fantoccio viene azionato

manualmente e, grazie ad una puleggia a sezione variabile appositamente progettata e realizzata, consente di eseguire movimenti respiratori variabili sia in frequenza che in ampiezza.

Per porsi in condizioni sperimentali analoghe a quelle di un trattamento tramite *CyberKnife*, la posizione dei primi tre marcatori viene letta in tempo reale, mentre la posizione del quarto viene resa disponibile all'algoritmo ad una frequenza imposta dell'utente, simulando di fatto l'acquisizione delle immagini radiografiche.

In modalità *online* il prototipo opera analogamente a come è stato descritto precedentemente, quindi esegue un addestramento iniziale delle reti neurali e successivamente stima, *frame* per *frame*, la posizione del bersaglio (quarto marcatore) in funzione della posizione dei primi tre. In concomitanza con l'acquisizione della posizione del quarto marcatore, viene valutata l'accuratezza della stima e, se necessario, le reti vengono riaddestrate.

I test eseguiti operando *online* mostrano che, acquisendo ad una frequenza di 60 Hz, il prototipo è in grado di effettuare la stima del segnale interno in tempo reale, dunque è possibile disporre di tale stima in corrispondenza di ogni *frame* acquisito, tranne nel caso in cui l'algoritmo stia eseguendo il riaddestramento delle reti, procedura che, in media, comporta la sospensione della stima per circa 0,57 secondi.

In conclusione, i risultati ottenuti sia in termini di accuratezza che in relazione alla possibilità di operare in tempo reale e di integrare il prototipo all'interno di sistemi commerciali, mostrano che le reti neurali rappresentano uno strumento potenzialmente utile nell'ambito dello sviluppo di sistemi di *tracking* per l'adroterapia. Seppure il prototipo realizzato, allo stato di sviluppo attuale, necessita sicuramente di un processo di ottimizzazione riguardante sia l'architettura delle reti neurali sia le modalità con le quali queste vengono addestrate, esso potrebbe essere un valido ausilio per la realizzazione di un sistema di *tracking*.

1. Summary

External beam radiotherapy aims at treating tumors by means of high energy electromagnetic waves, such as X-Rays. The energy transferred by the incident wave to irradiated tissues damages the cellular genetic material, preventing reproduction of irradiated cells. Since not only the tumor tissue but also healthy tissues are damaged by radiation, the fundamental principle of radiation therapy is to cause the highest possible damage to the target while, at the same time, sparing surrounding healthy structures.

For this purpose, radiotherapy treatments need to be carefully planned as a function of detailed information about target position and its volumetric extension.

A recent technique for radiotherapy, known as hadrontherapy, consist in irradiation of tumor lesions through accelerated charged particles, such as protons and carbon ions. The interaction between charged particles and matter is characterized, in terms of dose deposition, by the so-called “Bragg curve (or peak)” [Figure 1.1].

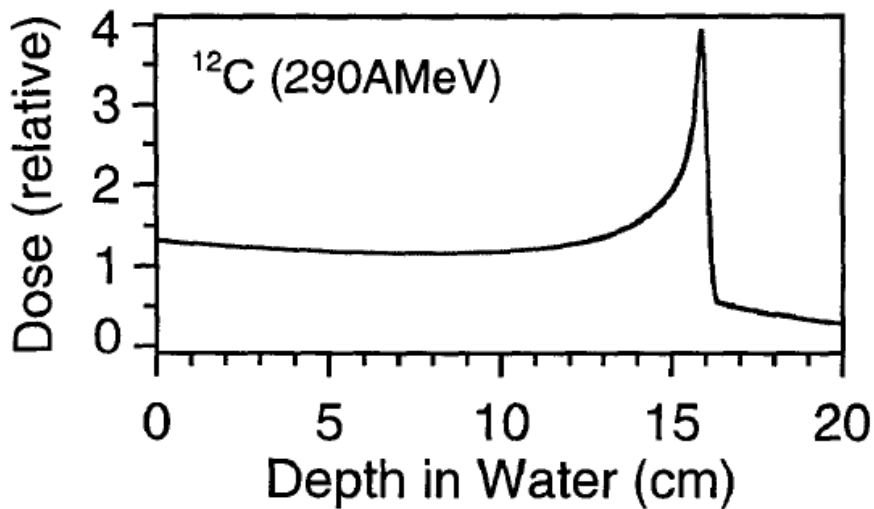


Figure 1.1: Bragg curve for carbon ions accelerated at 290 MeV/u.

This curve shows dose deposition as a function of penetration depth and has an initial plateau where the delivered dose is extremely reduced, followed by a sharp peak placed immediately before the depth where particles stop. Beyond this depth, dose delivery is limited and mainly due to light fragments traveling beyond. Moreover, the depth of the Bragg peak can be selected by adjusting the beam energy, in order to match maximal dose deposition within the target.

Compared with photon radiotherapy, hadrontherapy shows two main advantages. The first, which is a direct consequence of dose deposition properties, is the ability to deliver a high dose even to deep seated tumors while decreasing, at the same time, the damages to healthy tissues

placed proximal and distal to the target. The main consequence of this advantage is the possibility to accurately conform dose deposition to the target volume, thus providing highly selective treatments.

The second advantage is the increased relative biological effectiveness (RBE) compared with photons beams. This implies that a given physical dose delivered by particles (especially by carbon ions) causes increased biological damage to irradiated tissues with respect to photons, allowing for a more efficient tumor control.

However, these advantages can be fully exploited only if the treatment is delivered with very high accuracy, i.e. by ensuring that the beam hits the target at any time during irradiation. If accuracy is not guaranteed, the delivered treatment may cause severe collateral damage and an inefficient tumor control.

It may be very difficult to achieve the desired accuracy when treating lesions affected by organ motion, that is the movement due to physiological processes such as respiration and heart beat. Image guidance systems for radiotherapy (IGRT, Image Guided RadioTherapy) and hadrontherapy (IGHT, Image Guided HadronTherapy) were developed to decrease target localization uncertainties both due to patient positioning errors and organ motion. These systems allow for the acquisition of anatomical images directly in treatment room and immediately before irradiation. In this way the target can be accurately localized before being irradiated.

However, image guidance techniques are not able to reduce inaccuracies due to fast and cyclic target movements, such as those caused by respiration. A possible strategy for the compensation of these movements is to track the target by steering the beam. For this purpose, it is necessary to estimate the internal position of the target using external surrogate signals through an internal/external correspondence model. In fact, it is not possible to directly detect target position with very frequent radiographic acquisitions or through fluoroscopy, as this will result in an unacceptable over-dosage to the patient.

Currently this kind of solution has achieved successful clinical application only in photon radiotherapy, where the robotic treatment device called CyberKnife is available. This system uses an external surrogate signal, that is the position of three markers placed on the body surface, to estimate internal position of the target (internal signal). The system periodically acquires stereoscopic X-ray images, where the tumor is localized, in order to build an internal/external correspondence model and to evaluate the prediction accuracy.

On the other hand, no tumor tracking system is clinically available for hadrontherapy treatments. In the case of treatments with particles beams, in fact, such a system requires

greater complexity because motion compensation has not to be limited, as in photon radiotherapy, only to the transversal plane but it must also take into account movements in the longitudinal direction through beam energy modulation. When such a tracking system will be available, it will be a valuable tool for treating moving targets.

Among the centers that may profit from the use of such a system it is worth mentioning the Italian national center for hadrontherapy (CNAO, Centro Nazionale di Adroterapia Oncologica, Pavia, Italy), where clinical activity will start in 2012.

As a consequence, the aim of the proposed project is to develop methods for tumor tracking in hadrontherapy treatment. Specifically we aimed at the realization of a prototype able to learn the correlation between external surface markers and internal tumor trajectory. Such correlation can subsequently be used to estimate target position as a function of the external signal. Moreover, the proposed prototype has also to be able to update the correlation model when the prediction error, measured periodically through control images, is above a user defined threshold.

We opted for a black box approach, since there are no physical models able to explain the correlation between the body surface movement and internal target motion so that it is both accurate and real-time. Thus we considered two different computational methods, both able to solve a black box problem: artificial neural networks (ANN) and state variables models.

Data used for project development and validation are extracted from a database containing detailed and quantitative information about treatment delivered with the CyberKnife system at the Georgetown University Medical Center di Washington, DC.

Figure 1.2 shows the procedure followed for the development of the proposed project.

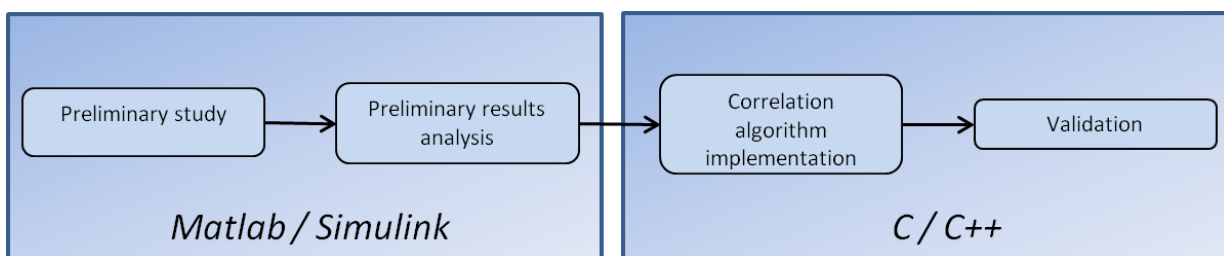


Figure 1.2: Block diagram showing the procedure followed to develop the project.

The first phase of the proposed project is a preliminary study in Matlab/Simulink environment, aimed at identifying the most performing computational method in learning the existing internal/external correlation. For this preparatory study, we selected as external signal the position of a single marker, while the internal signal is the target position as predicted by the dedicated CyberKnife module, called “modeler”.

A single Simulink model has been built to include a two-layers network and a state model. For the ANN model, the learning procedure for internal/external correlation was implemented through the Levenberg-Marquart method, whereas, for the state model, the identification of its parameters was performed by means of least-square minimization. In both cases the learning is done by considering 500 examples (input /output pairs) where each registered position of the external marker is associated with the contemporary prediction made by the modeler.

In order to conduct the preliminary study, eight fractions from the above mentioned database were considered, extracting from each of these fractions 5000 samples (corresponding to about 200 seconds) of the external signal (position of a marker) and the corresponding samples of the internal signal (estimation calculated by the modeler). Then, the ANN is trained and the state model parameters are calculated using the first 500 pairs from the 5000 considered. Next, for each sample, internal signal was estimated as a function of the external signal.

The performance assessment of the two methods was done by calculating, sample by sample, the difference between the prediction made by ANN /state model and the one performed by the modeler. The results for the eight considered cases, are shown in Table 1.1.

Method/Coordinate	Average difference [mm]		
	X	Y	Z
Neural Network	0,4885	0,4166	0,3455
State Model	0,9089	0,6993	0,4281

Table 1.1: Average values of the difference signal obtained for the neural network and the state model.

These results show that, in relation to the considered cases, the ANN method is able to learn the internal/external correlation more effectively than the state model.

We therefore designed a correlation algorithm based on neural networks, working in C / C ++ environment.

Data on CyberKnife treatments were used also for developing this algorithm: in particular, it was considered an external signal consisting of the position of three markers placed on the patient's body surface, while the internal signal was given by the position of the target detected by those X-ray images that the system acquires before irradiation (to build the model correlation model) and during treatment (to verify prediction the accuracy).

The main feature of the developed prototype is the ability to build and train three distinct neural networks, each one estimating the target coordinates according to the position of a single

external marker. The ANNs architecture, at user’s discretion, may or may not include a hidden layer consisting of any number of neurons. The output layer may have a linear or nonlinear (hyperbolic tangent) activation function.

Initial training is done considering 500 examples obtained by interpolation of the input /output pairs made available by those images that CyberKnife acquires for the construction of the correspondence model. After initial training, the three networks predict target position for each sample of the external signal recorded in the considered treatment and their outputs are averaged to have a single prediction.

In addition, in correspondence to the acquired control images, the algorithm checks the accuracy of its predictions and calculates the deviation between the predicted target position and the actual one. If the difference is above a threshold set by the user, neural networks are retrained by updating the training set with a *first in first out* logic, that is excluding the older pair and inserting that obtained with the last captured image. All the calculated estimates are recorded in a text format file, to be easily exported and analyzed in Matlab.

The accuracy of predictions calculated by the implemented model was compared to the accuracy of the modeler through the calculation of the three-dimensional difference between predicted and actual target position, measured in correspondence with every control image. For this comparison, 20 fractions, extracted from the database collected at the Georgetown University Medical Center, were considered. These fractions were divided into two subsets: one containing those cases where the modeler shows the largest average targeting error (“worst”) and one, named “control”, containing randomly selected cases where worst ones were obviously excluded. The overall average values of targeting error (TE), i.e. the 3D distance between the estimated target position and its actual position, calculated for the modeler and neural networks are shown in Table 1.2.

<i>Subset</i>	<i>Average modeler TE [mm]</i>	<i>Average neural networks TE [mm]</i>
<i>control</i>	1,6675	1,3870
<i>worst</i>	6,6086	4,8084

Table 1.2: Average targeting error calculated in each subset for the modeler and the neural networks.

These results show that, relatively to the 20 considered cases and to the points where control images are acquired, the developed model is able to provide predictions of the target position

with greater accuracy than the modeler. The difference between distributions of targeting errors was statistically significant (Wilcoxon test) for both control ($p = 2.1066 \times 10^{-5} < 0.01$) and worst cases ($p = 0.0012 < 0, 01$).

In order to integrate the prototype into a commercial system, such as the optical tracking system SMART-D (BTS Spa) installed at CNAO, it was modified to be able to operate online. In this way, data are not imported from a database but are provided to the algorithm at the same time they are acquired.

It was possible to test the online operation of the prototype thanks to a phantom for the simulation of respiratory motion, on which we decided to place four markers: three of those simulate the external signal, while the fourth, supported by elastic elements with variable stiffness, simulates the movement of the target to be tracked.

This phantom is operated manually and, thanks to a specifically designed and manufactured variable section pulley, is able to perform respiratory movements that are adjustable in frequency and amplitude.

In order to simulate a treatment delivered by CyberKnife, the algorithm reads the position of the first three markers in real time, while the position of the fourth (the target) is made available at a frequency set by the user, simulating X-ray images acquisition. In online mode the prototype works as described above, that is performing an initial training and then calculating, frame by frame, the position of the target (the fourth marker) as a function of the position of the first three markers. In conjunction with the acquisition of the fourth marker, prediction accuracy is evaluated and, if necessary, networks are retrained.

Tests about online operations, show that, acquiring at a frequency of 60 Hz, the prototype is capable of calculating internal signal in real time; hence it is possible to have such a prediction for each acquired frame, unless the algorithm is retraining one or more networks. This procedure, on average, results in unavailable prediction samples for about 0,57 seconds.

In conclusion, the results in terms of accuracy, real time operational capabilities and integration into a commercial system, show that neural networks represent a potentially useful method for the development of tumor tracking in hadron therapy. Although the prototype at its current development stage certainly requires an optimization process involving both the architecture of neural networks and the procedure used for training, it could be a useful tool for the implementation of a tracking system in hadrontherapy treatments.

2. Introduzione

In questo capitolo si vogliono introdurre i concetti fondamentali inerenti alla radioterapia convenzionale e all'adroterapia. Verranno presentati i principi fisici e gli effetti biologici di entrambe, impostando un confronto tra le possibili tecniche utilizzate.

Nel paragrafo successivo verrà esposta una panoramica introduttiva riguardo alla radioterapia e ai suoi principi fondamentali. Si ritiene utile fornire tale panoramica dal momento che l'adroterapia condivide la maggior parte dei principi e delle problematiche caratteristiche della radioterapia a fasci esterni.

2.1. Introduzione alla radioterapia a fasci esterni

Per radioterapia si intende il trattamento delle neoplasie mediante l'utilizzo di radiazioni ionizzanti ad elevata energia [1,2]. Le radiazioni ionizzanti, infatti, sono in grado di depositare all'interno dei tessuti un'energia sufficiente a distruggere o danneggiare il materiale genetico contenuto nelle cellule, rendendo quest'ultime incapaci di crescere e di riprodursi.

Il danneggiamento del materiale genetico può essere diretto o indiretto: nel primo caso le radiazioni determinano ionizzazioni direttamente a livello delle molecole di DNA, alterandone la struttura chimica e conseguentemente la funzionalità. Nel secondo caso, invece, la radiazione ionizza le molecole d'acqua, causando la formazione di radicali liberi idrossilici. Tali composti chimici aggressivi interagiscono con il DNA, danneggiandolo.

Le finalità che un trattamento radioterapico si pone possono essere di diversa natura. A questo proposito si può distinguere tra trattamenti radicali, palliativi o profilattici. Nel caso di trattamenti radicali la terapia ha lo scopo di permettere al paziente la guarigione completa tramite la distruzione di tutte le cellule neoplastiche. Nel caso di trattamenti palliativi lo scopo è invece quello di migliorare la qualità della vita del paziente colpito da tumore, cercando di ridurre i sintomi della malattia quali dolore e compressione. Nell'ultimo caso, invece, si irradiano dei volumi che non presentano più malattia neoplastica macroscopica, perché è già stato effettuato un trattamento chirurgico o chemioterapico, ma nei quali si vuole prevenire la futura crescita di cellule tumorali recidive o di metastasi.

A seconda della posizione della sorgente irraggiante rispetto al paziente, si può distinguere un trattamento in cui la sorgente radiante è posta all'interno del paziente, o brachiterapia, dalla radioterapia a fasci esterni. In questo caso, il fascio di radiazioni è prodotto da una apparecchiatura

specifica, come un acceleratore lineare, e viene indirizzato selettivamente sul bersaglio da trattare. Di seguito ci si riferirà esclusivamente a quest'ultimo caso.

Come già detto, la radioterapia prevede l'irraggiamento di un volume neoplastico, il *target*, tramite radiazione ionizzante, ovvero onde elettromagnetiche ad alta frequenza, alle quali, come è noto, è associata un'elevata energia.

Le radiazioni più comunemente utilizzate sono i raggi X, la cui frequenza è compresa tra 10^{16} e 10^{21} Hz [Figura 2.1]. Tali radiazioni sono prodotte da transizioni energetiche dovute ad elettroni in rapido movimento e vengono generati da macchine chiamate acceleratori lineari.

Secondariamente vengono impiegati anche raggi γ , la cui frequenza è compresa tra 10^{18} e 10^{23} Hz [Figura 2.1]. A differenza dei raggi X, quest'ultimi sono prodotti dal decadimento di isotopi radioattivi di elementi come l'Iridio o il Cobalto 60.

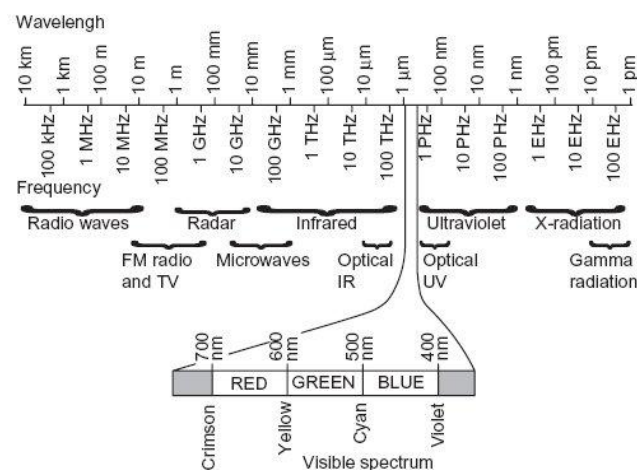


Figura 2.1: Spettro elettromagnetico.

Dal momento che la radiazione ionizzante è in grado di danneggiare, oltre alle cellule tumorali, anche quelle sane, causando effetti collaterali che possono comprendere anche l'insorgenza di lesioni tumorali nei tessuti sani irradiati, il principio fondamentale della radioterapia consiste nel causare il maggior danno possibile alla lesione, risparmiando, nel contempo, i tessuti sani circostanti. Di conseguenza, il trattamento radioterapico deve essere pianificato ed eseguito con metodologie e tecnologie volte a garantire che tale principio sia rispettato. Ciò si traduce, in sostanza, nel garantire che il fascio radiante depositi la dose richiesta (definita come l'energia assorbita per unità di massa, misurata in Gray [J/Kg]) sul *target* con la maggior accuratezza possibile. Eventuali inaccurately si traducono, infatti, in un sovradosaggio dei tessuti sani (che comporta in questi ultimi effetti collaterali indesiderati) e, contemporaneamente, in un sottodosaggio del bersaglio, con conseguente riduzione di efficacia del trattamento. L'evoluzione delle metodologie e delle tecnologie impiegate in radioterapia è stata ed è tutt'ora

indirizzata verso la massimizzazione dell'accuratezza con la quale il trattamento viene somministrato.

Trattamenti convenzionali

I trattamenti convenzionali prevedono l'individuazione del target tramite immagini CT, l'esecuzione di un piano di trattamento e quindi la somministrazione della dose nella zona del bersaglio. L'iter di un trattamento radioterapico, generalmente valido sia per i trattamenti convenzionali sia per quelli non convenzionali, può essere riassunto mediante il seguente schema a blocchi [Figura 2.2].

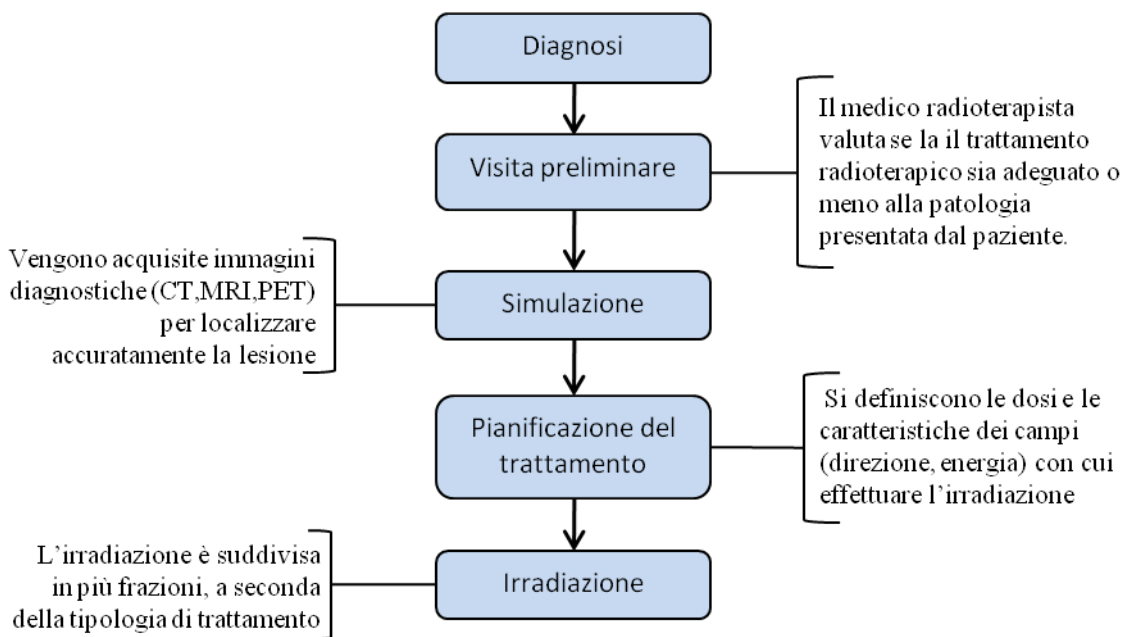


Figura 2.2: Schema a blocchi dell'iter di un generico trattamento di radioterapia.

Tali trattamenti, ancora largamente praticati per la loro affidabilità e semplicità di esecuzione, presentano lo svantaggio di una precisione più limitata e, spesso non permettono la somministrazione di dosi elevate nei trattamenti che le richiederebbero per garantire un controllo ottimale della malattia. Tali dosi, infatti, sarebbero eccessivamente tossiche per gli organi sani colpiti dal fascio radiante. È il caso, per citare un esempio molto comune, del tumore alla prostata, che non può essere colpito con dosi elevate per via della prossimità con organi particolarmente sensibili alle radiazioni quali la vescica e il retto.

In questo tipo di trattamento, infatti, il fascio radiante ha una sezione rettangolare o circolare mentre le lesioni sono spesso caratterizzate da forme irregolari: ciò implica che i raggi, seppure centrati sul bersaglio, necessariamente vadano a colpire anche organi sani circostanti.

Oltre a ciò, la mobilità degli organi fa sì che le lesioni (soprattutto nel distretto toraco-addominale) si spostino di diversi millimetri dal punto in cui erano state individuate in fase di

pianificazione, sia tra una seduta e la successiva sia durante la seduta stessa: questo implica che il fascio radiante non sia sempre perfettamente centrato sul bersaglio [3].

L'incertezza della posizione del *target*, inoltre, costringe ad accettare margini di sicurezza di diversi millimetri, grazie ai quali si può essere certi di colpire il bersaglio con il fascio radiante. Tali margini vengono sommati al volume visibile del bersaglio (*Gross Target Volume*, GTV) sia per considerare eventuali parti non visibili della lesione tumorale (*Clinical Target Volume*, CTV), sia per tener conto di eventuali incertezze di localizzazione del bersaglio (*Planning Target Volume*, PTV) [Figura 2.3].

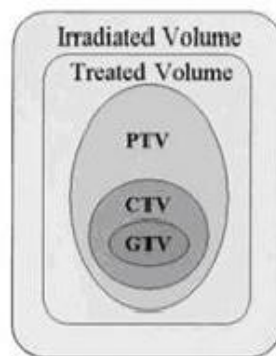


Figura 2.3: Definizioni ICRU 50 di GTV, CTV, PTV.

Lo svantaggio associato a tali margini è però riscontrabile nella necessità di irradiare zone più ampie di quanto sarebbe necessario per colpire la lesione, se la sua posizione fosse precisamente nota. Tutto ciò si traduce in trattamenti che somministrano dosi limitate e colpiscono sistematicamente i tessuti sani. Questa schematizzazione è applicabile anche ai trattamenti che utilizzano adroni, con alcune evidenti differenze che verranno espone nei paragrafi successivi.

2.2. Introduzione all'Adroterapia

2.2.1. Cenni storici

L'adroterapia è un tipo di trattamento in cui il fascio radiante è costituito da adroni, ovvero particelle formate da quark. Rientrano quindi in questa categoria le irradiazioni con protoni, pioni, ioni (che sono aggregati di nucleoni, a loro volta adroni) e neutroni [4].

Le potenzialità mostrate dalle particelle nel superare i limiti di carattere fisico e radiobiologico della radioterapia convenzionale furono evidenziate in un lavoro di Robert Wilson del 1946 [5], che studiò il profilo di dose in funzione della profondità per i protoni prodotti dal ciclotrone di Berkeley ed osservò un rapido aumento della deposizione di energia al termine della *range*

della particella. Tale fenomeno era già stato osservato nel 1903 da William Bragg per le particelle α [6] ed è conosciuto, appunto, come profilo (o picco) di Bragg.

Alla luce di tali lavori, risultò evidente che l'aumento della deposizione di energia al termine della traccia della particella permetteva di trasportare una dose elevata non solo ai tumori superficiali, bensì anche a quelli posti a profondità maggiori. Inoltre, il ridotto *scattering* laterale delle particelle pesanti e il loro *range* limitato consentiva un notevole vantaggio in termini di selettività della deposizione di dose.

In aggiunta a tali fattori strettamente legati alla fisica della particelle, si notò che l'aumento della densità di ionizzazioni osservata al termine del *range* della particella, implicava, per gli ioni pesanti più che per i protoni, un incremento nell'efficacia radiobiologica della radiazione [7].

I principi fisici e le considerazioni radiobiologiche qui soltanto accennate a scopo introduttivo, verranno riprese ed approfondite nei paragrafi 2.5.1 e 2.5.2, mentre di seguito verranno tracciate le principali tappe dell'evoluzione dell'adroterapia [4,8].

I primi trattamenti radioterapici furono effettuati nel 1938-1939, utilizzando neutroni veloci prodotti dal ciclotrone presente presso il Lawrence Berkeley Laboratory (LBL, USA). A seguito del lavoro di Wilson, nel 1954, sempre all'LBL, fu utilizzato per la prima volta un fascio di protoni per il trattamento di tumori.

L'adroterapia, ancor più della radioterapia convenzionale, necessita di una precisa localizzazione del volume bersaglio, cosa che difficilmente era possibile disponendo soltanto delle radiografie a raggi X come unico strumento per ottenere immagini volte alla diagnosi e alla pianificazione dell'irradiazione. Per questo motivo, soltanto i tumori collocati in zone facilmente identificabili nelle radiografie erano trattabili tramite adroterapia.

Fu soltanto negli anni '70 che, con l'avvento di tecniche di *imaging* volumetriche e a risoluzione millimetrica come la CT, si ebbe la possibilità di eseguire trattamenti adroterapici su un maggior numero di tumori. Inizialmente, tali trattamenti venivano effettuati in impianti, come l'LBL, dedicati principalmente alla ricerca nucleare e, solo in parte, all'adroterapia oncologica.

Il primo centro dedicato esclusivamente all'adroterapia fu Loma Linda (California, USA), dove si effettuano trattamenti con protoni. A Chiba, in Giappone, fu completata nel 1984 la prima unità di trattamento con ioni pesanti e, a tutt'oggi, il centro HIMAC (*Heavy Ion Medical Accelerator in Chiba*) ha accumulato una notevole esperienza clinica nel trattamento con ioni carbonio. In Europa [9] spiccano i centri PSI (Villigen, Svizzera) e GSI (Darmstadt, Germania). In quest'ultimo centro, oltre ai protoni si utilizzano anche fasci di ioni carbonio.

Sempre in Europa, negli ultimi anni, si è assistito alla nascita di due moderni centri intermanti dedicati all'adroterapia con protoni e ioni carbonio: il primo, in Germania, chiamato HIT (*Heidelberg Ion-Beam Therapy Center*), che è già in funzione e rileva completamente l'attività clinica che si svolgeva presso il centro GSI, il quale, ad oggi, svolge soltanto attività di ricerca. Il secondo, chiamato CNAO (Centro Nazionale di Adroterapia Oncologica), a Pavia, in Italia.

2.2.2. Principi fisici

Scopo di questo paragrafo è di fornire una sintetica descrizione dei principi fisici che regolano il comportamento delle particelle e la loro interazione con la materia[10,11], trascurando, per il momento, gli effetti che tali particelle hanno sulla materia biologica. Successivamente si discuterà delle tecnologie utilizzate per la generazione del fascio e per la deposizione della dose[11].

Curva di Bragg

Come già anticipato, le particelle cariche sono caratterizzate da una curva di deposizione di energia in funzione della profondità della materia attraversata regolata dal "picco di Bragg". Per poter mostrare come questo andamento dipenda della fisica dell'interazione tra la particella e la materia attraversata, è necessario, per prima cosa, definire il trasferimento lineare di energia (LET, *Linear Energy Transfer*):

$$LET \triangleq - \frac{dE}{dx} \quad (2.1)$$

che rappresenta l'energia che la particella cede al mezzo attraversato per unità di lunghezza della traccia, ricordando che per traccia si intende la traiettoria della particella all'interno della materia attraversata. Si noti che il segno meno vuole indicare proprio che l'energia trasferita viene ceduta dalla particella alla materia. Il LET viene generalmente misurato in KeV/ μ m.

In Figura 2.4 viene mostrato l'andamento del LET in funzione dell'energia per diverse particelle, dal protone allo ione uranio.

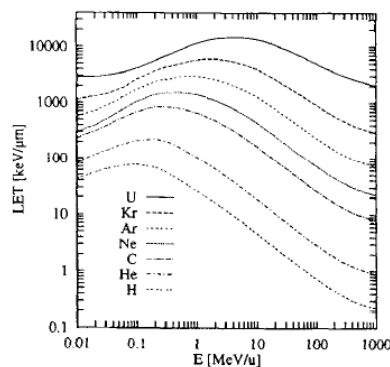


Figura 2.4: andamento del LET in funzione dell'energia per diverse particelle, dal protone allo ione uranio.

Nel campo delle energie utilizzate per i trattamenti adroterapici (diverse centinaia di MeV/u), il LET può essere calcolato tramite la formula di Bethe-Bloch, riportata di seguito:

$$\frac{dE}{dx} = \frac{4\pi e^4 Z_{eff}^2 N}{m_e v^2} \ln \frac{2mv^2}{I(1-\beta^2)} + \text{termini relativistici} \quad (2.2)$$

dove e ed m_e sono rispettivamente la carica e la massa dell'elettrone, Z è il numero atomico del bersaglio, N la densità elettronica del bersaglio e v la velocità della particella. Infine, Z_{eff} rappresenta la carica efficace della particella e può essere approssimata con la seguente relazione ricavata empiricamente da Barkas[12]:

$$Z_{eff} = Z(1 - e^{-125\beta z^{-2/3}}) \quad (2.3)$$

ricordando che $\beta = v/c$, con c velocità della luce nel vuoto.

L'andamento di Z_{eff} in funzione dell'energia è riportato in Figura 2.5.

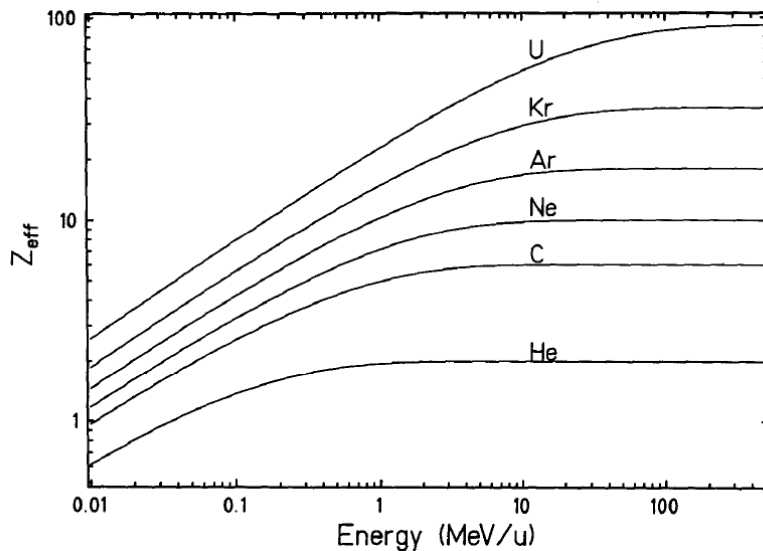


Figura 2.5: Carica efficace della particella in funzione della sua energia.

Volendo identificare i fattori che maggiormente influenzano il LET si può semplificare l'equazione 2.2 scrivendo:

$$LET \propto \frac{Z_{eff}^2}{v^2} \quad (2.4)$$

dove si evidenzia soltanto la dipendenza dal quadrato della carica efficace e dall'inverso del quadrato della velocità della particella.

Nel campo di energie utilizzate per i trattamenti adroterapici, la particella carica interagisce prevalentemente con gli elettroni della materia attraversata e il tempo di interazione influenza direttamente l'intensità dell'interazione stessa: maggiore è il tempo di interazione maggiore è la forza di interazione tra particella ed elettroni del bersaglio, e viceversa. La considerazione

appena esposta giustifica la dipendenza del LET dall'inverso del quadrato della velocità della particella, riscontrabile nell'equazione 2.4 .

Poiché l'energia della particella è proporzionale al quadrato della sua velocità, si può affermare che il LET sia inversamente proporzionale all'energia della particella:

$$LET \propto \frac{Z_{eff}^2}{E} \quad (2.5)$$

come si osserva in Figura 2.4, dove si evidenzia una diminuzione del LET all'aumentare dell'energia.

Per giustificare l'incremento del LET alle basse energie fino ad un valore massimo, è necessario osservare (in Figura 2.5) come varia la carica efficace del proiettile in funzione dell'energia. Si osserva che, alle alte energie, la carica efficace coincide con il numero atomico dell'elemento considerato, dal momento che non vi è nessun elettrone che si lega con lo ione. Al diminuire dell'energia, il proiettile tende a raccogliere elettroni dalla materia in cui si muove: ciò comporta una rapida diminuzione della carica efficace e, di conseguenza, del LET. Si voglia ora tracciare l'andamento della deposizione di energia in funzione della profondità della particella all'interno della materia attraversata [Figura 2.6]. Si ottiene il seguente andamento [13]:

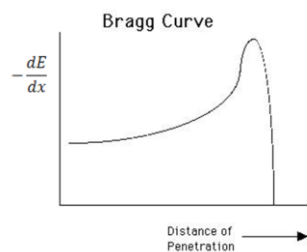


Figura 2.6: Andamento della deposizione di energia in funzione della profondità di penetrazione.

Un andamento analogo si ottiene per la deposizione di dose, sempre in funzione della profondità di penetrazione [Figura 2.7]:

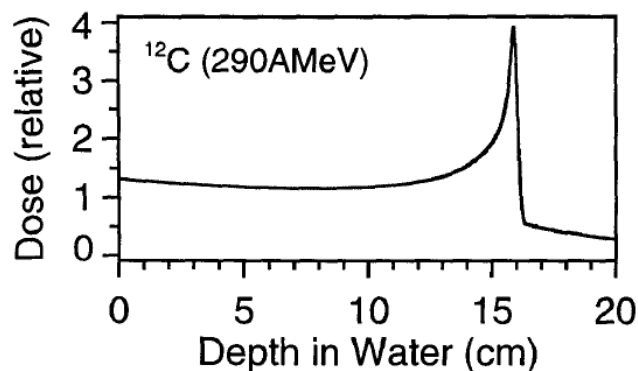


Figura 2.7: Dose in funzione della profondità in acqua per un fascio di ioni carbonio con energia iniziale di 290 MeV/u

Per commentare quest'ultimo andamento, noto come "curva o picco di Bragg", si ricorda che, come è ovvio, alle basse profondità di penetrazione sono associate alte energie delle particelle e che queste perdono progressivamente energia mentre attraversano la materia.

Osservando il grafico in Figura 2.7, si osserva un plateau iniziale seguito da un rapido incremento della dose depositata, a sua volta seguito da un altrettanto rapido decremento.

Il plateau iniziale si giustifica ricordando che alle alte energie le particelle mostrano uno scarso LET e, conseguentemente, depositano una dose ridotta. Al crescere della profondità di penetrazione (quindi al decrescere dell'energia della particella) si osserva un rapido aumento del LET e quindi della dose depositata. In ultimo, quando l'energia della particella diminuisce ulteriormente, si assiste alla rapida diminuzione del LET dovuta alla diminuzione della carica efficace dello ione, fenomeno che causa la diminuzione della dose depositata e che porta all'arresto della particella.

E' opportuno evidenziare che le misurazioni effettuate con camere di ionizzazione mostrano curve di Bragg con andamento molto più dolce di quelle ottenute teoricamente tramite la formula di Bethe-Bloch, con una distribuzione quasi gaussiana della deposizione di energia[14].

Scattering laterale

Fino ad ora si è considerato il comportamento della particella soltanto in direzione longitudinale, ovvero parallelamente alla sua direzione principale di propagazione all'interno del mezzo.

Altrettanto importante ai fini terapeutici è determinare il comportamento della particella in direzione trasversale, ovvero perpendicolarmente all'asse del fascio.

Idealmente una particella si propaga all'interno del mezzo lungo una linea retta e senza che la sua traiettoria subisca deviazioni. In realtà, a causa dell'interazione Coulombiana elastica tra le particelle e i nuclei della materia attraversata, la traccia non è perfettamente rettilinea ma subisce delle deviazioni. Tale fenomeno è detto *scattering* laterale ed è descritto dalla teoria di Moliere.

Come risulta intuitivo, l'angolo di *scattering* è inversamente proporzionale alla quantità di moto del proiettile [15], dunque, a parità di energia, le particelle più pesanti sono caratterizzate da una traccia più rettilinea. Inoltre, verso la fine della traccia, cioè quando la velocità (e quindi la quantità di moto) della particella diminuisce, lo *scattering* laterale diventa più significativo e la traccia più erratica.

È evidente che il fenomeno appena descritto causa un allargamento del fascio trasversalmente alla direzione di propagazione: ciò deve essere considerato quando si pianifica l'irradiazione di

un volume bersaglio, in quanto le particelle, deviando dalla direzione di propagazione ideale, potrebbero colpire e danneggiare tessuti circostanti al *target* e/o al canale di ingresso, ovvero tessuti che il fascio attraversa prima di raggiungere il bersaglio.

Quanto precedentemente detto riguardo allo *scattering* laterale viene mostrato in Figura 2.8 [16], dove l'allargamento trasversale del fascio è riportato in funzione della profondità di penetrazione.

Si noti, a conferma di quanto detto, che i protoni, avendo una massa minore degli ioni carbonio, sono soggetti a uno *scattering* laterale più accentuato.

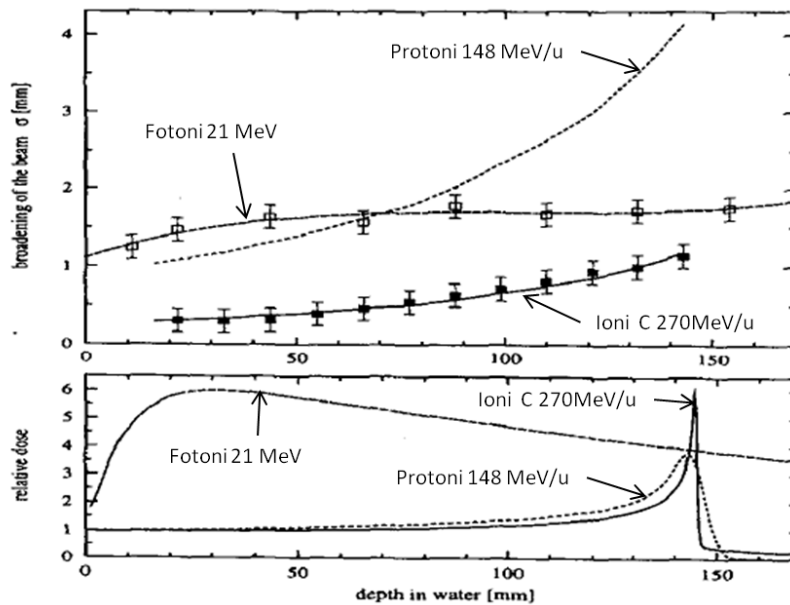


Figura 2.8: Pannello A: allargamento trasversale del fascio in funzione della profondità di penetrazione.

Pannello B: deposizione di dose in funzione della profondità di penetrazione. Vengono considerati sia fotoni, sia protoni, sia ioni carbonio.

Frammentazione nucleare

Un fenomeno tipico dell'interazione tra particelle e un materiale assorbente, come ad esempio il corpo umano, è la frammentazione nucleare, dovuta agli urti e alle reazioni nucleari che avvengono tra gli ioni del fascio e i nuclei del mezzo attraversato. Tale fenomeno produce particelle più leggere di quelle presenti nel fascio originario, la cui presenza non può essere trascurata nella pianificazione di un trattamento adroterapico.

Alle elevate energie che vengono utilizzate in ambito terapeutico, i fenomeni di frammentazione sono fortemente influenzati da fattori geometrici e le collisioni più probabili, e quindi più frequenti, sono quelle dette "periferiche" (in cui si ha solo parziale sovrapposizione tra particella incidente e nucleo colpito). Sono, invece, molto più rare le collisioni "centrali".

Non volendo approfondire i fenomeni e le reazioni nucleari che avvengono a seguito degli urti, si sottolinea soltanto che i frammenti prodotti, essendo più leggeri delle particelle originarie, mostrano un percorso nella materia leggermente diverso da queste ultime, sia longitudinalmente sia trasversalmente.

Per prima cosa si osserva come i frammenti leggeri siano caratterizzati da un *range* maggiore rispetto a quello delle particelle originarie, vale infatti la formula:

$$R_{fr} = R_{pr} \frac{Z_{pr}^2}{M_{pr}} \frac{M_{fr}}{Z_{fr}^2} \quad (2.6)$$

Dove R indica il *range* della particella, Z il numero atomico e M la massa, i pedici (pr) ed (fr) indicano rispettivamente il proiettile e il frammento.

L'aumento del *range* che si osserva nei frammenti leggeri comporta un'incertezza sul *range* complessivo del fascio radiante, che, non essendo formato da ioni di un unico elemento, non è univocamente determinato ma è influenzato dalla natura aleatoria del fenomeno della frammentazione nucleare. Tale incertezza sul *range* si traduce, come è evidente, in un'incertezza sulla profondità alla quale le particelle rilasciano la maggior dose relativa. Nelle parti distali delle curve di Bragg (oltre il picco), infatti, si osserva una "coda" nella deposizione di dose dovuta appunto al contributo dei frammenti leggeri [Figura 2.9]. Tale coda può essere così rilevante da depositare una dose distale intollerabile per un trattamento adroterapico, come avviene nel caso degli ioni neon. Per limitare la dose indesiderata rilasciata dalla coda del picco di Bragg, una soluzione, talvolta utilizzata a livello clinico, consiste nell'impiantare chirurgicamente una "calza" in tessuto gore-tex tra il bersaglio e i tessuti ad esso posteriori (dal punto di vista del fascio incidente) : in questo modo, i frammenti leggeri che generano la coda si arrestano nel gore-tex e non depositano dose nei tessuti sani.

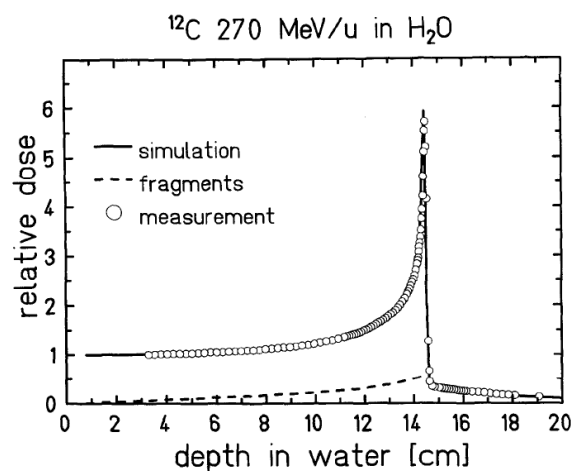


Figura 2.9: curva di Bragg per un fascio di ioni carbonio con energia di 270 MeV/u in acqua: è evidenziato il contributo di frammenti leggeri (linea tratteggiata).

Oltre all'incertezza introdotta in direzione longitudinale, la frammentazione nucleare comporta una maggiore rilevanza dello *scattering* laterale, in quanto, come già detto, tale fenomeno è tanto più rilevante quanto più è leggera la particella.

In linea generale, il fenomeno della frammentazione rende critico l'inserimento di materiali assorbitori sulla linea del fascio, spesso sfruttati per sagomare la sezione del fascio o modularne l'energia. L'inserimento di tali dispositivi, infatti, causa un peggioramento delle caratteristiche del fascio dovuto all'aumentare del numero di frammenti.

La frammentazione nucleare, però, non introduce soltanto svantaggi e problematiche nei trattamenti di adroterapia. Nel caso in cui venga utilizzato un fascio di ioni carbonio, tale fenomeno causa la rimozione di uno o due neutroni dalla particella, generando due isotopi emettitori di positroni, il ^{11}C e il ^{10}C [17]. Tali isotopi hanno un *range* e una velocità paragonabile a quella del ^{12}C .

L'emissione di due fotoni gamma coincidenti risultanti dall'annichilazione dei positroni emessi da tali isotopi consente la rilevazione *in-situ* (tramite uno scanner PET) dei punti, all'interno del copro, in cui le particelle si arrestano e rilasciano la massima dose relativa.

Risulta evidente come questa possibilità lasci spazio alla realizzazione di sistemi in grado di monitorare la distribuzione di particelle all'interno del corpo del paziente, potendo quindi osservare in tempo reale la deposizione di dose e valutare di conseguenza l'efficacia del trattamento.

2.3. Generazione del fascio radiante

In questo paragrafo si andranno a descrivere le tecnologie utilizzate per la generazione del fascio radiante. Per prima cosa si descriveranno brevemente gli acceleratori lineari utilizzati in radioterapia [18], mentre di seguito, dopo un richiamo sul moto di particelle cariche in un campo magnetico, si accennerà ai due acceleratori di particelle più utilizzati in ambito adroterapico: il ciclotrone [19] e il sincrotrone [20].

Acceleratori lineari

Gli acceleratori lineari (*Linac*) vengono utilizzati nell'ambito della radioterapia per la generazione di fasci di elettroni o di raggi X ad alta energia. La componente principale di un *Linac* consiste in un insieme di cavità risonanti disposte in linea ed operanti nel campo delle microonde. Generalmente, tutti gli acceleratori utilizzati in ambito medico sono basati su cavità lunghe tra 2,5 e 5 m, con diametro di circa 10 cm ed operanti ad una frequenza di circa 3000Hz. Tali cavità hanno lo scopo di generare un fascio di elettroni accelerati, utilizzato direttamente per la terapia con elettroni o per la generazione di raggi X. In quest'ultimo caso, un magnete

deflette il fascio di elettroni in modo che questo incida su un *target* di tungsteno, generando raggi X di *bremsstrahlung* con energie fino ai 20 MeV. Tali raggi vengono poi collimati ed estratti dalla testata dell'acceleratore.

Le unità di trattamento vengono installate con geometrie isocentriche, ovvero in una struttura rotante ad un grado di libertà, chiamata *gantry*, che contiene l'intera guida d'onda [Figura 2.10]. In tal modo è possibile irradiare il paziente con campi provenienti da diverse direzioni complanari.



Figura 2.10: Acceleratore lineare con *gantry* isocentrico.

Il ciclotrone

Il ciclotrone è un acceleratore circolare che basa il suo funzionamento sulle caratteristiche del moto di una particella carica all'interno di un campo magnetico. Tale moto è regolato dalla forza di Lorentz, che viene quindi richiamata di seguito.

Si consideri un campo magnetico uniforme e statico, di intensità B_0 e diretto lungo un generico asse z , assunto perpendicolare al piano descritto dagli assi x ed y :

$$\vec{B} = B_0 \vec{u}_z \quad (2.7)$$

La forza che agisce su una particella avente carica q , che si muove all'interno del campo B con velocità v è data dalla relazione:

$$\vec{F} = q\vec{v} \times \vec{B} \quad (2.8)$$

ed è nota come forza di Lorentz. Si noti che la forza è perpendicolare al campo magnetico, quindi non ha componente lungo z . Questo implica che la velocità della particella lungo z rimane costante e quindi coincidente con quella iniziale. Si noti inoltre che la forza è perpendicolare anche al vettore velocità, quindi risulta sempre perpendicolare alla traiettoria

della particella e il vettore velocità non varia in modulo ma soltanto in direzione. Queste considerazioni permettano di affermare che la particella si muove all'interno del campo magnetico con un moto elicoidale, eventualmente circolare se la velocità iniziale lungo l'asse z fosse nulla.

Si consideri ora un sistema di riferimento in coordinate cilindriche, centrato nel centro dell'orbita della particella: il bilancio di forza in direzione radiale consiste nell'eguagliare la forza di Lorentz con la forza centrifuga, scrivendo:

$$qv_{\theta}B_0 = \frac{mv_{\theta}^2}{r} \quad (2.9)$$

dove v_{θ} è la velocità azimutale della particella, m la massa della particella e r il raggio dell'orbita.

Dall'equazione 2.9 si ricava il raggio dell'orbita e la pulsazione che caratterizza il moto circolare della particella.

$$r = \frac{mv_{\theta}}{qB_0} \quad (2.10)$$

$$\omega = \frac{qB_0}{m} \quad (2.11)$$

Osservando le relazioni 2.10 e 2.11 si nota come all'aumentare della velocità della particella, e quindi della sua energia, aumenti il raggio dell'orbita. La pulsazione, invece, non è influenzata dalla velocità della particella.

Entrambe queste considerazioni sono fondamentali per descrivere il funzionamento di un ciclotrone.

Il ciclotrone, rappresentato schematicamente in Figura 2.11, è costituito da due cavità metalliche a forma di "D" (chiamate appunto *dee*) accostate ed immerse in un campo magnetico verticale (perpendicolare alle *dee*) generato dai poli di un magnete. Tra le due *dee* vi è un spazio vuoto chiamato *gap* di accelerazione.

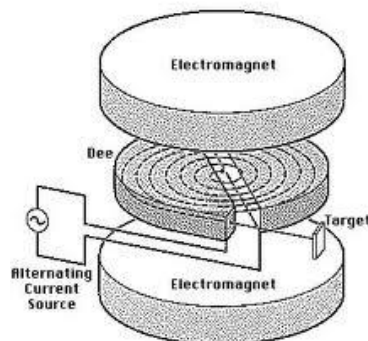


Figura 2.11: Schema delle parti fondamentali di un ciclotrone.

Le due *dee* vengono polarizzate con una tensione alternata avente pulsazione corrispondente a quella caratteristica del moto circolare della particella che si vuole accelerare all'interno di un campo magnetico di intensità pari a quello generato dai magneti (si veda la relazione 2.11).

Il ciclo di accelerazione è il seguente [Figura 2.12].

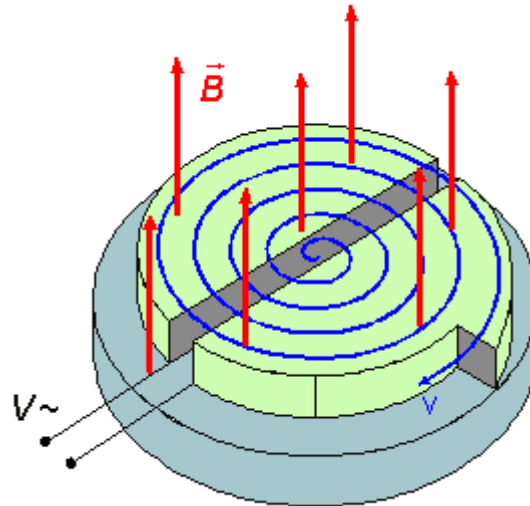


Figura 2.12: Accelerazione di una particella in un ciclotrone. In blu: traiettoria della particella. In rosso: campo magnetico.

Una sorgente di ioni, posta al centro della macchina, emette le particelle da accelerare (con carica positiva), le quali vengono attratte dalla *dee* polarizzata negativamente, subendo una prima accelerazione. Quando la particella entra nella *dee* il suo moto diventa circolare uniforme, essendo soggetta soltanto alla forza di Lorentz e alla forza centrifuga. Compiuta una rotazione pari a metà circonferenza, la particella esce dalla *dee* e attraversa il *gap* di accelerazione, dove, appunto, subisce un'ulteriore accelerazione, in quanto viene attratta dalla seconda *dee*, che, contemporaneamente, viene polarizzata negativamente. La pulsazione della tensione alternata con cui vengono polarizzate le *dee*, infatti, fa sì che la particella, ogni volta che attraversa il *gap*, sia accelerata. Man mano che la particella viene accelerata, il raggio della sua orbita aumenta e, raggiunta l'energia desiderata, la particella viene estratta dalla macchina deflettendo il fascio con opportuni campi.

L'energia del fascio estratto, oltre che dal raggio della macchina, è limitata dal fenomeno relativistico per il quale, all'aumentare della velocità della particella, anche la sua massa aumenta. Ciò comporta uno sfasamento tra la pulsazione del moto circolare della particella (che decresce secondo la relazione 2.11) e quello della tensione di polarizzazione che, invece, rimane costante.

Il ciclotroni producono un fascio la cui intensità è molto stabile, ma la cui energia può essere variata soltanto introducendo filtri assorbitori, non potendo variare semplicemente e velocemente l'energia con cui il fascio viene estratto dalla macchina.

In campo terapeutico, i ciclotroni vengono utilizzati principalmente per accelerare protoni, in quanto l'accelerazione di ioni pesanti, come gli ioni carbonio, richiederebbe macchine dal peso e dal costo eccessivo (a cause delle dimensioni richieste ai poli dei magneti), rendendo preferibili i sincrotroni.

Il sincrotrone

Il sincrotrone, esattamente come il ciclotrone, è un acceleratore circolare e si basa sui principi fisici appena descritti.

La principale differenza tra i due consiste nel fatto che, mentre nel ciclotrone il raggio dell'orbita aumenta all'aumentare dell'energia della particella, nel sincrotrone la particella si muove sempre sulla stessa orbita, definita dalla geometria costruttiva della macchina. Ciò implica una serie di differenze riguardanti sia la costruzione della macchina sia il suo funzionamento.

Da un punto di vista costruttivo, il sincrotrone è costituito principalmente dai seguenti elementi funzionali. Innanzitutto vi è una sorgente dalla quale vengono estratti gli ioni da accelerare. Tipicamente, tali particelle sono ottenute tramite ionizzazione di un gas: ad esempio gli ioni carbonio sono ottenuti ionizzando l'anidride carbonica. Successivamente all'estrazione, le particelle vengono pre-accelerate attraverso un acceleratore lineare per essere poi immesse nell'anello del sincrotrone tramite un'apposita linea di immissione.

L'anello è costituito da una camera a vuoto circolare, all'interno della quale orbita il fascio. Per garantire che il fascio rimanga confinato all'interno di tale camera, è necessario agire su quest'ultimo con campi magnetici volti a focalizzarlo e defletterlo. Per fare ciò, attorno alla camera a vuoto sono disposti dei magneti, che vengono classificati in base al numero di poli utilizzati per generare il campo. Si hanno magneti di dipolo (con 2 poli) la cui funzione è di deflettere il fascio per mantenerlo su un'orbita chiusa, si hanno poi magneti di quadrupolo (con 4 poli), volti a focalizzare il fascio all'interno della camera, facendo sì che vi rimanga confinato. In ultimo, i magneti di sestupolo (con 6 poli) garantiscono un'elevata tolleranza del sistema di focalizzazione.

L'accelerazione del fascio è ottenuta inserendo sulla linea una o più cavità risonanti a radiofrequenza, la cui frequenza può essere modulata su un ampio intervallo.

Risulta chiaro che, per mantenere il fascio su un'orbita a raggio costante e aumentarne l'energia, è necessario che contemporaneamente sia aumentato anche il campo magnetico che mantiene il fascio in orbita. Si ricorda, infatti, che, per mantenere r costante, un aumento di v_θ (e quindi dell'energia della particella) deve essere compensato da un aumento di B_0 .

Non soltanto deve essere variato il campo magnetico, bensì anche la frequenza di risonanza della cavità di accelerazione deve essere variata ad ogni ciclo, visto che la pulsazione caratteristica del moto della particella cresce all'aumentare di B_0 , dunque cresce all'aumentare dell'energia della particella. Se tale frequenza non fosse modulata ciclo per ciclo in modo da coincidere con la frequenza di rotazione del fascio, non si avrebbe accelerazione.

In ultimo, terminato il ciclo di accelerazione, il fascio viene estratto dall'anello e deviato, attraverso appositi magneti, su una linea di estrazione.

In Figura 2.13 è mostrato schematicamente il sincrotrone installato presso il CNAO di Pavia, con evidenziate i componenti appena descritti.

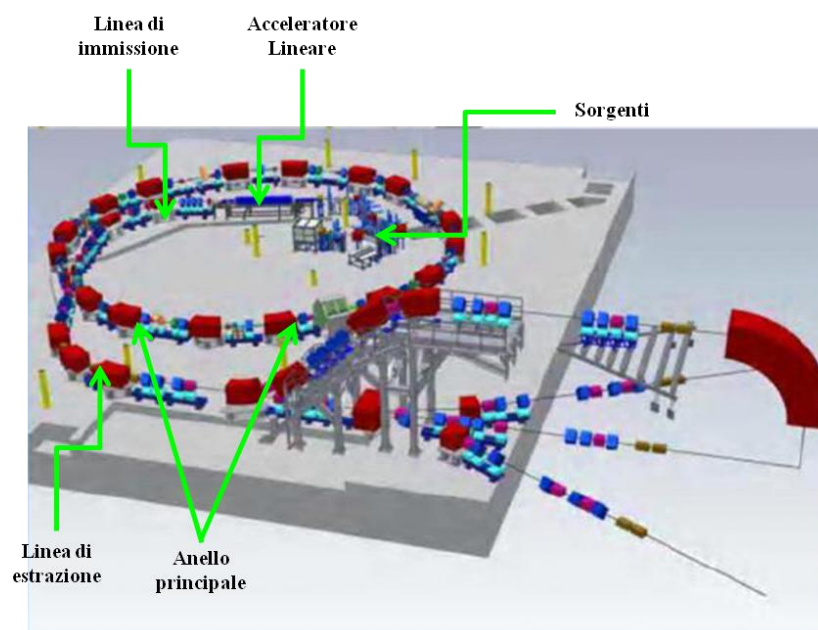


Figura 2.13: Rappresentazione schematica del sincrotrone del CNAO.

Dal momento che sia il campo magnetico sia la frequenza della cavità risonante possono essere modulate, il sincrotrone permette di accelerare ioni fino a velocità relativistiche, non soffrendo del problema dello sfasamento tra pulsazione del fascio e pulsazione della radiofrequenza di accelerazione.

Un ulteriore vantaggio del sincrotrone rispetto al ciclotrone consiste nella possibilità di estrarre il fascio dalla macchina all'energia desiderata. Quest'ultima viene modulata variando il numero

di cicli di accelerazione subiti dalle particelle (che corrisponde al numero di rivoluzioni all'interno dell'anello). Tali variazioni di energia, benché siano possibili, non sempre vengono sfruttate in ambito terapeutico per variare la profondità di penetrazione del fascio, dal momento che introducono notevoli complessità nel controllo della macchina. Inoltre, i tempi richiesti per la variazione d'energia spesso non sono compatibili con le esigenze cliniche, rendendo preferibile l'estrazione di un fascio ad energia costante, successivamente modulata da filtri assorbitori.

Come già accennato, i sincrotroni sono gli acceleratori più utilizzati per l'accelerazione di ioni pesanti come lo ione carbonio, in quanto, pur essendo più grandi e complessi dei ciclotroni, le dimensioni dei magneti richiesti sono minori di quelle che sarebbero necessarie per realizzare un ciclotrone capace di accelerare le stesse particelle. In fine, i sincrotroni mantengono la capacità di accelerare particelle leggere come i protoni.

In conclusione, bisogna notare che gli sforzi richiesti per la costruzione di un sincrotrone sono notevolmente maggiori di quelli necessari per realizzare un ciclotrone, sia dal punto di vista economico sia dal punto di vista logistico ed organizzativo. Un ciclotrone, infatti, ha generalmente un diametro nell'ordine dei metri e un costo per la realizzazione nell'ordine delle centinaia di migliaia di Euro. Un sincrotrone, invece, ha generalmente un diametro di alcune decine di metri, mentre il costo è nell'ordine dei milioni di Euro. Oltre a ciò, la maggiore complessità del sincrotrone comporta tempi di realizzazione notevolmente superiori e la necessità di prevedere ampi spazi aggiuntivi per ospitare i sistemi di controllo e di alimentazione della macchina.

2.4. Strategie di deposito della dose

Dopo aver descritto le macchine utilizzate per generare il fascio radiante, si descriveranno le più comuni strategie e le tecnologie utilizzate per depositare la dose nel volume bersaglio, sia nell'ambito della radioterapia (paragrafo 2.4.1) con fotoni sia nell'ambito dell'adroterapia (paragrafo 2.4.2).

2.4.1. Trattamenti non convenzionali nella radioterapia con fotoni

Per fronteggiare le problematiche esposte relativamente ai trattamenti convenzionali (paragrafo 2.1) sono state sviluppate tecniche dette "non convenzionali" il cui scopo ultimo è quello di aumentare l'accuratezza con cui il fascio radiante colpisce il *target*, aumentando la dose somministrata e dunque l'efficacia della terapia, e contemporaneamente risparmiando i tessuti sani con diminuzione degli effetti collaterali.

Si descrivono di seguito le principali tecniche che permettono questi risultati, quali la radioterapia conformazionale, ad intensità modulata e guidata da immagini.

Radioterapia conformazionale

La radioterapia conformazionale consiste nell'irradiare il volume tumorale con un fascio il cui profilo segua quanto più fedelmente possibile la forma del bersaglio [21].

Requisito fondamentale per questo tipo di trattamento è dunque conoscere non solo la posizione della lesione ma anche la sua forma nelle tre dimensioni. Ciò è possibile grazie ai moderni strumenti diagnostici usati in fase di pianificazione come la tomografia computerizzata (CT), la risonanza magnetica (MRI) o la *Positron Emission Tomography* (PET), tecniche in grado di acquisire immagini volumetriche dettagliate della lesione e degli organi sani circostanti. Di notevole importanza è poi il ruolo ricoperto dai software di pianificazione che permettono all'operatore una visualizzazione tridimensionale della lesione e degli organi adiacenti. Grazie a queste tecnologie si è in grado di risalire al profilo della lesione osservato sotto qualunque angolazione e in particolare sotto quelle dalle quali si somministrerà la dose.

Nota dunque il profilo del bersaglio, il fascio radiante viene conformato su quest'ultimo utilizzando un micro-collimatore multilamellare applicato sulla testa dell'acceleratore [Figura 2.14] o già integrato in essa. Tale dispositivo contiene un serie di coppie di lamelle di tungsteno che vengono automaticamente disposte in modo da conformare il campo di radiazione alla forma del target.

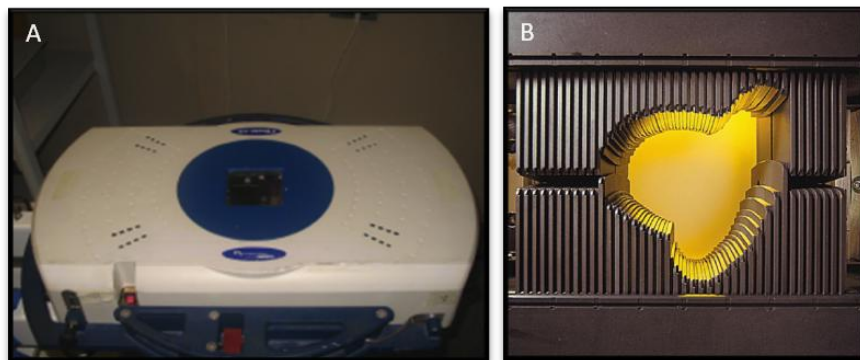


Figura 2.14: Pannello A: Esempio di un micro-collimatore multi lamellare da applicare alla testa dell'acceleratore lineare (Modello M3, BrainLAB). Pannello B: Dettaglio delle coppie di lamelle.

La maggiore precisione del fascio garantisce un danneggiamento minore degli organi sani, riducendo gli effetti collaterali del trattamento e contemporaneamente permettendo di somministrare al bersaglio una dose più elevata rispetto a un trattamento convenzionale. Per citare un esempio, si è osservato che trattamenti conformazionali per il cancro alla prostata hanno presentato una tossicità a lungo termine per il retto del 5%, rispetto al 15% osservato in trattamenti convenzionali [22].

Radioterapia ad intensità modulata

La radioterapia ad intensità modulata si può considerare una evoluzione della radioterapia conformazionale, in quanto il fascio radiante non è modellato solamente nel profilo bensì anche in intensità.

Per ottenere ciò, il fascio radiante viene suddiviso al suo interno in numerosi fasci di piccola sezione, ognuno con una sua intensità, permettendo di colpire il bersaglio con intensità diverse da punto a punto. L'utilizzo di numerosi fasci è poi utile a conformare la dose molto precisamente, permettendo, ad esempio, di modellare il fascio su una superficie curva. Ciò è particolarmente utile quando la massa neoplastica avvolge un organo a rischio [23].

La modulazione di intensità è ottenuta con collimatori dinamici che modificano l'intensità del fascio durante l'erogazione, mediante il movimento della coppie di lamelle.

2.4.2. Sistemi per il rilascio della dose in adroterapia

La problematica che, nell'ambito dell'adroterapia, richiede lo sviluppo di apposite strategie per la deposizione della dose consiste nel fatto che, di norma, le larghezza del picco di Bragg e la sezione trasversale del fascio radiante sono molto minori delle dimensioni del volume da irradiare: è necessario quindi agire sul fascio, sia in direzione longitudinale sia sul piano trasversale, per garantire un'uniforme ed accurata irradiazione del *target*.

A seconda di come si agisce sul fascio, si è soliti dividere i sistemi di deposito della dose in due categorie: i sistemi passivi e i sistemi attivi.

Sistemi passivi

I sistemi puramente passivi modulano il fascio sia longitudinalmente che trasversalmente tramite l'interposizione sulla linea di dispositivi assorbitori e dispersori. I primi hanno lo scopo di modulare l'energia del fascio estratto dall'acceleratore: maggiore è l'assorbimento minore è l'energia con cui le particelle incidono sul corpo del paziente, con la conseguente riduzione del *range* della particella e lo spostamento del picco di Bragg verso profondità minori.

I secondi, invece, incrementando il fenomeno dello *scattering* laterale, causano un aumento della sezione trasversale del fascio, e possono essere abbinati ad ulteriori dispositivi volti ad ottenere una particolare distribuzione trasversale di energia, ad esempio gaussiana o piatta.

Si descrivono di seguito alcune soluzioni tecnologiche utilizzate per realizzare i sistemi appena citati.

La variazione di energia di un fascio ad energia fissata può essere ottenuta inserendo dispositivi in grado di variare l'assorbimento a seconda delle esigenze legate principalmente dalla

profondità a cui si trova il bersaglio da irradiare. Per fare ciò è stato proposto l'utilizzo di una colonna d'acqua ad altezza variabile o di piatti di materiale a basso numero atomico o tessuto-equivalente [24]. Nel primo caso, un pistone varia l'altezza di una colonna d'acqua che viene attraversata dal fascio in direzione longitudinale: maggiore è l'altezza della colonna maggiore è il tratto in cui il fascio perde energia attraversando l'acqua. Nel secondo caso il concetto è analogo, solo che lo spessore dell'assorbitore, e conseguentemente l'energia persa dal fascio, viene variato agendo sul numero di piatti che lo compongono.

La modulazione di energia viene sfruttata, non solo per poter trattare tumori posti a differenti profondità, bensì anche per allargare in direzione longitudinale il profilo di deposizione di dose. Come già detto, la larghezza di un singolo picco di Bragg è molto inferiore di quella del volume bersaglio: per questo motivo la pratica dei trattamenti adroterapici che utilizzano sistemi passivi prevede la sovrapposizione di numerosi picchi di Bragg collocati a profondità diverse, allo scopo di ottenere, nel volume bersaglio, un profilo di dose in funzione della profondità che sia il più possibile uniforme. Tale profilo prende il nome di *Spread Out Bragg Peak* (SOBP) ed è presentato in Figura 2.15.

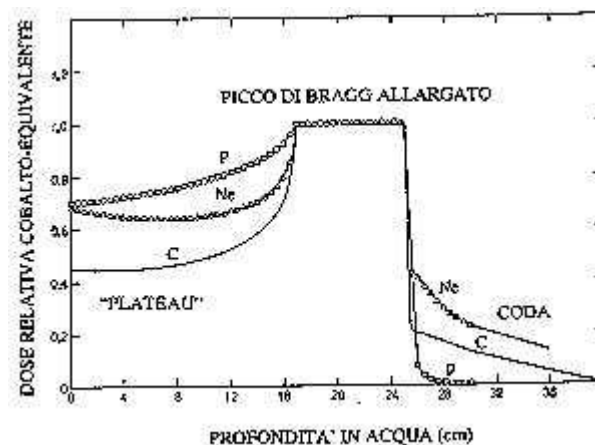


Figura 2.15: Picco di Bragg allargato per protoni, ioni neon e ioni carbonio.

Per ottenere uno SOBP, avendo a disposizione un fascio ad energia fissata, si utilizzano dispositivi rotanti a forma di elica o a spirale [Figura 2.16], realizzati in modo che il fascio attraversi una spessore differente a seconda della posizione angolare del dispositivo. In questo modo si ottengono rapide variazioni di energia che permettono di ottenere picchi di Bragg a diverse profondità in rapida successione temporale.

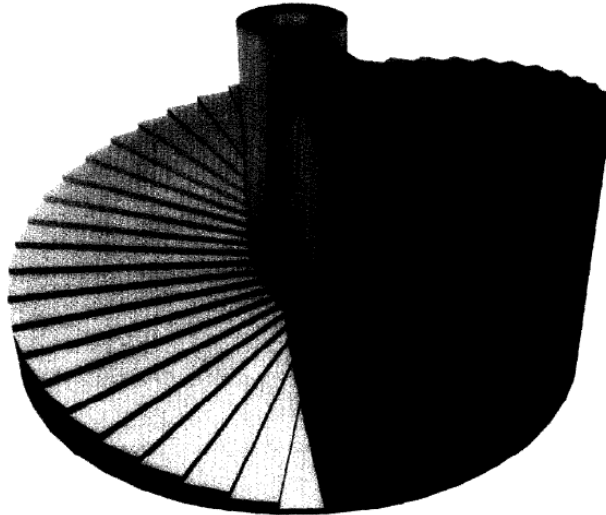


Figura 2.16: Modulatore di energia a spirale.

In ultimo, un'ulteriore modulazione di energia si rende necessaria per compensare le differenze di densità che caratterizzano i tessuti attraversati dal fascio prima di raggiungere il bersaglio. Dalla formula di Bethe-Bloch (equazione 2.1) si osserva che la deposizione di energia all'interno di un mezzo è proporzionale alla densità di quest'ultimo. Ne consegue che, a parità di energia iniziale, le particelle che attraversano tessuti più densi avranno un *range* minore, e viceversa. Per far fronte a tale fenomeno, si utilizzano compensatori personalizzati per ogni paziente e sagomati sfruttando le informazioni di densità ricavate dalle immagini CT. Tali compensatori hanno uno spessore maggiore dove si prevede che il fascio attraversa tessuti meno densi, e viceversa. La Figura 2.17 mostra un compensatore utilizzato per un trattamento al fegato.



Figura 2.17: Compensatore per un trattamento al fegato.

Per quanto riguarda i dispersori, una possibile realizzazione è presentata in Figura 2.18 [25].

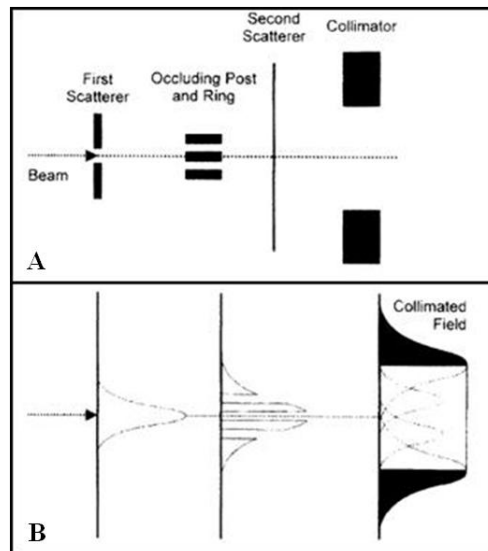


Figura 2.18: Pannello A: Sistema dispersore per ottenere una distribuzione trasversale di energia piatta.

Pannello B: schematizzazione della distribuzione di energia.

Si nota come il fascio, una volta attraversato il primo filtro dispersore, assuma un profilo gaussiano. Successivamente, un sistema di anelli occlude la parte centrale del fascio e le parti esterne. Il fascio così sagomato attraversa un secondo filtro dispersore ed un collimatore, dal quale emerge una distribuzione trasversale di energia uniforme (con profilo piatto), data dalla sovrapposizione di più profili gaussiani. Si nota anche che tale distribuzione decresce abbastanza rapidamente ai bordi.

In conclusione, i sistemi passivi, pur garantendo una semplice ed abbastanza efficace modulazione del fascio, hanno diversi svantaggi, tra i quali la necessità di avere a disposizione fasci con energia iniziale elevata (gran parte della quale viene assorbita prima che il fascio raggiunga il paziente) e la contaminazione del fascio con neutroni e, se si utilizzano ioni pesanti, con frammenti leggeri.

Per far fronte a tali svantaggi e per ottenere una maggiore flessibilità nella deposizione di dose, sono state sviluppate tecnologie basate su sistemi parzialmente o completamente attivi.

Sistemi semi-attivi e attivi

Prima di passare a descrivere il funzionamento dei sistemi puramente attivi, si descriveranno sistemi ibridi, detti semi-attivi, nei quali la distribuzione della dose sul piano trasversale avviene tramite magneti. Tali magneti sono disposti in modo da generare due campi perpendicolari, in grado di deflettere il fascio sul piano trasversale. A seconda di come vengono controllati i magneti è possibile ottenere distribuzioni circolari o rettangolari. Nel primo caso, a entrambi i magneti è applicata una tensione alternata alla stessa frequenza, ma in modo tale che

tra i due vi sia uno sfasamento di 90° . In tal modo un magnete è controllato da una senoide, mentre l'altro è controllato con una cosenoide, dando origine ad un campo magnetico che devia il fascio su una traiettoria circolare e perpendicolare al suo asse, il cui raggio dipende dall'intensità del campo stesso.

Nel secondo caso, invece, un magnete viene controllato con un segnale a rampa lentamente crescente, mentre all'altro viene applicato un segnale veloce a forma di “dente di sega”. Tale tecnica, detta *raster scanning*, muove il fascio a zig-zag per ottenere una distribuzione di dose omogenea all'interno di un rettangolo [Figura 2.19].

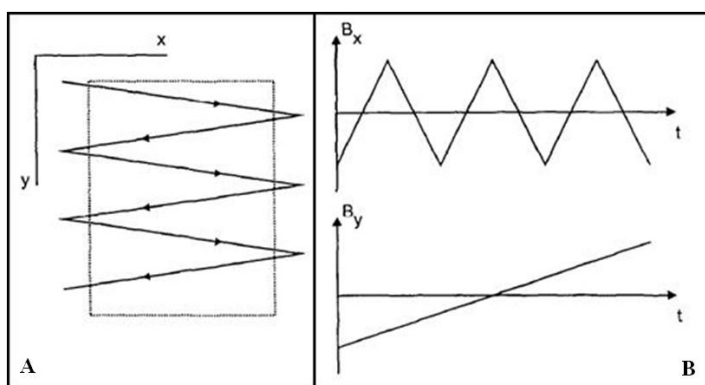


Figura 2.19: raster scanning. Pannello A: andamento a zig-zag del fascio radiante. Pannello B: segnali di controllo dei magneti.

Per quanto riguarda i sistemi puramente attivi, la tecnica di irradiazione è la seguente. In fase di pianificazione, il volume bersaglio viene diviso in unità elementari, dette *voxel*, ognuna delle quali è caratterizzata da tre coordinate: una profondità (z) e una posizione sul piano trasversale al fascio (x,y). Inoltre, per ogni *voxel*, si definisce la dose che deve essere deposita in esso, allo scopo di ottenere un effetto biologico uniforme in tutto il volume. È chiaro che la dose depositata in ogni *voxel* non può essere pianificata indipendentemente da quella rilasciata nei *voxel* circostanti: basti pensare al fatto che il fascio, per raggiungere lo strato più profondo del bersaglio, attraversa tutto il volume di trattamento, rilasciando agli starti più superficiali un dose non trascurabile che andrà considerata in fase di pianificazione.

Le dimensioni dei *voxel* in cui si divide il volume bersaglio sono un parametro rilevante ai fini dell'efficacia del trattamento. È importante che le loro dimensioni (sia sul piano trasversale sia in direzione longitudinale) siano sufficientemente piccole per poter garantire una distribuzione di dose uniforme nel volume di trattamento, al fine di evitare la presenza di “punti freddi”, ovvero zone del bersaglio che ricevono una dose insufficiente per la completa inattivazione delle cellule. Tali “punti freddi”, infatti, possono dar luogo alla nascita di tumori recidivi. In aggiunta, utilizzando *voxel* di piccole dimensioni, si ottiene una distribuzione di dose

maggiormente conformata al volume da irradiare, risparmiando con maggior facilità i tessuti sani circostanti.

Bisogna notare, però, che l'utilizzo di *voxel* dalle dimensioni eccessivamente ridotte comporta un notevole aumento del tempo richiesto dall'irradiazione, dal momento che ogni *voxel* viene irradiato singolarmente.

La tecnica di irradiazione, benché possa essere diversa da sistema a sistema, è, in linea generale, la seguente [Figura 2.20].

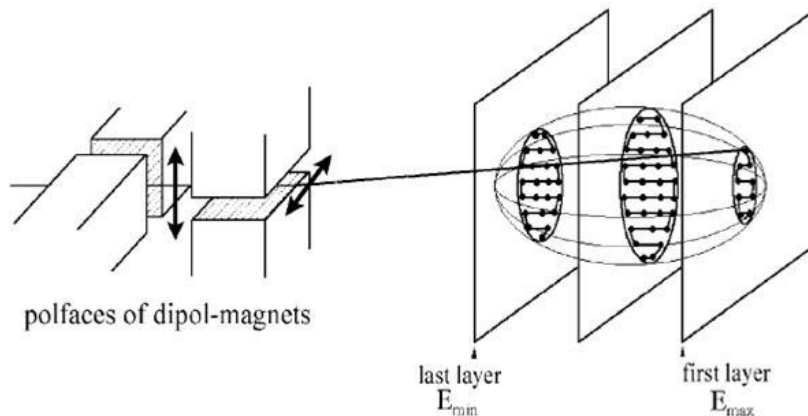


Figura 2.20: Strategia di deposito della dose con sistema attivo [25].

i primi *voxel* ad essere irradiati sono i più profondi, per i quali è richiesta la maggiore energia. Impostata tale energia, tutti i *voxel* della *slice* più profonda vengono irradiati, muovendo il fascio trasversalmente grazie ai magneti di *steering*. La dose rilasciata in ogni *voxel* viene misurata, ad esempio valutando la fluenza delle particella che lo colpiscono e, quando viene raggiunta la dose pianificata, si irradia il *voxel* successivo. Terminata la *slice* più profonda, si riduce l'energia del fascio in modo da poter trattare la *slice* adiacente. Tale procedimento si ripete per tutte le *slice*.

Utilizzando la tecnica appena descritta, l'irradiazione può essere interrotta nel passaggio tra un *voxel* e il successivo: in questo caso è necessario disporre di sistemi in grado di interrompere il fascio molto rapidamente [26]. In alternativa, l'irradiazione può continuare nel passaggio tra un *voxel* e il successivo, consentendo tempi di trattamento più limitati ma causando la deposizione di dose non desiderata. Al fine di minimizzare tale dose indesiderata, i magneti di scansione devono deflettere il fascio radiante in tempi estremamente rapidi (qualche centinaia di microsecondi) [27].

In ultimo, il *pattern* di scansione, invece di prevedere l'irradiazione *voxel* per *voxel*, può essere basato sulla tecnica del *raster scanning*. Utilizzando tale tecnica, la velocità con cui viene mosso il fascio, e quindi le tempistiche del trattamento, devono essere calcolate in accordo che

l'intensità del fascio stesso: è chiaro che, a parità di dose che deve essere depositata, un fascio molto intenso deve essere mosso più velocemente di un fascio ad intensità minore [25].

In conclusione i sistemi attivi garantiscono una maggiore flessibilità ed accuratezza nel deposito della dose, necessitando, però, di tecnologie *hardware* e *software* decisamente più complesse rispetto a quelle richieste per i sistemi passivi.

2.5. Confronto tra radioterapia con fotoni e con adroni

Nel paragrafo 2.2.2 si è discussa la fisica sulla quale si basano i trattamenti di adroterapia e la tecnologia che questi utilizzano. Si vogliono ora evidenziare le principali differenze che vi sono tra tali trattamenti e quelli condotti tramite fasci di fotoni, sia sul piano degli effetti biologici prodotti dalla radiazione [10,11] sia sul piano della deposizione di dose [11].

2.5.1. Effetti biologici

Scopo di questo paragrafo è di fornire una descrizione di come gli effetti biologici prodotti da un campo radiante siano dipendenti dalla natura della radiazione. Per prima cosa si definiranno gli strumenti utilizzati per quantificare tali effetti e la loro dipendenza dalle caratteristiche della radiazione, di seguito si approfondirà lo studio dei suddetti parametri nel caso delle particelle più utilizzate in adroterapia: i protoni [28] e gli ioni carbonio[10,29].

Quantificazione degli effetti biologici

Il principio fondamentale della radioterapia, ovviamente valido anche per l'adroterapia, richiede che sia inflitto il massimo danno possibile al tumore, risparmiando, nel contempo, i tessuti sani. Ottenere il massimo danneggiamento possibile del tumore richiede l'ottimizzazione del trattamento, non solo in termini di accuratezza con la quale si rilascia la dose, ma anche in termini di efficacia biologica della radiazione, ovvero della capacità della radiazione di danneggiare il tessuto tumorale. Risulta quindi evidente la necessità di quantificare l'efficacia biologica della radiazione.

Per fare ciò è necessario, per prima cosa, definire in modo quantitativo il danno arrecato al tessuto biologico e, successivamente osservare come il danno arrecato varia rispetto a quello che si ottiene con un campo radiante di riferimento.

Il metodo utilizzato per quantificare il danno biologico consiste nel sottoporre colture cellulari al campo radiante di interesse e valutare, in funzione della dose assorbita, la percentuale di cellule sopravvissute. In questo modo si ottiene un grafico [Figura 2.21], detto curva di

sopravvivenza, le cui caratteristiche dipendono, oltre che dalla natura delle cellule analizzate, anche dalle caratteristiche della radiazione a cui sono state esposte.

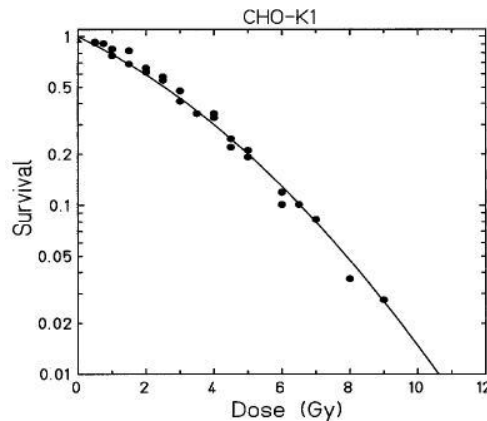


Figura 2.21: Curva di sopravvivenza per una coltura di cellule CHO irradiate con raggi X[30].

Il grafico mostra, a dosi basse, una spalla che è dovuta alla capacità delle cellule di riparare il danno subito. A dosi più elevate, invece, l'andamento è pressoché lineare, indicando che la percentuale di cellule sopravvissute decresce esponenzialmente con la dose assorbita (si noti che la percentuale di sopravvivenza è espressa in scala logaritmica). Tale comportamento può essere approssimato da un andamento lineare-quadratico, descritto dalla seguente relazione:

$$S = S_0 e^{-(\alpha D + \beta D^2)} \quad (2.12)$$

Dove S/S_0 indica la percentuale di cellule sopravvissute, e D la dose assorbita. α e β , invece, sono parametri che rappresentano la risposta alla dose della particolare coltura cellulare.

Confrontando curve di sopravvivenza della stessa coltura cellulare ma ottenute a seguito dell'esposizione a campi di natura diversa, è possibile quantificare l'efficacia biologica di una radiazione. Tale quantificazione è espressa da un parametro detto efficacia biologica relativa (RBE, *Relative Biological Effectiveness*), che indica il rapporto tra la dose assorbita nell'irradiazione con il campo di riferimento rispetto a quella assorbita nell'irradiazione con il campo di interesse, a parità di effetto biologico (equazione 2.13). Tale effetto è generalmente quantificato come la sopravvivenza del 10% delle cellule irradiate (in questo caso, si parla quindi di RBE_{10}). Il campo di riferimento è tipicamente costituito da raggi X a 220 KeV [31].

$$RBE = \frac{D_{ref}}{D} \quad (2.13)$$

Si nota che un campo con $RBE > 1$ ha una maggiore efficacia biologica rispetto al campo di riferimento, in quanto, per ottenere lo stesso effetto, è stata assorbita una dose minore.

Secondo la definizione appena fornita, è immediato ricavare il valore di RBE_{10} di un campo di interesse (ad esempio di particelle), semplicemente confrontando le curve di sopravvivenza ottenute per il campo di interesse e per quello di riferimento [Figura 2.22].

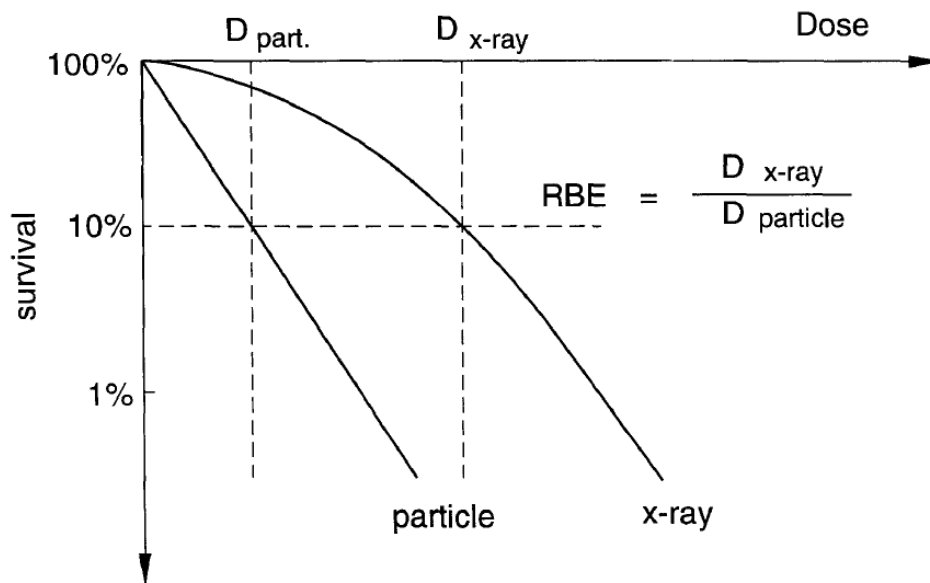


Figura 2.22: Calcolo dell' RBE_{10} dalle curve di sopravvivenza per un generico fascio di particelle.

Un ulteriore modo per calcolare l' RBE , sempre basandosi sul modello lineare-quadratico, consiste nell'effettuare il rapporto tra i coefficienti α delle curve ottenute per il campo di interesse e per quello di riferimento:

$$RBE_{\alpha} = \frac{\alpha}{\alpha_{ref}} \quad (2.14)$$

Poiché il coefficiente α rappresenta la pendenza iniziale della curva di sopravvivenza, si ha un $RBE > 1$ qualora il campo di interesse mostri una curva di sopravvivenza che decresce più rapidamente di quella associata al campo di riferimento, cioè nel caso in cui $\alpha > \alpha_{ref}$.

Come già accennato, il calcolo dell' RBE è fondamentale per l'ottimizzazione di un trattamento di adroterapia, nel quale, ai fini del controllo della malattia, non conta tanto la dose fisica assorbita dal tessuto tumorale, bensì la dose biologica (o dose equivalente). Tale dose, che solitamente è espressa in Gray equivalenti (Gye), coincide con la dose fisica (espressa in Gray) pesata per l' RBE :

$$D_e = D \times RBE \quad (2.15)$$

Si descriverà ora come l' RBE dipende dalle caratteristiche del fascio di particelle utilizzato, quali la dose depositata, il LET ed il numero atomico della particella. In ultimo si discuterà la dipendenza dalle caratteristiche dalla coltura cellulare.

Dal momento che l'RBE dipende dall'effetto biologico che si vuole osservare (equazione 2.12), esso risulta dipendente dalla dose somministrata al campione. L'RBE risulta quindi maggiore per dosi basse mentre tende asintoticamente ad 1 al crescere della dose.

L'RBE, inoltre, risulta fortemente dipendente dal LET della radiazione, visto che tale valore rappresenta l'energia che la particella cede alla materia attraversata. Tale dipendenza è mostrata in Figura 2.23, prendendo come esempio un fascio di ioni carbonio, ed è commentata in seguito [32].

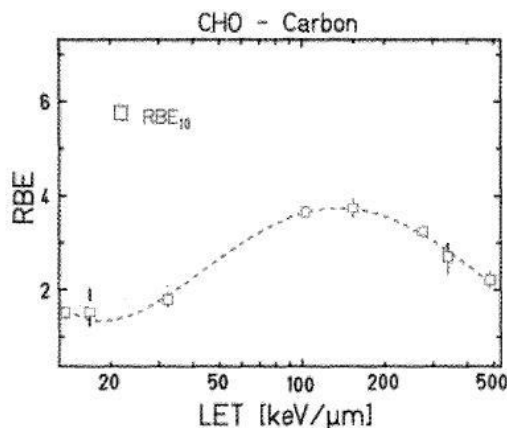


Figura 2.23: Andamento dell'RBE₁₀ in funzione del LET, per un fascio di ioni carbonio che irradia una coltura di cellule CHO.

Si osserva che, a bassi valori di LET, ovvero quando la particella trasferisce alla materia un'energia ridotta, gli eventi di ionizzazione sono meno frequenti e, di conseguenza, il danno arrecato alla coltura è minore. Al crescere del LET, i fenomeni di ionizzazione diventano più frequenti, causando un aumento dell'RBE. Per LET molto elevati, ovvero al termine della traccia, le ionizzazioni risultano estremamente concentrate in una regione di spazio limitata, con l'effetto di causare un danno elevato ma concentrato, "sprecando" la dose per il danneggiamento di una singola cellula. Ciò comporta una riduzione dell'RBE ad elevati LET. Oltre che dal LET, l'RBE dipende dal numero atomico della particella che costituisce il fascio radiante [Figura 2.24].

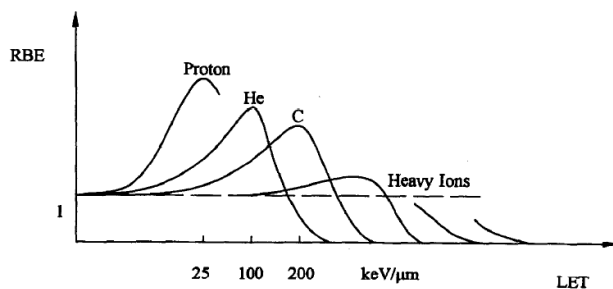


Figura 2.24: Andamento dell'RBE in funzione del LET, per varie particelle.

Si nota come, all'aumentare del numero atomico, il valore massimo di RBE diminuisca e la sua posizione si sposti verso LET più elevati. Tale comportamento può essere spiegato assumendo che l'RBE non dipenda soltanto dal LET ma anche dall'energia della radiazione, dalla quale dipende la distribuzione radiale degli eventi di ionizzazione attorno alla traccia della particella. In ultimo, dopo aver discusso la dipendenza dell'RBE dalle caratteristiche del fascio radiante, si vedrà come questo dipende dalle caratteristiche della cellule irradiate e, in particolare, dalla capacità di riparare i danni subiti.

Come già detto, tale capacità di riparazione è evidenziata da un'ampia spalla visibile nelle curva di sopravvivenza, mentre l'assenza di tale capacità comporta una decrescita esponenziale della percentuale di cellule sopravvissute, che si manifesta in una curva di sopravvivenza dall'andamento lineare.

Tale comportamento fa sì che l'RBE, mostrato in funzione del LET, sia più elevato in colture con elevata capacità di riparazione, come rappresentato in Figura 2.25.

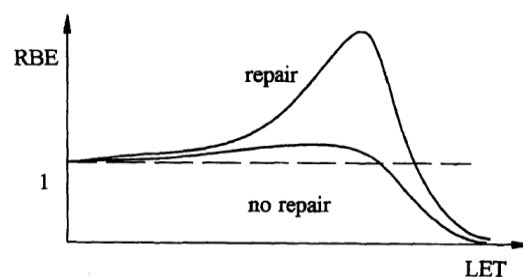


Figura 2.25: Confronto tra l'RBE (in funzione del LET) di una coltura cellulare con forte capacità di riparazione e quello di una coltura con scarsa capacità di riparazione.

Tale dipendenza è di notevole importanza dal punto di vista clinico, dal momento che differenti tessuti biologici sono spesso caratterizzati da differenti capacità di riparare i danni causati dalle radiazioni incidenti. Ciò comporta una notevole differenziazione tra vari tessuti in termini di RBE e quindi di efficacia del trattamento somministrato.

In ultimo, la risposta biologica delle cellule è fortemente dipendente dalla quantità di ossigeno presente nel tessuto in analisi. In particolare si osserva che i tessuti non ossigenati mostrano una minore sensibilità alla radiazione, che si traduce in curve di sopravvivenza la cui pendenza è minore se paragonata a quella di una curva riferita ad un tessuto ossigenato. Per esprimere quantitativamente tale effetto, si introduce un parametro, detto *Oxygen Enhancement Ratio* (OER), definito (equazione 2.16) come il rapporto tra la dose necessaria a produrre un certo effetto biologico (D) e la dose che produrrebbe lo stesso effetto se il tessuto fosse completamente ossigenato in aria a pressione atmosferica (D_0).

$$OER = \frac{D}{D_0} \quad (2.16)$$

Un valore di OER prossimo ad 1 indica che il tessuto in analisi risponde alla radiazione in modo indipendente all'ossigenazione, mentre valori di $OER > 1$ indicano una dipendenza (tanto maggiore quanto maggiore è l'OER) della risposta dall'ossigenazione.

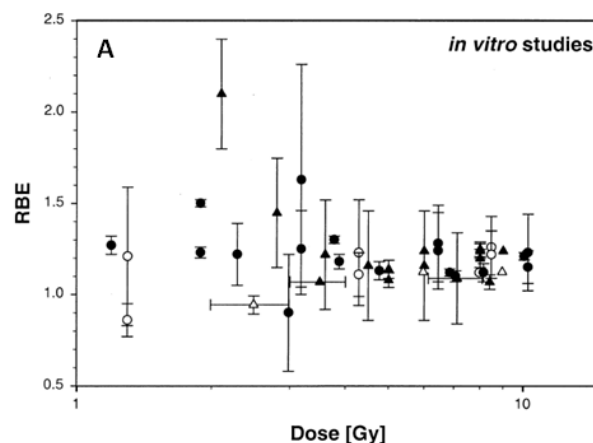
Fissato un certo effetto biologico (ad esempio la sopravvivenza al 10%), l'OER mostra una dipendenza dal LET della radiazione: più precisamente decresce al crescere del LET. Tale dipendenza verrà approfondita successivamente, nel caso particolare di un fascio di ioni carbonio.

Effetti biologici: i protoni

Nella pratica clinica della terapia con fasci di protoni, il valore di RBE utilizzato per la pianificazione del trattamento non ha mai considerato distinzioni riguardo alla dose somministrata per frazione, alla posizione nel picco di Bragg allargato, all'energia iniziale del fascio o al tessuto considerato. Per questo motivo si è soliti parlare di "RBE generico", il cui valore è fissato ad 1.0 o 1.1.

Con il crescere dell'esperienza clinica relativa ai trattamenti con protoni e in vista della crescita del numero di centri che somministreranno tale terapia, l'adeguatezza dell'utilizzo di un singolo valore di RBE per tutti i trattamenti è stata messa in discussione a seguito di studi che hanno ottenuto valori sperimentali di RBE differenti sia per sistemi *in-vitro* sia per sistemi *in-vivo*. Per questo motivo Paganetti *et al.* [28] hanno proposto un studio volto a valutare se le variazioni osservate nell'RBE siano significative da un punto di vista clinico.

La Figura 2.26 mostra i valori sperimentali di RBE misurati *in-vitro* e *in-vivo*, dove si può osservare la variabilità di cui si è appena parlato.



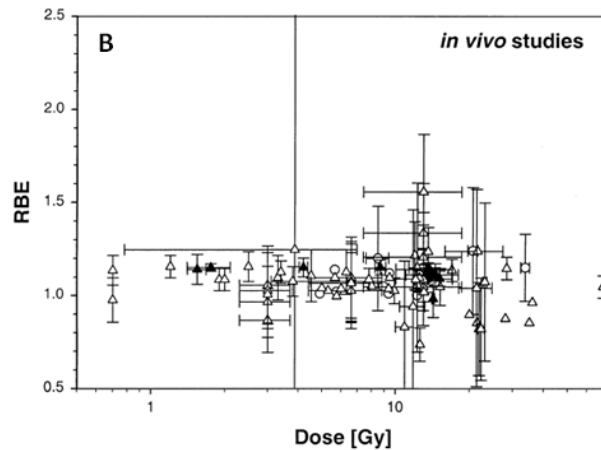


Figura 2.26: RBE in funzione della dose per frazione. Grafico A: studi *in-vitro*. Grafico B: studi *in-vivo*. I cerchi rappresentano fasci con energia < 100 MeV, i triangoli fasci con energia > 100 MeV.

Lo studio di Paganetti *et al.* ha mostrato valori di RBE medi pari a 1.22 ± 0.02 per sistemi *in-vitro*, e 1.10 ± 0.01 per sistemi *in-vivo*, mentre per quanto riguarda la dipendenza di tale parametro dalla posizione nel SOBP, si è osservato un modesto incremento all'aumentare della profondità considerata. Alla luce di questi dati, gli autori concludono che l'utilizzo di un valore di RBE "generico" rimane una pratica adeguata, come dimostrato dal fatto che l'esperienza clinica non dimostra che l'utilizzo di un unico RBE (pari a 1.1) comporti un incremento nella frequenza delle complicazioni a seguito del trattamento.

Effetti biologici: gli ioni carbonio

Si andrà ora ad approfondire la descrizione degli effetti biologici nel caso in cui si utilizzino ioni carbonio, descrivendo il comportamento dell'RBE in funzione della profondità nel tessuto, dell'ossigenazione e della capacità di riparazione dei danni mostrata dal tessuto in analisi.

Per quanto riguarda la dipendenza dell'RBE dalla profondità di penetrazione, si ricordi che man mano che la particella penetra nel tessuto perde energia e si ha un incremento del LET. Come già detto, ad un incremento del LET della radiazione corrisponde un incremento dell'RBE, fino ad un valore massimo che, per gli ioni carbonio, si manifesta intorno a $200 \text{ KeV}/\mu\text{m}$ [33].

L'incremento del LET comporta anche l'incremento della dose rilasciata, dando origine al "picco di Bragg". Risulta quindi evidente la dipendenza dell'RBE dalla profondità di penetrazione.

Per quanto riguarda gli ioni carbonio, l'RBE assume valori prossimi ad 1 nel canale di ingresso (cioè la parte della traccia precedente al picco di Bragg), mentre raggiunge valori fino a 3 al termine del *range* della particella.

Lo studio di tale andamento è risultato fondamentale per identificare nello ione carbonio la particella più adatta ai trattamenti di adroterapia, in quanto un RBE vicino ad 1 nel canale di ingresso limita considerevolmente i danni ai tessuti sani, mentre i valori assunti al termine del *range* garantiscono un efficace danneggiamento del volume tumorale. Un altro ione valutato, l'argon, non risultò adeguato proprio perché mostrava RBE pari a 3 nel canale di ingresso, cosa che comportava un notevole danneggiamento dei tessuti sani con conseguenti effetti collaterali e a lungo termine inaccettabili. In Figura 2.27 è proposto un confronto tra l'RBE (in funzione della profondità di penetrazione) di varie particelle, dove si può osservare come al cresce del numero atomico crescano l'RBE nel canale di ingresso.

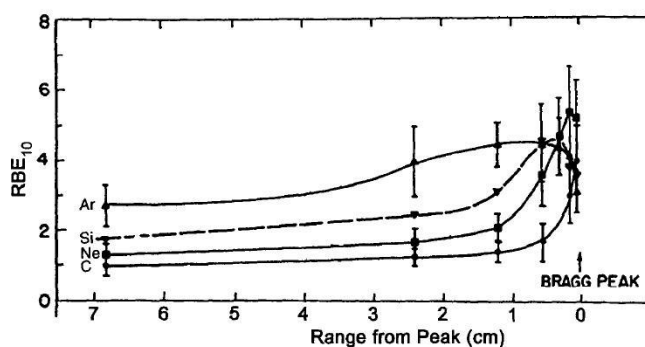


Figura 2.27: RBE in funzione della distanza dal picco di Bragg, per gli ioni carbonio, neon, silicio ed argon.

Come già detto, ai fini di massimizzare l'efficacia del trattamento, è necessario garantire che l'intero volume bersaglio sia danneggiato dalla radiazione in modo uniforme, ovvero che la dose biologica somministrata sia uguale in ogni punto del volume. Per ottenere un profilo di dose biologica uniforme lungo il picco di Bragg allargato, è necessario tenere in considerazione l'incremento di RBE al termine della traccia, diminuendo di conseguenza la dose fisica rilasciata in questa regione, secondo l'equazione (15). Un confronto tra dose fisica e dose biologica (per differenti colture cellulari) è mostrato in Figura 2.28 [34], dove si può notare il decrescere della dose fisica all'aumentare della profondità di penetrazione.

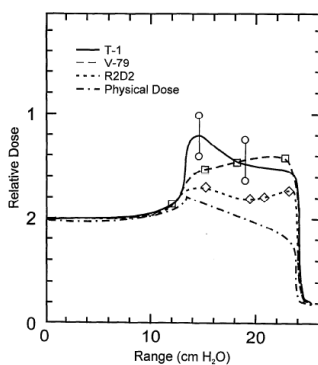


Figura 2.28: Dose fisica e dose biologica (per tre differenti colture cellulari) in funzione della profondità di penetrazione in acqua.

Come già anticipato, la presenza di ossigeno fa sì che i tessuti irradiati mostrino una maggiore sensibilità alle radiazioni. Tale sensibilità è quantificata dall'OER ma, come risulta intuitivo, influisce anche sull'RBE. L'ossigenazione, infatti, agisce alterando la risposta biologica del tessuto considerato.

Al crescere del LET della radiazione, l'OER decresce e, per quanto riguarda gli ioni carbonio, assume valori prossimi ad 1 per LET maggiori di 200 KeV/μm, come mostrato in Figura 2.29 [35].

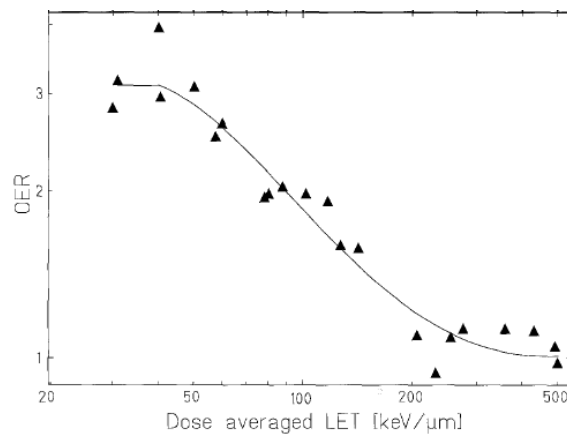


Figura 2.29: OER in funzione del LET per una coltura di cellule V79 irradiate da ioni carbonio

Tale andamento influisce direttamente sull'RBE: se si confronta tale parametro (in funzione del LET) per la stessa coltura cellulare ma in diverse condizioni di ossigenazione, si osserva che la curva ottenuta in condizioni di assenza di ossigeno assume valori decisamente superiori a quelli ottenuti in presenza di ossigeno [Figura 2.30].

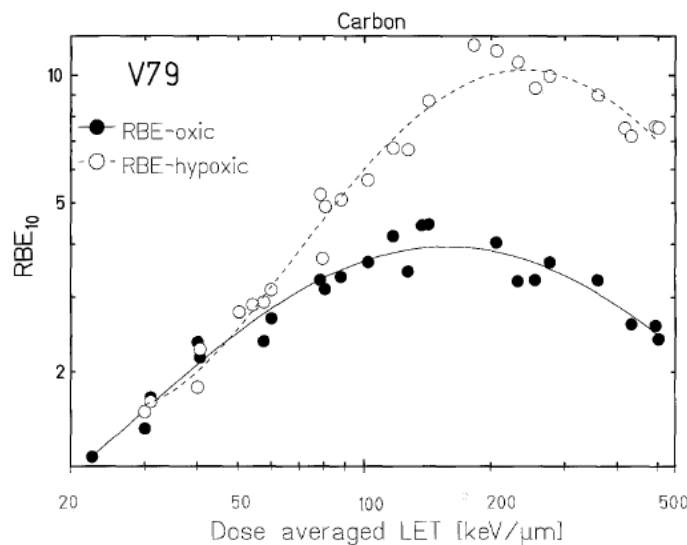


Figura 2.30: RBE in funzione del LET per una coltura di cellule V79, in presenza ed in assenza di ossigeno.

Tale differenza tra le due curve si giustifica ricordando che, per LET maggiori di 200 KeV/μm, l'OER è prossimo ad 1, ovvero si è in condizioni per le quali l'ossigenazione non influisce sulla risposta biologica del tessuto al fascio di ioni carbonio. Si ricordi anche che l'RBE è definito relativamente ad un fascio di raggi X ad energia fissata, per il quale l'effetto ossigeno rimane costante, mentre diminuisce per il fascio di ioni carbonio: per questo motivo le cellule irradiate in assenza di ossigeno mostrano un RBE maggiore, la cui differenza rispetto a quello misurato nella coltura in presenza di ossigeno, è attribuibile soltanto alla diminuzione di OER.

In ultimo si riprende la discussione relativa alla dipendenza dell'RBE dalla capacità di riparazione dei danni mostrata dalla coltura cellulare, presentando alcuni dati ottenuti nell'irradiazione con ioni carbonio. A conferma di quanto già detto riguardo a tale dipendenza, la Figura 2.31 mostra l'andamento dell'RBE in funzione del LET e della profondità di penetrazione, per due colture cellulari: una in grado di riparare i danni subiti (CHO-K1) e una che mostra scarsa capacità di riparazione (xrs-5). Si nota come l'RBE di quest'ultima coltura non mostri il tipico andamento con un massimo a 200 KeV/μm, bensì rimanga pressoché costante in funzione del LET e, di conseguenza, anche in funzione della profondità di penetrazione.

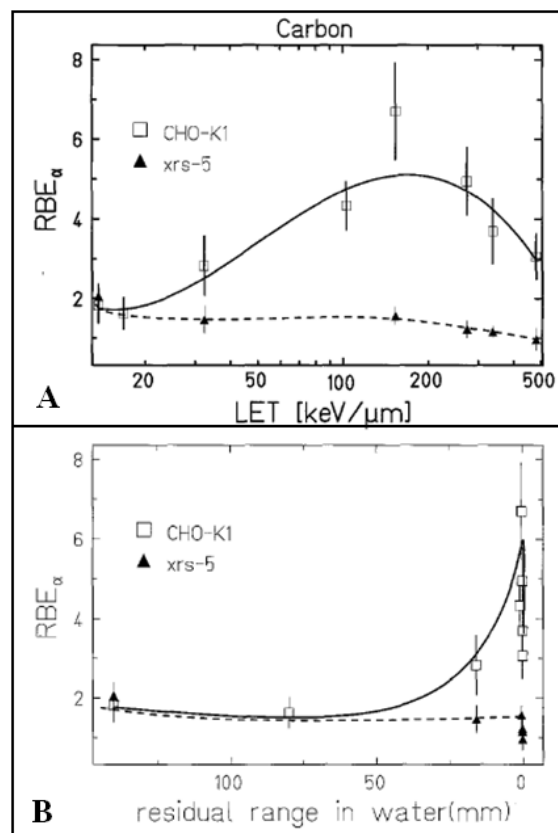


Figura 2.31: RBE in funzione del LET (Pannello A) e del range residuo in acqua (Pannello B), per colture di cellule CHO-k1 e xrs-5.

La relazione tra capacità di recupero e RBE è mostrata in Figura 2.32, in cui tale capacità è quantificata nel rapporto β/α , calcolato a partire dalla curve di sopravvivenza. Si ricordi che la capacità di recupero è evidenziata da una pronunciata spalla nella curva di sopravvivenza, che comporta una prevalenza del parametro β sul parametro α . Dunque un elevato rapporto β/α indica una elevata capacità di riparare i danni subiti.

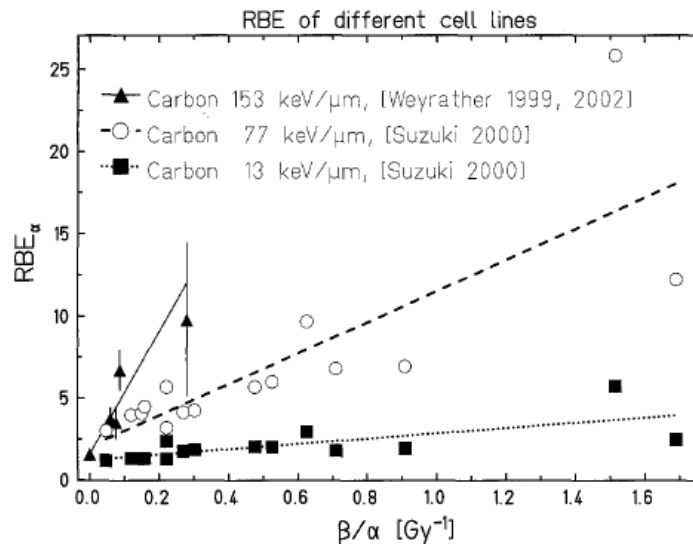


Figura 2.32: RBE $_{\alpha}$ in funzione del rapporto β/α , per tre differenti valori di LET.

In ultimo, si noti che all'aumentare del LET aumentano i valori di RBE misurati e la loro correlazione con la capacità di riparazione diventa più significativa.

2.5.2. Deposito della dose

In questo paragrafo si descriveranno le principali differenze tra un trattamento con fotoni ed un con particelle cariche, dal punto di vista della fisica e della geometria della deposizione della dose all'interno dei tessuti.

I profili di dose in funzione della profondità di penetrazione per fotoni a tre diverse energie e per ioni carbonio a due diverse energie sono mostrati in Figura 2.33.

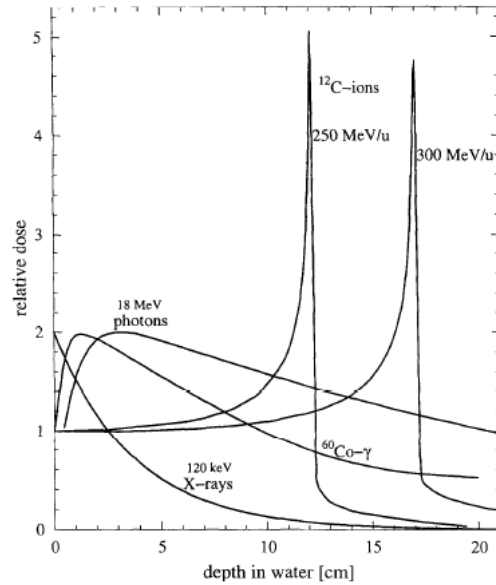


Figura 2.33: Dose relativa in funzione della profondità di penetrazione in acqua per Raggi X a 120 KeV, fotoni γ emessi dal ^{60}Co , fotoni di Bremsstrahlung a 18 MeV, ioni carbonio a 250 MeV/u e a 300 MeV/u.

I raggi X, essendo una radiazione elettromagnetica, sono caratterizzati da una perdita di energia in funzione dello spessore di materiale attraversato regolata da una funzione esponenziale decrescente, secondo la legge di Lambert-Beer:

$$I(z) = I(0)e^{-\mu\rho z} \quad (2.17)$$

dove z rappresenta la profondità di penetrazione, $I(0)$ l'intensità iniziale del fascio, μ e ρ rispettivamente il coefficiente di assorbimento e la densità del materiale attraversato. Il trasferimento di energia avviene per effetto fotoelettrico e, per energie elevate, anche per effetto Compton.

Si noti che l'energia del fascio (e quindi la dose) decresce tanto più rapidamente quanto maggiore è il coefficiente di assorbimento e la densità del mezzo. È quindi chiaro che la maggior dose relativa viene depositata negli strati superficiali del corpo del paziente, ovvero sulla pelle, che è un tessuto molto sensibile alle radiazioni, mentre alla profondità a cui si colloca il bersaglio la dose ceduta è minore.

Tale problematica viene limitata (ma non completamente risolta) utilizzando fasci ad energia maggiore. In questo caso gli elettroni Compton, cioè quelli che la radiazione "strappa" al mezzo, vengono scatterati prevalentemente nella direzione del fascio, trasportando parte dell'energia dalla superficie ad una profondità maggiore, risparmiando la pelle. Oltre al ciò, i fasci ad alta energia subiscono una minore attenuazione all'interno del mezzo, dunque la dose depositata decresce meno rapidamente in funzione della profondità. Per questi motivi, i trattamenti di radioterapia utilizzano fasci ad alta energia (dell'ordine dei MeV).

Un ulteriore svantaggio dell'irradiazione con fotoni consiste nel fatto che l'andamento esponenziale dell'energia depositata fa sì che il fascio depositi dose anche oltre il bersaglio, andando a colpire anche i tessuti sani posizionati posteriormente ad esso.

Per far fronte alle problematiche introdotte dalla deposizione di dose appena descritta, i trattamenti di radioterapia con fotoni prevedono l'utilizzo di tecniche volte a limitare il più possibile la dose assorbita dai tessuti sani, massimizzando quella ceduta al bersaglio. A tal fine, la pratica clinica prevede l'utilizzo di numerosi fasci radianti provenienti da diverse direzioni: così facendo la dose depositata si somma nel volume bersaglio e resta limitata per i tessuti anteriori e posteriori ad esso. Tali fasci sono necessariamente complanari se si utilizzano acceleratori lineari isocentrici, ovvero il cui movimento è limitato alla rotazione attorno all'asse cranio-caudale del paziente. Se, invece, si utilizzano acceleratori dotati di più gradi di libertà, i fasci radianti possono colpire il bersaglio provenendo da qualsiasi direzione.

Per quanto riguarda i fasci di particelle cariche, il profilo di dose è regolato dalla curva di Bragg (paragrafo 2.2.2). Come già accennato, tale profilo risulta particolarmente favorevole per trattamenti adroterapici, in quanto le particelle non soltanto rilasciano la massima dose relativa ad una determinata profondità (regolabile agendo sull'energia), ma anche risparmiano sia in tessuti posti anteriormente al bersaglio, sia quelli posteriori ad esso, in quanto il loro percorso all'interno del mezzo è limitato.

Per questo motivo, l'utilizzo di più fasci provenienti da direzioni differenti non è una priorità nell'ambito di trattamenti di adroterapia, mentre risulta di fondamentale importanza, ai fini dell'efficacia del trattamento, l'accurata determinazione della profondità del *target*, in base alla quale si deve modulare l'energia delle particelle incidenti.

Considerando ora il comportamento del fascio radiante sul piano trasversale, sia nei trattamenti con fotoni sia in adroterapia, la localizzazione del bersaglio sul piano trasversale è fondamentale, in quanto permette di collimare il fascio sul *target* con la massima accuratezza possibile. Nel caso dei trattamenti adroterapici, le caratteristiche della traccia delle particelle influiscono sull'accuratezza con cui il bersaglio viene colpito. Senza considerare eventuali errori di localizzazione del bersaglio o di posizionamento del paziente, è chiaro che una traccia prevalentemente rettilinea assicura una maggiore accuratezza sul piano trasversale. Al contrario, particelle con traiettorie erratiche hanno una maggiore probabilità di deviare significativamente dal percorso pianificato danneggiando tessuti esterni al volume del bersaglio.

Dunque, la selettività geometrica, cioè la capacità di confinare accuratamente la dose all'interno del volume bersaglio, è maggiore nel caso di trattamenti con particelle rispetto a

trattamenti con fotoni, anche considerando il piano trasversale e non soltanto la direzione longitudinale.

In conclusione, i fasci di particelle, se confrontati con i fotoni, consentono di eseguire trattamenti potenzialmente più accurati. La necessità di garantire tale accuratezza, però, introduce nei trattamenti di adroterapia problematiche non presenti o non particolarmente rilevanti nell'ambito della radioterapia con fotoni, che saranno discusse nel prossimo capitolo.

3. Strategie di trattamento per lesioni mobili

Nel paragrafo precedente si è affermato che i trattamenti di adroterapia permettono un'elevata selettività geometrica nella deposizione della dose. Benché questo aspetto sia un indiscutibile vantaggio, comporta anche l'accentuarsi di alcune problematiche tipiche della radioterapia, nonché l'introduzione di altre non presenti nei trattamenti con fotoni.

La finalità di una terapia che utilizza radiazioni ionizzanti per lesioni neoplastiche consiste nel raggiungere la massa tumorale con la maggiore accuratezza possibile. Eventuali inaccurately, infatti, implicano che parte dell'energia del fascio sia rilasciata all'interno di tessuti sani invece che nel volume bersaglio, causando contemporaneamente sia effetti collaterali di rilevante entità sia una minore efficacia del trattamento in termini di controllo della malattia.

Risulta evidente che, in un trattamento adroterapico, la maggiore selettività geometrica potenziale costituisce un vantaggio solo se si garantisce un'accuratezza estremamente elevata, sicuramente maggiore di quella richiesta in un trattamento di radioterapia convenzionale. I danni causati da un'inaccuratezza della medesima entità sono molto maggiori in un trattamento di adroterapia piuttosto che in un trattamento con fotoni, proprio in virtù della maggiore selettività geometrica e della maggiore efficacia biologica mostrata dai fasci di particelle, come descritto nei paragrafi 2.5.1 e 2.5.2.

Inoltre, in un trattamento di adroterapia, l'accuratezza non dipende soltanto dal posizionamento del paziente rispetto al fascio, bensì dipende significativamente anche delle caratteristiche del fascio stesso, soprattutto in termini di energia (profondità del picco di Bragg). Per tale motivo è richiesta una precisa localizzazione tridimensionale del volume bersaglio, oltre che sul piano perpendicolare all'asse del fascio, anche in direzione parallela a quest'ultimo.

Infine, la natura dei tessuti attraversati, ed in particolare la loro densità, assume un ruolo non trascurabile in un trattamento di adroterapia se confrontato con un trattamento di radioterapia con raggi X. In quest'ultima tipologia di trattamento, infatti, l'informazione sulla densità dei tessuti è utilizzata per calcolare l'attenuazione subita dal fascio all'interno del corpo del paziente ma non influisce sulla geometria della deposizione di dose. In adroterapia, invece, la densità dei tessuti attraversati modula il *range* delle particelle, agendo sulla profondità del picco di Bragg e quindi sulla geometria con la quale viene depositata l'energia del fascio.

Di seguito si presenterà il principale fenomeno che limita l'accuratezza dei trattamenti di radioterapia e adroterapia extra-cranica, ovvero l'*organ motion*. Successivamente si descriveranno le tecniche di Radioterapia e Adroterapia Guidata da Immagini (IGRT e IGHT), riportando alcuni esempi delle soluzioni tecnologiche utilizzate per la loro implementazione. In

ultimo, riferendosi allo stato dell'arte di tali tecnologie e alle esigenze cliniche, verrà presentato lo scopo del progetto sul quale verte il presente elaborato.

3.1. Organ Motion

Per “*organ motion*” si intende il movimento che alcuni organi compiono in condizioni basali per effetto di processi fisiologici [36]. Basandosi su questa definizione si può dedurre che il fenomeno si presenti nei distretti corporei nei quali i processi fisiologici implicano spostamenti o deformazioni delle strutture interne. È quindi evidente che il distretto cranico non sia soggetto ad *organ motion* mentre lo siano le zone toraciche ed addominali, dove l'apparato digerente, l'apparato respiratorio, il sistema cardiocircolatorio e l'apparato escretore generano, durante la loro attività, spostamenti di entità non trascurabile.

Dal momento che le lesioni tumorali sono solidali con la struttura sulla quale si trovano, anch'esse risultano soggette a movimento qualora si trovino in posizione toracica o addominale; l'ampiezza di tale movimento dipende dalla collocazione anatomica del tumore.

Analizzando la problematica dal punto di vista di un trattamento di radioterapia od adroterapia, si è soliti classificare i fenomeni di *organ motion* su base temporale, distinguendo tra *organ motion* inter-frazione ed *organ motion* intra-frazione [37].

L'*organ motion* inter-frazione è dovuto a processi il cui effetto è apprezzabile solo tra una seduta di trattamento e la successiva (cioè in un tempo dell'ordine dei giorni) e non all'interno della seduta stessa. Si ritiene che questa tipologia di *organ motion* interessi principalmente organi che sono parte dell'apparato digerente o che sono prossimi ad esso. Un esempio emblematico si trova osservando la mobilità della prostata, la cui posizione, che dipende in gran parte dal riempimento di retto e vescica, varia giorno per giorno a seconda delle funzionalità intestinali e dell'apparato escretore del paziente.

L'*organ motion* intra-frazione, invece, è causato da processi il cui effetto è apprezzabile anche nell'arco temporale ricoperto da una singola seduta di trattamento. In questo caso si ritiene che i principali responsabili del movimento siano il ciclo respiratorio, la contrazione cardiaca e la pulsazione dell'aorta, tutti fenomeni che si ripetono con frequenza dell'ordine dei secondi.

Considerando ad esempio le lesioni polmonari, soggette contemporaneamente ai tre fattori citati, sono riportati movimenti medi di 3,9 mm in direzione cranio-caudale e 2,4 mm in direzione latero-laterale ed antero-posteriore, con massimi di 12 mm in direzione cranio-caudale e 5 mm in direzione latero-laterale ed antero-posteriore [3]. Tali spostamenti, data la loro entità, non possono essere trascurati.

Poiché il fenomeno dell'*organ motion* costituisce una notevole fonte di incertezza riguardo alla posizione del bersaglio da irradiare, esso contribuisce a limitare l'efficacia dei trattamenti, sia facendo sì che il fascio radiante non risulti costantemente centrato sul *target*, sia imponendo l'adozione di margini di sicurezza e/o la riduzione della dose somministrata per evitare di danneggiare i tessuti sani circostanti alla lesione.

Oltre a ciò, l'*organ motion* comporta anche significative variazioni delle caratteristiche dei tessuti e principalmente della loro densità [38]. Si pensi, ad esempio, alla densità polmonare, che diminuisce all'aumentare del volume d'aria inspirato. Come già detto, nell'ambito dell'adroterapia, la densità dei tessuti attraversati dal fascio di particelle è un dato fondamentale in quanto influisce sul *range* di quest'ultime.

3.2. Radioterapia guidata da immagini

La radioterapia guidata da immagini (IGRT, *Image Guided Radiation Therapy*) è una tecnica basata sull'acquisizione di immagini del paziente direttamente in sala di trattamento, allo scopo di minimizzare, ad ogni seduta, l'errore derivato dall'incertezza nella localizzazione del bersaglio. Tale errore è causato sia da un diverso posizionamento del paziente rispetto alla fase di simulazione sia dal movimento al quale sono soggetti gli organi interni, ovvero all'*organ motion*.

Per ridurre gli effetti di tali errori, vengono acquisite immagini della zona da trattare negli istanti immediatamente precedenti all'irraggiamento del paziente. A tal fine vengono generalmente utilizzate tecniche basate sui raggi X, sistemi di localizzazione ottica od ultrasuoni. Di seguito vengono presentati alcuni esempi significativi di queste tecniche.

- Acquisizioni radiografiche stereoscopiche

Tale tecnica prevede l'acquisizione di immagini radiografiche tramite un sistema di *imaging* a raggi X costituito da due coppie sorgente-detettore. Tali coppie sono installate in sala di trattamento in modo tale che le immagini acquisite risultino ortogonali tra di loro.

Le immagini acquisite precedentemente all'irradiazione vengono confrontate con ricostruzioni digitali, chiamate *Digital Reconstructed Radiography* (DRR), ottenute dal software di pianificazione in accordo con la geometria del trattamento. Per facilitare il confronto e la stima delle correzioni da applicare alla posizione del paziente per riportarlo nella posizione nominale, molto spesso vengono inseriti chirurgicamente dei reperi radio-opachi in prossimità del tumore [Figura 3.1].



Figura 3.1: Schermata del software di gestione del trattamento, in cui si osservano i reperi radiopachi impiantati nel bersaglio (in questo caso, la prostata).

Da questo confronto si ottiene lo scostamento del bersaglio rispetto alla posizione in cui era durante la simulazione. In base a questo *shift*, la posizione del paziente viene corretta in modo che ad ogni seduta l'isocentro della macchina coincida con l'attuale posizione del volume bersaglio.

- ***In-room CT***

In questo caso, l'acquisizione delle immagini di controllo avviene attraverso uno scanner CT installato direttamente in sala di trattamento [Figura 3.2]. Con tale tecnica non solo è possibile localizzare il bersaglio attraverso immagini di tipo volumetrico, ma è anche possibile effettuare valutazioni di tipo dosimetrico immediatamente prima dell'irradiazione [39].



Figura 3.2: Sistema di IGRT basato *in room CT*.

- Fluoroscopia

Per fluoroscopia si intende l'acquisizione in tempo reale di immagini a raggi X. Tale tecnica, oltre alla localizzazione del bersaglio precedentemente all'irradiazione, permette di monitorarne la posizione durante il trattamento, risultando particolarmente utile nel caso di lesioni mobili. Il principale svantaggio associato alle tecniche fluoroscopiche consiste nella somministrazione di dose non desiderata al paziente.

Dal punto di vista tecnologico, un fluoroscopio consiste in un sorgente di raggi X e in un detettore (come un intensificatore d'immagine o un *flat panel*) in grado di acquisire immagini in tempo reale. La coppia sorgente-detettore può essere installata su un arco a C o direttamente sull'acceleratore lineare, come avviene ad esempio nel sistema Varian® On-Board Imager® [40][Figura 3.3].



Figura 3.3 Sistema Varian® On-Board Imager®, si nota la coppia sorgente-detettore installata sull'acceleratore lineare.

- Sistemi di localizzazione ottica

Tali sistemi sono costituiti da almeno due telecamere ad infrarossi in grado di localizzare nello spazio la posizione di marcatori riflettenti (od emittenti).

L'utilizzo di sistemi di localizzazione ottica in IGRT prevede che, precedentemente alla CT di pianificazione, vengano posti sulla superficie corporea del paziente dei marcatori riflettenti e radiopachi. In tal modo è possibile identificare la posizione del bersaglio nel sistema di riferimento definito dalla configurazione di marcatori esterni (sistema locale). Immediatamente prima dell'irradiazione, i marcatori devono essere riposizionati esattamente nella posizione in cui si trovavano durante la CT di pianificazione. Successivamente, il sistema di localizzazione acquisisce la loro posizione spaziale e, nota le coordinate del bersaglio nel sistema di riferimento locale, è possibile calcolare

gli *shift* da applicare al fine di portare il paziente in posizione nominale, ovvero quella in cui il *target* è collocato nell'isocentro dell'acceleratore.

In alternativa all'utilizzo di marcatori esterni, la localizzazione ottica può avvenire tramite l'acquisizione di *patch* di superficie corporea, ad esempio tramite la localizzazione di uno *spot* laser che viene riflesso da quest'ultima. In questo caso, l'allineamento del paziente avviene effettuando un *matching* tra la superficie acquisita in fase di pianificazione e quella acquisita immediatamente prima del trattamento.

- Tecniche ultrasonografiche

Tali tecniche prevedono l'acquisizione *in room* di immagini ecografiche in modalità B-Mode tramite un'apposita sonda sulla quale sono installati marcatori riflettenti. In questo modo è possibile conoscere posizione ed orientamento della sonda nel momento in cui vengono acquisite le immagini trasversali e sagittali della zona da irradiare. Tali immagini possono essere sovrapposte ad altre ricavate dalla CT di pianificazione, ovviamente in accordo con la geometria dell'acquisizione ultrasonografica, ottenendo di conseguenza le correzioni da effettuare per portare il paziente nella posizione pianificata.

In conclusione, le tecniche IGRT rendono massima la probabilità che il fascio radiante sia accuratamente centrato sul bersaglio, dunque i margini di sicurezza imposti dalla necessità di essere sicuri di comprendere il bersaglio nella zona irradiata possono essere ridotti. È così possibile aumentare la dose erogata durante la singola frazione e, conseguentemente, ridurre il numero di frazioni alle quali il paziente si deve sottoporre.

3.3. Adroterapia guidata da immagini

Le tecniche di *Image Guided Hadrontherapy* (IGHT) consistono nell'estensione delle tecniche di radioterapia guidata da immagini (IGRT) all'adroterapia.

Esattamente come nella radioterapia con fotoni, anche in adroterapia queste tecniche sono volte a massimizzare l'accuratezza del trattamento garantendo la minimizzazione degli errori di set-up. Tali errori sono introdotti durante la procedura di posizionamento del paziente prima dell'irradiazione e sono definiti come lo scostamento tra la posizione attuale del paziente e la posizione "nominale", ovvero quella stabilita in fase di pianificazione.

Se garantire un'elevata accuratezza è estremamente importante in radioterapia, in adroterapia è assolutamente indispensabile, per i motivi già elencati all'inizio del presente capitolo. È dunque

chiaro che le tecniche IGHT rivestono un ruolo fondamentale nei trattamenti con fasci di particelle.

Per assicurare l'accuratezza necessaria per un efficiente trattamento adroterapico, i sistemi di IGHT richiedono un elevato livello di complessità e l'integrazione di diversi sottosistemi. In generale, un moderno sistema di IGHT comprende i seguenti sottosistemi: un sistema di *imaging in-room*, cioè allestito in sala di trattamento, un sistema di localizzazione ottica e, infine, un dispositivo per il posizionamento automatico del paziente.

Di seguito si descriveranno, da un punto di vista generale, le funzionalità associate ai sistemi appena citati, per poi fornire un panoramica delle modalità con le quali tali sistemi sono installati presso i principali centri di adroterapia.

3.3.1. Soluzioni tecnologiche

Il sistema di *imaging in-room* è, in generale, costituito da un dispositivo di acquisizione di immagini a raggi X. Ad esempio può essere composto da almeno due sorgenti e due detettori posti ortogonalmente tra di loro per poter effettuare acquisizioni stereoscopiche, oppure può consistere in uno scanner tomografico.

Lo scopo di questo sistema è di acquisire immagini radiografiche (o tomografiche) del paziente in sala di trattamento. Tali immagini, se confrontate con il set di immagini diagnostiche (CT, MRI, PET) ottenuto in fase di pianificazione del trattamento, permettono di individuare e quantificare accuratamente eventuali errori di set-up, espressi in termini di tre traslazioni lungo gli assi cartesiani e di tre rotazioni attorno a tali assi.

Inoltre, le immagini ottenute permettono anche la localizzazione di eventuali marcatori radiopachi impiantati nella lesione.

Nel caso in cui si utilizzi un sistema basato su coppie sorgente-detettore, l'acquisizione di immagini RX, oltre che bidimensionale, può avvenire in modalità tridimensionale, qualora l'installazione di tali coppie sia in grado di ruotare attorno al paziente. In tal caso il sistema può acquisire immagini tomografiche sia con tecnica *Cone Beam CT* (CBCT), in cui il sistema compie un'intera rotazione attorno al paziente, sia con tecnica *Digital Tomosynthesis* (DTS), in cui la rotazione è ridotta ad un angolo inferiore a 360° , con il vantaggio di limitare la dose aggiuntiva assorbita dal paziente e il tempo di acquisizione, limitando, però, la risoluzione delle immagini ottenute [41].

Il sistema di localizzazione ottica comprende un set di emettitori o marcatori riflettenti nell'infrarosso che vengono collocati sulla superficie corporea del paziente e una serie di almeno due telecamere IR che rilevano la posizione di tali marcatori [Figura 3.4].



Figura 3.4: Rappresentazione schematica del funzionamento di un sistema di localizzazione ottica con marcatori riflettenti.

Il sistema è completato da un apposito software che ricostruisce la posizione spaziale dei marcatori a partire dai segnali provenienti dalla telecamera. Tale sistema ha come scopo sia quello di garantire un primo posizionamento del paziente (prima dell'acquisizione delle immagini RX) sia di monitorare la posizione del paziente in tempo reale, assicurandosi che questa non subisca alterazioni durante l'intero trattamento. Tale controllo, essendo in tempo reale, non può essere delegato efficacemente ad un sistema di *imaging in-room*. Quest'ultimo, infatti, sebbene possa essere in grado di acquisire immagini RX in continua, presenta lo svantaggio di depositare dosi indesiderata (in quanto non terapeutica) nel corpo del paziente. Per questo motivo e per l'elevata accuratezza, i sistemi ottici risultano preferibili a quelli radiografici per assolvere alle funzionalità appena citate.

Inoltre, il posizionamento del paziente deve essere effettuato tramite un sistema di movimentazione robotizzato. Da un punto di vista meccanico, il sistema deve godere di sei gradi di libertà, ovvero deve essere in grado di eseguire tre traslazioni e tre rotazioni nello spazio, garantendo, per ognuno di questi movimenti, un'elevata accuratezza. In aggiunta, poiché il paziente deve essere vincolato rigidamente al sistema di posizionamento, questo dovrà comprendere dispositivi di aggancio per il supporto dove si colloca il paziente (ad esempio un lettino o una sedia).

In ultimo, il sistema di posizionamento deve essere in grado di svolgere movimenti in modo automatico, controllato dalle informazioni provenienti sia dal piano di trattamento, sia dal sistema di *imaging in-room*, sia dal sistema di localizzazione ottica.

Sebbene gli strumenti appena descritti vengano utilizzati in modo differente nei diversi centri che propongono trattamenti di adroterapia, non si può prescindere dall'loro utilizzo per la corretta localizzazione e il monitoraggio del paziente.

Fatta eccezione per CNAO, i cui sistemi e le cui scelte progettuali saranno descritte nel paragrafo 3.3.2, le soluzioni tecnologiche adottate nei principali centri di adroterapia sono:

- PSI (Paul Scherrer Institute, Villigen, Svizzera)

In questo centro, dove si effettua adroterapia con protoni, il paziente viene preparato e posizionato esternamente alla sala di trattamento [42].

Il paziente viene quindi preparato per il trattamento in una sala di preparazione dedicata, dove viene immobilizzato su di un lettino con sistemi di immobilizzazione personalizzati. Attraverso un apposito supporto mobile viene portato in sala di posizionamento, dove è installato uno scanner CT analogo a quello che viene utilizzato per la pianificazione del trattamento. Il lettino viene vincolato al tavolo dello scanner tramite lo stesso meccanismo di posizionamento presente in sala di trattamento, in modo da garantire un'accurata riproducibilità della posizione del paziente. L'acquisizione tomografica viene utilizzata per il posizionamento, e il paziente viene successivamente portato in sala di terapia. Ulteriori verifiche vengono effettuate con un sistema di *imaging* radiografico installato direttamente sul *gantry* isocentrico, con il quale viene somministrata la dose [Figura 3.5].

Recentemente i sistemi di IGHT presso il PSI sono stati integrati dal sistema di localizzazione ottica Vision RT, adibito all'acquisizione in tempo reale di superfici 3D.



Figura 3.5: Sala Gantry 2 presso il PSI

- HIT (Heidelberg Ion-Beam Therapy Center, Heidelberg, Germania)

A differenza di quanto avviene presso il PSI, presso l'HIT il posizionamento del paziente avviene direttamente in sala di terapia. Al tal fine è installato un sistema di IGHT sviluppato dalla Siemens [43] che comprende un dispositivo di posizionamento robotizzato e un sistema di *imaging* a raggi X installato su un arco a C vincolato al soffitto della sala [Figura 3.6].



Figura 3.6: Sala di trattamento presso l'HIT. Sono visibili sia il sistema di posizionamento robotizzato sia l'arco a C.

Il trattamento tramite fasci di protoni o ioni carbonio viene somministrato con fasci fissi o tramite un *gantry* isocentrico.

- HIMAC (Heavy Ion Medical Accelerator in Chiba, Chiba, Giappone) e GHMC (Gunma University Heavy Ion Medical Center, Gunma, Giappone)

Presso tali centri, che eseguono trattamenti sia con fasci di protoni che di ioni carbonio, il sistema di *imaging* utilizzato per il posizionamento del paziente consiste in uno scanner CT installato in sala di trattamento. Analogamente a quanto descritto riguardo agli altri centri, il posizionamento avviene tramite un lettino robotizzato [Figura 3.7]



Figura 3.7: sistema di IGHT presso al GHMC. Si notano lo scanner CT e il lettino robotico.

Presso HIMAC, il sistema di localizzazione ottica viene utilizzato per l'implementazione di un sistema di *gating* respiratorio. Tale sistema esegue un monitoraggio della fase respiratoria del paziente durante l'intero trattamento, utilizzando una telecamera che osserva il movimento di un sensore posto sul paziente [Figura 3.8] . In tal modo è possibile irradiare il bersaglio soltanto durante un preciso intervallo del ciclo respiratorio, riducendo di conseguenza le inaccuranze introdotte dall'*organ motion* dovuto alla respirazione.



Figura 3.8: Telecamera e sensore utilizzati per il *gating* respiratorio presso HIMAC.

Un sistema di *gating* (Anzai Medical) è presente anche presso il GHMC, dove un sensore laser esegue il monitoraggio in tempo reale della distanza tra un supporto fisso e la superficie corporea del paziente [Figura 3.9]. In tal modo si ottiene un segnale che può essere utilizzato per il *gating*.



Figura 3.9: Dispositivo per il *gating* respiratorio utilizzato presso GHMC.

3.3.2. Implementazione di sistemi di *image guidance* presso CNAO

Il Centro Nazionale di Adroterapia Oncologica (CNAO) di Pavia è il primo centro italiano che eseguirà trattamenti di adroterapia con protoni e ioni carbonio. È dotato di tre sale di trattamento: due sale sono caratterizzate da un unico fascio fisso orizzontale (sale laterali) mentre una terza sala (sala centrale) disporrà di un fascio orizzontale e di uno verticale, entrambi fissi. Sono inoltre previsti sviluppi futuri per un'ulteriore linea di fascio dedicata alla ricerca e all'utilizzo di un *gantry* isocentrico. Nelle sale laterali sono state già installate le soluzioni di IGHT, mentre nella sala centrale è al momento in fase di studio una soluzione idonea agli ingombri geometrici della linea verticale.

Pertanto, verranno descritti in questo paragrafo i sistemi per IGHT installati nelle due sale laterali di CNAO [44].

Come già detto nel paragrafo 3.3.1, un possibile sistema IGHT comprende tre sottosistemi integrati [Figura 3.10]: un sistema di posizionamento del paziente o PPS (*Patient Positioning System*), un sistema di verifica della posizione del paziente o PVS (*Patient Verification System*) e un sistema di localizzazione ottica o OTS (*Optical Tracking System*).

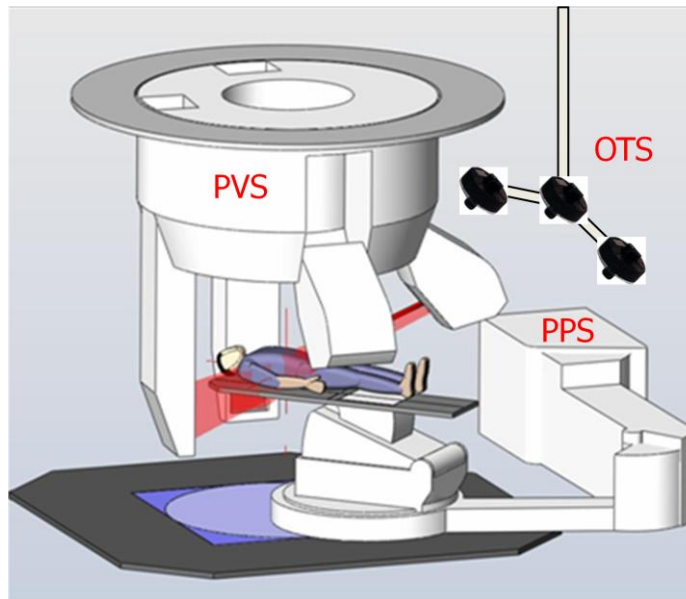


Figura 3.10: Rappresentazione schematica dei tre sottosistemi integrati nelle sale di trattamento laterali del CNAO.

Il sistema di posizionamento del paziente (PPS) consiste in un braccio robotico in grado di vincolarsi automaticamente sia ad un lettino sia ad una apposita sedia di trattamento [Figura 3.11].

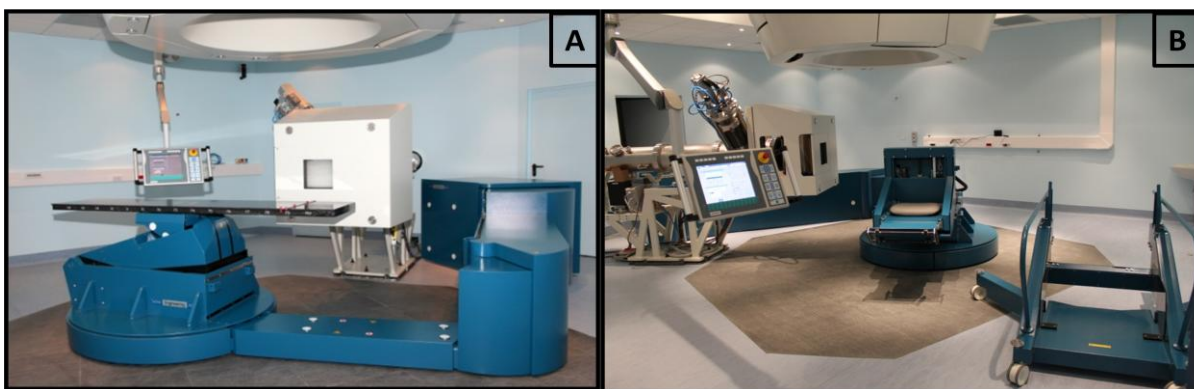


Figura 3.11: Pannello A: PPS vincolato al lettino. Pannello B: PPS vincolato alla sedia.

Il sistema è dotato di sei gradi di libertà ed è in grado eseguire i movimenti con elevata accuratezza: l'errore massimo di posizionamento per i movimenti traslatori è di 0.3 mm e di 0.1° per quelli rotatori [Figura 3.12]. Inoltre, gli spostamenti risultano particolarmente fluidi in quanto il sistema si muove su di un cuscino di aria compressa che viene generato tra il braccio e la piattaforma di granito sulla quale si muove.

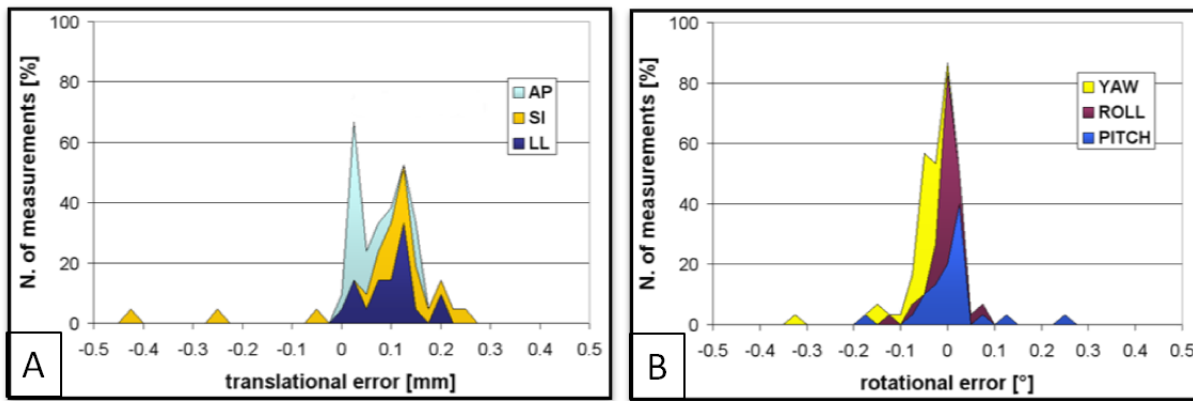


Figura 3.12: Distribuzione degli errori di posizionamento traslazionali e rotazionali del PVS installato presso il CNAO [45].

Il sistema di controllo comprende un'interfaccia per l'utente sia in sala di trattamento sia nella sala di controllo locale e consente il movimento automatizzato del braccio robotico sulla base delle informazioni contenute nel piano di trattamento o fornite da PVS o OTS.

In ultimo, vi è integrato un sistema che, qualora venisse rilevata una collisione durante il movimento, ne interrompe automaticamente l'esecuzione.

L'iter del trattamento presso CNAO prevede che il paziente venga immobilizzato sul lettino (o sulla sedia nel caso di lesioni al distretto testa collo) in apposite stanze esterne a quella di trattamento. Successivamente, uno o due operatori accompagnano il paziente in sala di trattamento, dove il braccio robotico, automaticamente, procede all'ancoraggio del lettino. A questo punto il sistema di trasporto viene rimosso e il paziente viene allineato al fascio, nella posizione nominale di trattamento. Al termine della sessione, il braccio robotico riporta il lettino in una posizione predefinita di sgancio, viene accoppiato il sistema di trasporto e il lettino viene sganciato e riportato nelle sale di preparazione, concludendo di fatto la seduta di trattamento.

Il sistema di verifica della posizione del paziente (PVS) è un sistema di acquisizione di immagini RX costituito da due coppie sorgente-dettettore, in cui la sorgente consiste in un tubo radiogeno operante ad una tensione di 150 kV ed il dettore in un pannello in silicio amorfo con risoluzione di 2232x3200 pixel. Il sistema è alloggiato in un cilindro posizionato verticalmente sopra l'isocentro della sala [Figura 3.13], capace di ruotare attorno all'asse verticale di 180° con un errore di posizionamento massimo di 0.1°. Al momento dell'utilizzo, i tubi radiogeni e i detettori vengono estratti dal cilindro e posizionati con un massimo errore traslazionale di 0.15 mm.

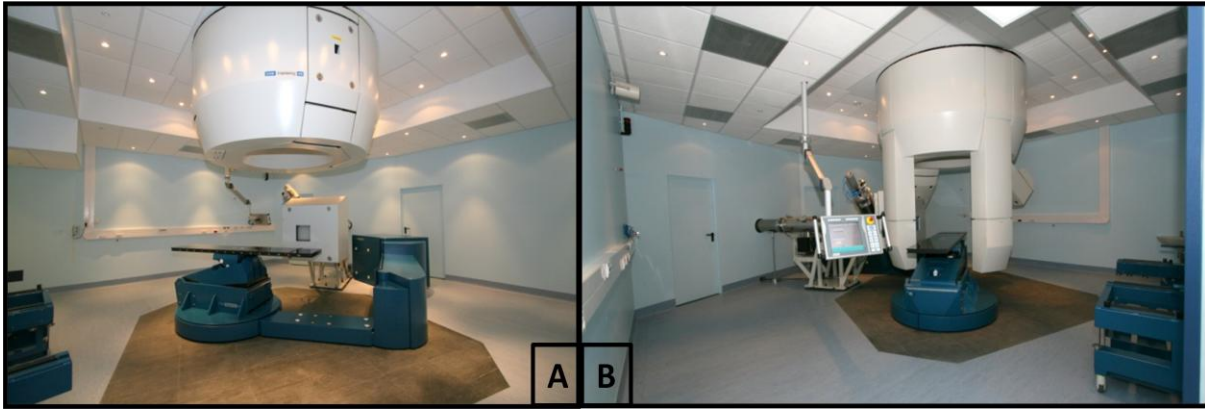


Figura 3.13: PVS. Pannello A: i tubi radiogeni e i detettori sono ritratti nel cilindro di supporto. Pannello B: i tubi radiogeni e i detettori sono estratti per l'utilizzo.

Le due coppie sorgente-detettore possono essere utilizzate sia per l'acquisizione simultanea di due immagini radiografiche sia per l'acquisizione di proiezioni singole.

Il PVS è completato da un software responsabile dell'acquisizione e della elaborazione delle immagini RX. Tale software implementa la generazione automatica delle DRR (*Digital Reconstructed Radiography*) dal piano di trattamento, in accordo con l'angolo dal quale sono state acquisite le immagini di controllo. Tali immagini vengono confrontate con le DRR tramite algoritmi di registrazione, in modo tale da fornire il vettore di correzione da applicare alla posizione del paziente (tramite il PPS) per portarlo in posizione nominale.

In ultimo, grazie alla capacità del cilindro di ruotare attorno all'asse verticale, le coppie sorgente-detettore del PVS possono essere potenzialmente utilizzate anche per l'acquisizione di immagini volumetriche *in-room*, sia con tecniche CBCT sia con tecniche DTS.

Il sistema hardware del localizzatore ottico (OTS) è basato sul sistema commerciale di *motion capture* SMART-D (BTS Spa) e consiste in tre telecamere IR [Figura 3.14] digitali, aventi una risoluzione 512x512 pixel ed operanti ad una frequenza di 70 Hz. L'errore massimo di localizzazione è inferiore a 0.5 mm in un volume di lavoro di 1 m³.

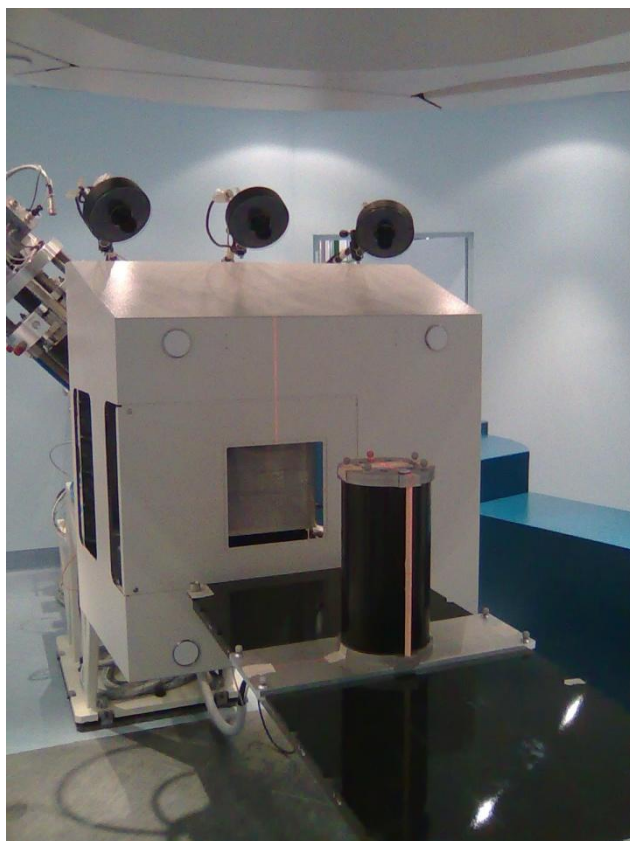


Figura 3.14: OTS installato presso CNAO. Sono visibili le tre telecamere del sistema SMART-D.

Il sistema viene utilizzato sia per il controllo in tempo reale della posizione del paziente durante l'intero trattamento, sia per il calcolo del vettore di correzione per il posizionamento preliminare del paziente. Tali correzioni possono essere calcolate utilizzando algoritmi basati sia su punti sia su superfici: nel primo caso il sistema ricostruisce la posizione spaziale dei marcatori riflettenti posti sulla superficie corporea del paziente, mentre, nel secondo caso, vengono acquisiti spot laser proiettati sul corpo del paziente al fine di ottenere una *patch* di superficie.

In ultimo, l'OTS consente l'eventuale implementazione di sistemi di *gating* respiratorio e di *tracking*.

Il software di controllo del localizzatore è stato ideato e sviluppato dal gruppo di Bioingegneria del Politecnico di Milano.

3.4. Localizzazione dinamica del bersaglio

Nei paragrafi precedenti sono state esposte le principali problematiche che riguardano l'accuratezza di un trattamento di adroterapia, identificando nelle tecniche di IGHT una possibile soluzione volta alla minimizzazione degli errori di set-up. Oltre a tali errori, anche le

inaccuratezze introdotte dall'*organ motion* inter-frazione vengono significativamente ridotte grazie a queste tecniche. Tali fenomeni, infatti, si manifestano, come già detto, in periodi di tempo dell'ordine dei giorni e dunque, una volta che il loro effetto è stato compensato all'inizio della frazione di trattamento, non introducono ulteriori incertezze sulla localizzazione del bersaglio.

Al contrario, l'*organ motion* intra-frazione, manifestandosi su periodi dell'ordine dei secondi o dei minuti, causa incertezze che si manifestano durante l'irradiazione e che non possono essere compensate con sistemi che agiscono soltanto all'inizio della frazione, come nel caso tecniche IGHT.

Per limitare tale fenomeno sono state introdotte diverse soluzioni, come ad esempio il *gating* respiratorio, che consiste nel monitorare il ciclo respiratorio del paziente e far sì che il fascio radiante sia attivo solo durante una determinata fase del ciclo e disattivo negli altri istanti. Tale tecnica si basa sull'assunzione che il ciclo respiratorio del paziente non subisca significative variazioni durante il trattamento e che quindi, all'interno di un determinato intervallo di tale ciclo, la posizione della lesione sia nota e sia sempre la medesima per l'intera durata della frazione.

Un'ulteriore soluzione alla problematica precedentemente esposta consiste nello sviluppo di sistemi di *tracking*, ovvero sistemi volti ad inseguire il movimento del bersaglio, al fine di garantire che il fascio radiante sia sempre centrato su di esso. I due paragrafi seguenti esporranno lo stato dell'arte relativo al *tracking* sia in radioterapia con fotoni sia in adroterapia.

3.4.1. Il sistema *Synchrony*TM per *CyberKnife*.

Ad oggi, l'unico sistema di *tracking* ampiamente utilizzato nella pratica clinica è il modulo *Synchrony*TM *Respiratory Tracking System* (RTS) per il sistema di trattamento robotico *Cyberknife* [46], che sarà presentato di seguito.

Il sistema *CyberKnife* [47][Figura 3.15] è un dispositivo di trattamento che consiste in un robot industriale a sei gradi di libertà sul quale è montato un acceleratore lineare compatto. Un sistema di *imaging* a raggi X, costituito da due tubi radiogeni e da due detettori al silicio amorfo, permette di ottenere due immagini ortogonali del paziente direttamente in sala di trattamento.



Figura 3.15: Sistema *CyberKnife*, con evidenziati i componenti principali.

Il sistema *Synchrony*, oltre ai dispositivi elencati in precedenza, utilizza un sistema optoelettronico per misurare in tempo reale gli spostamenti indotti dalla respirazione a cui è soggetta la superficie corporea del paziente. Tale sistema è costituito da tre emettitori ad infrarossi posizionati, tramite una particolare maglietta elastica, sul torace o sull'addome del paziente e da una serie di rilevatori IR sospesi al soffitto, la cui funzione è quella di registrare istante per istante la posizione spaziale degli emettitori [48].

Tutte le componenti del sistema *Synchrony* sono rappresentate in Figura 3.16.

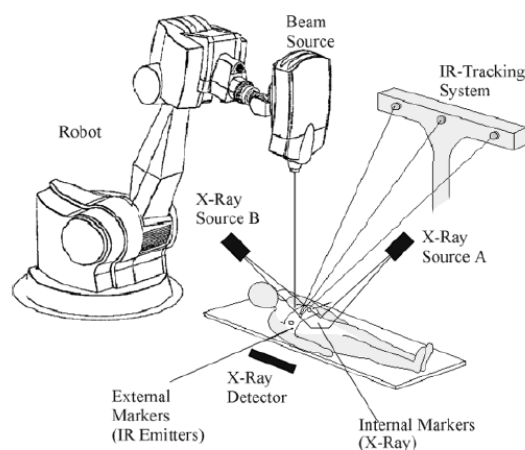


Figura 3.16: Rappresentazione schematica delle principali componenti del sistema *Synchrony*.

Il funzionamento del *Synchrony* RTS si basa sull'idea generale di stimare istante per istante la posizione tridimensionale del bersaglio utilizzando un modello di corrispondenza che correla il movimento esterno degli emettitori IR con il movimento interno della lesione.

Verrà di seguito descritta la procedura che viene seguita durante un tipico trattamento tramite *Synchrony*.

Prima dell'inizio dell'irradiazione il sistema di *imaging* integrato nel *Cyberknife* acquisisce una serie di immagini in rapida successione, in modo tale da campionare adeguatamente il ciclo respiratorio del paziente. In ogni immagine viene localizzata la posizione dei marcatori impiantati all'interno della lesione; così facendo si ottiene la traiettoria sulla quale si muove il bersaglio per l'effetto dell'*organ motion* intra-frazione. Contemporaneamente all'acquisizione delle immagini RX, il sistema optoelettronico rileva ad alta frequenza (circa 25 Hz) la posizione dei tre emettitori collocati sulla maglietta elastica. In questo modo il sistema dispone di informazioni dettagliate e sincronizzate temporalmente sul movimento indotto dalla respirazione sia sul *target* sia sulla superficie corporea.

Note tali informazioni, il software calcola un modello di corrispondenza che mette in relazione il movimento del bersaglio con quello degli emettitori. Il sistema permette di scegliere tra due modelli matematici diversi [49]: uno lineare e uno polinomiale del secondo ordine.

Durante l'irradiazione, una componente software denominata "*modeler*" provvede ad utilizzare il modello precedentemente calcolato per stimare la posizione del bersaglio in ogni istante.

Dal momento che il modello necessita di essere aggiornato per evitare che la sua precisione diminuisca nel tempo, immagini RX vengono acquisite periodicamente anche durante il trattamento e, sulla base delle nuove informazioni ottenute sulla posizione della lesione, il modello viene aggiornato utilizzando un metodo *first in first out*, ovvero mantenendo costante il numero di punti sui quali calcolare la correlazione e, una volta ottenuto un nuovo punto (tramite *imaging*), scartando il più vecchio.

Se si presenta la necessità, l'irradiazione viene interrotta e il modello viene calcolato nuovamente acquisendo un nuovo set di immagini.

3.4.2. Sistemi di *tracking* in adroterapia

I sistemi di *tracking* nascono come soluzioni volte a limitare il più possibile le inaccuranze introdotte dall'*organ motion* intra-frazione e hanno come scopo quello di far sì che il fascio radiante sia sempre collimato sul bersaglio, nonostante i suoi movimenti: a tal fine il fascio deve inseguire il *target* lungo la sua traiettoria.

Dal punto di vista tecnologico, per realizzare un sistema di *tracking* è quindi necessario disporre di due componenti: un sistema di erogazione della dose in cui il fascio radiante non sia fisso ma possa essere direzionato e un sistema in grado di fornire in tempo reale la posizione (nelle tre dimensioni) della lesione da irradiare.

Per quanto riguarda il sistema di erogazione del fascio, è evidente che sia gli acceleratori lineari a geometria isocentrica (per quanto riguarda la radioterapia) sia i sistemi passivi basati su

modulatori di energia e scatteratori (per quanto riguarda l'adroterapia) non permettono di inseguire efficacemente il movimento del *target*. Con tali sistemi, infatti, seppure sia possibile muovere il fascio in direzione trasversale tramite l'utilizzo di collimatori multi-lamellari, tale movimento risulta limitato in ampiezza e, di conseguenza, non adeguato ad inseguire un bersaglio lungo la sua intera traiettoria. Questo, invece, è possibile se si ricorre ad acceleratori lineari robotizzati (in radioterapia) e, in adroterapia, a sistemi attivi di deposizione della dose.

Il sistema che fornisce la posizione del bersaglio si deve basare su un modello in grado di ricavare tale informazione basandosi su un segnale surrogato, quale la posizione di marcatori posti sulla superficie corporea del paziente.

Nell'ambito dell'adroterapia, il *tracking* dinamico è al momento solo oggetto di ricerca e non trova applicazioni cliniche nei centri che praticano questa terapia. Esso può essere effettuato solo con sistemi attivi, pertanto viene chiamato *active scanning* dinamico. Un dispositivo di questo genere è stato sviluppato presso il centro GSI, dove è in fase sperimentale [50,51]. Lo schema del prototipo realizzato è rappresentato in Figura 3.17 e descritto brevemente di seguito.

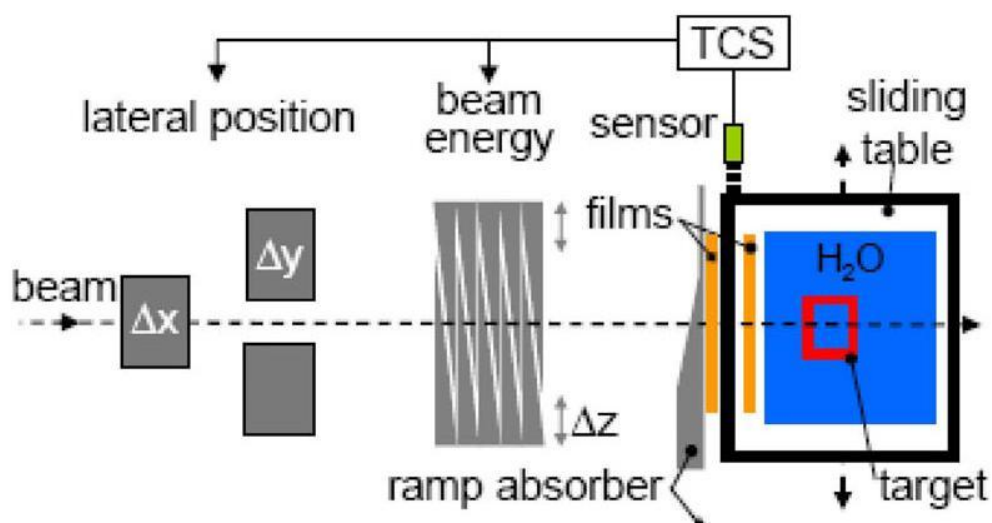


Figura 3.17: Set-up sperimentale del sistema di *tracking* realizzato presso il GSI.

Il bersaglio, immerso in acqua, è collocato su di una apposita piattaforma mobile, mentre per simulare le variazioni di *range* dovute al movimento del *target*, è stato inserito un filtro assorbitore a rampa fisso, così che il fascio, quando viene spostato in sul piano trasversale, attraversi uno spessore diverso e abbia quindi un diverso *range*. Le coordinate tridimensionali del bersaglio vengono calcolate da un sensore laser tramite triangolazione ed inviate al sistema di controllo del trattamento (TCS). Tale sistema, al fine di inseguire il movimento a cui il *target*

è sottoposto, agisce sui magneti di *steering* per orientare il fascio sul piano trasversale e su assorbitore a cuneo per modularne l'energia.

Utilizzando questo prototipo sono stati ottenuti risultati soddisfacenti sia dal punto di vista della geometria della deposizione di dose [50] sia dal punto di vista dosimetrico [51].

3.5. Scopo del progetto

Come già visto nei capitoli 2, l'accuratezza della localizzazione del bersaglio e della deposizione della dose su di esso risultano di estrema importanza al fine di controllare efficacemente la malattia limitando, nel contempo, il danneggiamento di tessuti sani.

Il fenomeno dell'*organ motion*, però, fa sì che la posizione dei tumori extra-cranici sia soggetta a variazioni sia a lungo termine (in alcuni giorni) sia a breve termine (in alcuni secondi), introducendo un'indeterminazione la cui entità è tale da compromettere l'efficacia della terapia.

Nell'ambito dei trattamenti adroterapici, i sistemi di IGHT (paragrafo 3.3) garantiscono un'accurata localizzazione del bersaglio immediatamente prima dell'inizio dell'irradiazione, limitando l'incertezza introdotta da movimenti su lunghi intervalli temporali (*organ motion* inter-frazione). Se, però, il bersaglio è soggetto a movimenti in un ridotto intervallo di tempo (*organ motion* intra-frazione), come ad esempio nel caso di lesioni collocate in prossimità di strutture che risentono fortemente del movimento respiratorio, la sua posizione varia durante l'irradiazione e tale variazione non è compensata dalle tecniche IGHT, dato che queste intervengono soltanto prima dell'irradiazione e non durante. Tale problematica risulta ancora più rilevante nel caso in cui si utilizzino sistemi di deposito della dose basati sull'*active scanning*, nei quali la selettività geometrica è più elevata di quella ottenuta con sistemi passivi.

In generale, dunque, l'*organ motion* intra-frazione limita la possibilità di trattare efficacemente con fasci di particelle le lesioni toraciche o addominali.

Per superare tale limitazione, risulta evidente la necessità di introdurre sistemi e tecnologie in grado di mitigare l'effetto dell'*organ motion* intra-frazione sull'accuratezza con la quale la dose viene depositata nel volume bersaglio. A tal fine, una possibile soluzione consiste nella realizzazione di un sistema di *tracking* (paragrafo 3.4), ovvero di un sistema nel quale il fascio radiante insegue il bersaglio lungo la sua traiettoria. Come già detto, sistemi di questo tipo sono, ad oggi, in fase sperimentale presso il GSI [50,51], dove è stato sviluppato un prototipo in cui l'energia e la direzione del fascio vengono variate per colpire un bersaglio in movimento la cui posizione, però, è esattamente nota, essendo misurata con un sensore laser.

In una generica applicazione clinica di un sistema di *tracking* in adroterapia è sì fondamentale disporre di un sistema che varia energia e direzione del fascio in accordo ai movimenti del

bersaglio, ma, prima di tutto, è necessario misurare o stimare accuratamente ed in tempo reale la sua posizione spaziale all'interno del corpo del paziente.

Il requisito di una misurazione in tempo reale, però, impedisce l'utilizzo di metodi basati sull'acquisizione in continua di immagini radiografiche, dato che ciò comporterebbe un sovradosaggio di entità intollerabile per la sicurezza del paziente.

È quindi necessario stimare la posizione del bersaglio sfruttando segnali facilmente misurabili il cui andamento temporale si correla con il movimento del *target*, come ad esempio la posizione, rilevata tramite un sistema ottico, di marcatori posti sulla superficie corporea del paziente.

Qualora si abbia a disposizione un tale segnale, è possibile stimare, istante per istante, la posizione interna della lesione da irradiare. Per far ciò è necessario un apposito software che correli tale posizione interna al segnale esterno misurato.

La correlazione deve essere stabilita all'inizio della terapia, acquisendo simultaneamente la posizione dei marcatori esterni (attraverso un sistema di localizzazione ottica) e della lesione interna (attraverso immagini RX in rapida successione) e deve essere aggiornata nel corso del trattamento tramite ulteriori immagini radiografiche di controllo. Il generico funzionamento del software appena descritto è mostrato in Figura 3.18.

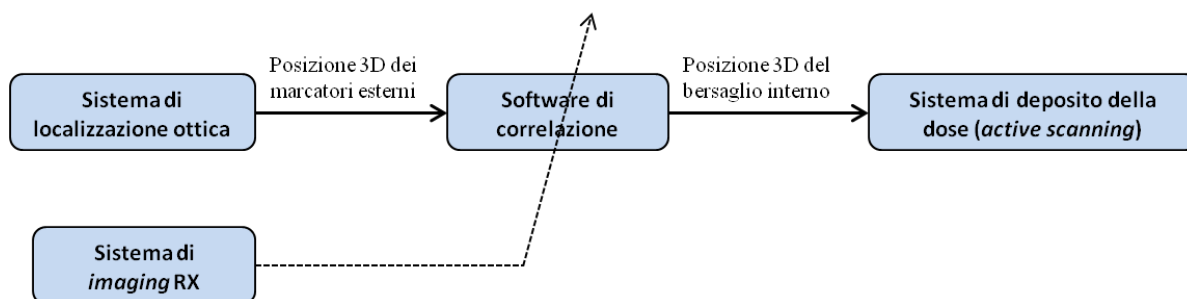


Figura 3.18: Schema a blocchi di un generico software di correlazione interno/esterno per un sistema *tracking*.

Lo scopo del progetto che verrà presentato consiste appunto nello sviluppo di metodi software in grado di stimare la posizione di una lesione interna a partire da un segnale che descrive il movimento della superficie corporea del soggetto. Le principali funzionalità che sono richieste agli strumenti implementati sono:

- Acquisire in tempo reale la posizione di marcatori esterni (segnale esterno).
- Acquisire la posizione “vera” del bersaglio (segnale interno), negli istanti in cui vengono ottenute le immagini RX.
- Apprendere la correlazione tra il segnale interno e quello esterno.
- Utilizzare la correlazione appresa per stimare il segnale interno in tempo reale.

- Confrontare la posizione interna stimata con quella “vera” ottenuta in corrispondenza dalle immagini RX di controllo.
- Se necessario, adattare la correlazione a fronte dei nuovi dati resi disponibili dalle immagini di controllo.

L'apprendimento della correlazione interno/esterno costituisce, come è evidente, la problematica centrale dell'intero lavoro e verrà affrontata ricorrendo a strumenti computazionali quali le reti neurali e i modelli di stato, che saranno descritti singolarmente nel capitolo 4.

In ultimo, il lavoro proposto comprende anche la validazione del suddetto software.

Il raggiungimento degli obiettivi appena citati renderebbe disponibile un prototipo che implementa tutte le principali funzionalità di un software di correlazione interno/esterno e, come tale, costituisce un punto di partenza per lo sviluppo di un modulo di *tracking* completamente integrabile nei sistemi di deposito della dose installati presso CNAO. Inoltre, tali sistemi, essendo basati sull'*active scannig*, risultano una piattaforma ideale per un sistema di inseguimento del bersaglio.

In conclusione, il prototipo che si vuole realizzare costituirà un valido strumento nell'ambito del trattamento di lesioni mobili, sia perché consentirà di disporre istante per istante della posizione interna del *target*, sia perché, nel momento in cui presso CNAO verranno trattate lesioni extra-craniche soggette ad *organ-motion* intra-frazione, potrà costituire un ausilio per lo sviluppo di un sistema di *tracking*.

4. Strumenti utilizzati

Nel paragrafo 3.5 sono stati esposti i principali obiettivi del lavoro sul quale verte questo elaborato. Prima di descrivere come tale lavoro è stato sviluppato, si ritiene utile fornire una panoramica sugli strumenti teorici e computazionali che sono stati utilizzati, ovvero le reti neurali e i modelli a variabili di stato. Verrà inoltre descritto il database dal quale sono stati prelevati i dati necessari allo sviluppo e alla validazione del lavoro proposto.

4.1. Reti neurali

4.1.1. Introduzione

L'origine delle reti neurali artificiali può essere ricondotta alla constatazione della superiorità del cervello umano, rispetto a qualsiasi calcolatore, nella soluzione di problemi non immediatamente riconducibili a calcoli aritmetici. Tale constatazione portò alla nascita del “connessionismo”, ovvero una branca delle scienze cognitive che ha come scopo quello di modellizzare il funzionamento della mente umana basandosi su una descrizione del cervello come costituito da reti di neuroni interconnesse [52].

In questo modo sono state poste le basi per lo sviluppo di algoritmi computazionali basati su processi tipici del sistema nervoso centrale, detti appunto “reti neurali artificiali” (ANN). Ad oggi, l'utilizzo delle ANN si è decisamente allontanato dagli studi di neurofisiologia per i quali erano state proposte e comprende la soluzione di svariati problemi, quali la predizione di un'uscita a fronte di un determinato ingresso, la classificazione di oggetti, l'approssimazione di funzioni matematiche, il riconoscimento di schemi all'interno di una serie di dati e il completamento di modelli conosciuti.

A seguito della loro versatilità, le ANN hanno trovato applicazione, oltre che negli studi riguardanti l'intelligenza artificiale, anche in ambito medico, dove possono essere sfruttate, ad esempio, per l'estrazione di informazione da una serie di dati grezzi o come strumento di supporto alla decisione [53].

Da un punto di vista teorico, le ANN vengono definite come uno strumento computazionale basato su algoritmi non lineari in grado di approssimare la soluzione di un problema *black-box* apprendendo da esempi. A partire da tale definizione si può immediatamente identificare sia la principale caratteristica delle ANN sia le loro possibili applicazioni in ambito ingegneristico.

Come risulta chiaro, le ANN si differenziano da ogni altro algoritmo proprio per la capacità di apprendere da esempi. La maggior parte degli strumenti computazionali, infatti, si basa su

processi di tipo deduttivo, in cui da una regola di valore generale vengono ricavati esempi particolari. Le ANN, invece, utilizzano processi di tipo induttivo, ovvero sono in grado di apprendere, a partire da un set di esempi, la regola che governa il sistema che li ha generati. La regola appresa può essere quindi sfruttata per il calcolo dell'uscita del sistema a fronte di ingressi noti.

Questa potenzialità risulta particolarmente utile, ad esempio, per la soluzione di problemi che non possono essere risolti in maniera deduttiva, come ad esempio quelli in cui il sistema del quale si vuole studiare il comportamento non può essere descritto da un modello matematico oppure quando tale modello, pur esistendo, è troppo complesso per essere implementato efficacemente. Questa tipologia di problemi viene definita "a scatola nera" (*black box*).

In ultimo si noti che l'utilizzo di metodi induttivi ha come risultato una soluzione approssimata del problema e mai una soluzione esatta, come invece nel caso di problemi risolti con metodi deduttivi.

In Figura 4.1 è proposto un confronto tra la modalità di soluzione di un problema in maniera deduttiva e la modalità di soluzione di un problema *black-box* tramite ANN.

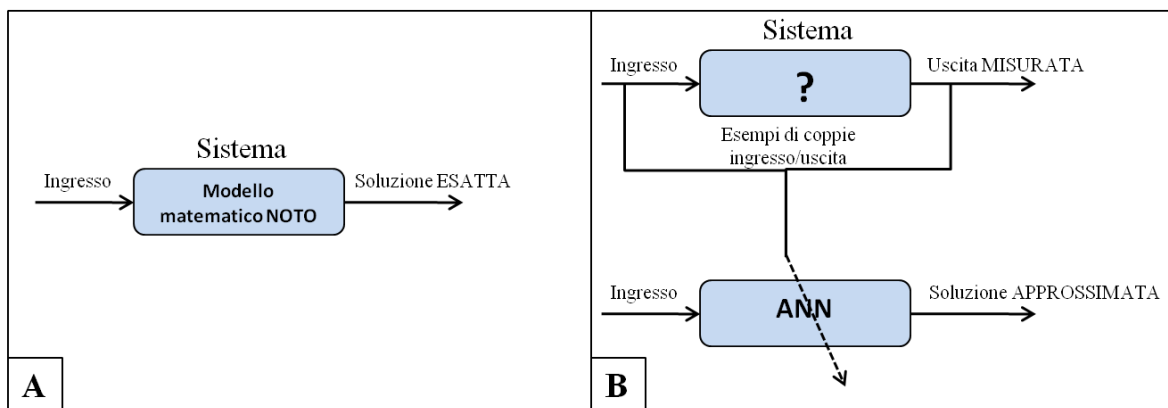


Figura 4.1: Pannello A: soluzione di un problema in maniera deduttiva. Pannello B: soluzione di un problema a scatola nera tramite ANN.

Come già detto, le ANN si ispirano al funzionamento del sistema nervoso centrale, la cui unità fondamentale è il neurone [Figura 4.2].

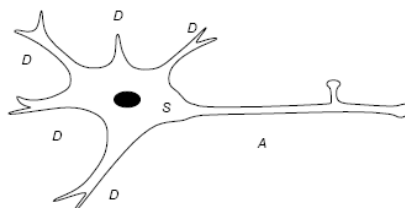


Figura 4.2: Rappresentazione schematica di un neurone. D: dendriti, S: Soma, A: Assone.

Tale cellula è caratterizzata da connessioni ramificate dette dendriti, grazie alle quali il nucleo (o soma) riceve gli stimoli provenienti da altri neuroni. A seguito degli stimoli ricevuti, il nucleo risponde con un stimolo che viene trasmesso attraverso l'assone verso altri neuroni.

È proprio sulla base della struttura di un neurone naturale che, nel 1943, Mc Culloch e Pitts proposero un modello di neurone [Figura 4.3] che calcola la somma pesata degli ingressi e produce un'uscita binaria (0 o 1) a seconda che tale somma superi o meno un valore di soglia.

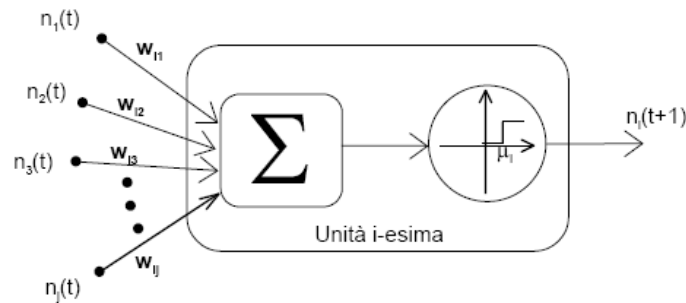


Figura 4.3: Modello di neurone di Mc Culloch e Pitts (MCP).

L'equazione che regola l'uscita di tale modello è la seguente:

$$n_i(t + 1) = \Theta[\sum_j w_{ij}n_j(t) - \mu_i] \quad (4.1)$$

dove η_i rappresenta l'uscita del neurone i-esimo, η_j gli ingressi (cioè le uscite dei neuroni precedenti), w_{ij} il peso della connessione tra l'ingresso j-esimo e il neurone i-esimo e μ_i la soglia di attivazione del neurone i-esimo. Infine Θ rappresenta la funzione di attivazione del neurone, che coincide con la funzione di Heaviside, così definita:

$$\Theta(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \quad (4.2)$$

Il neurone MCP fu sfruttato da Rosenblatt [54] per la progettazione di una vera e propria rete neurale, cioè uno strumento di calcolo costituito dall'interconnessione di più unità fondamentali (neuroni) e in grado di fornire, a fronte di un determinato ingresso, una uscita adeguata. La rete neurale proposta da Rosenblatt, chiamata "perceptrone semplice" [Figura 4.4], era composta da un singolo strato di neuroni MCP con un numero variabile di ingressi. Il numero di uscite, invece, coincide necessariamente con il numero di neuroni che costituiscono lo strato.

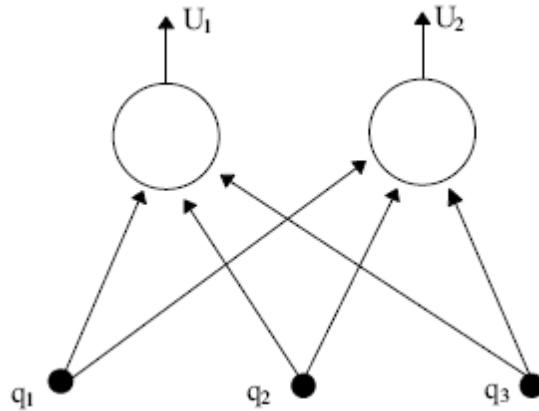


Figura 4.4: Schema di un perceptrone semplice avente 3 ingressi e 2 uscite.

Ogni uscita viene calcolata analogamente a come descritto per il neurone MCP, osservando che l'equazione 4.1 può essere generalizzata nel seguente modo:

$$U_i = g[\sum_j w_{ij}q_j] \quad (4.3)$$

dove $g(\cdot)$ rappresenta una generica funzione di attivazione (non necessariamente la funzione di Heaviside) e la soglia μ_i non è esplicitata in quanto il suo effetto può essere ottenuto aggiungendo un ingresso alla rete fittizio costante e pari ad 1 ($q_0 = 1$), al quale sono associati i pesi w_{i0} anch'essi costanti e pari all'opposto del valore di soglia ($w_{i0} = -\mu_i$).

Un caso particolare di perceptrone semplice è quello in cui si utilizzano unità a soglia [55], ovvero neuroni la cui funzione di attivazione coincide con la funzione segno, definita nel seguente modo:

$$\text{sgn}(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ -1 & \text{se } x < 0 \end{cases} \quad (4.4)$$

In questo caso l'equazione 4.3 può essere riscritta nel seguente modo (utilizzando una notazione vettoriale):

$$U_i = \text{sgn}[\mathbf{w} \cdot \mathbf{q}] \quad (4.5)$$

Si osserva che la proiezione del vettore dei pesi (\mathbf{w}) sul vettore degli ingressi (\mathbf{q}) deve avere lo stesso segno dell'uscita desiderata. Dunque, affinché la rete sia in grado di fornire uscite adeguate, deve necessariamente esistere un iperpiano che separi lo spazio delle uscite positive da quello delle uscite negative. Tale condizione, detta condizione di separabilità lineare, impone l'esistenza di un vettore di pesi $\bar{\mathbf{w}}$ che soddisfi la seguente relazione:

$$\bar{\mathbf{w}} \cdot \mathbf{q} = 0 \quad (4.6)$$

È quindi chiaro che questo tipo di perceptrone è in grado di risolvere soltanto problemi lineari, mentre non è in grado di apprendere la soluzione di problemi in cui lo spazio delle uscite positive e quello delle uscite negative sono separabili soltanto da curve non lineari.

La limitazione appena discussa fu superata nel 1974 da Werbos [56], il quale propose una rete neurale in cui, tra gli ingressi e le uscite, vi fosse più di uno strato di neuroni. Tale rete, chiamata appunto “perceptrone multi strato” (MLP, *Multi Layers Perceptron*), è costituita da uno strato d’uscita (formato da uno o più neuroni) e da un numero variabile di strati detti “nascosti”, in quanto non direttamente connessi all’uscita della rete. Anche gli strati nascosti possono essere formati da un numero variabile di neuroni. Per quanto riguarda la descrizione matematica delle reti MLP, essendo queste formate da unità riconducibili al perceptrone semplice, vale, per ogni neurone, la relazione 4.3.

Due esempi di MLP sono presentati in Figura 4.5.

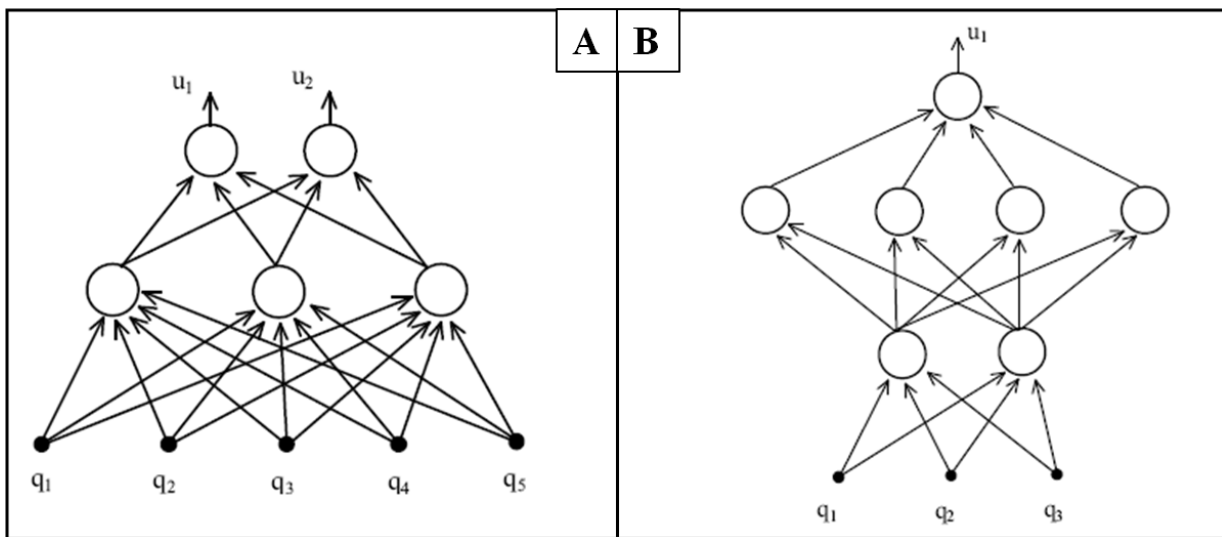


Figura 4.5. Pannello A: MLP con 3 neuroni nello strato nascosto e 2 neuroni nello strato d’uscita. Pannello B: MLP con due strati nascosti (da 2 e da 3 neuroni) e un neurone nello strato d’uscita.

L’introduzione degli strati nascosti, oltre che permettere la soluzione di problemi non lineari, permette lo sviluppo di reti le cui architetture possono essere le più svariate. A tal proposito si è soliti dividere le ANN in due grandi categorie: le reti *feed forward* e le reti retroazionate. Queste ultime, a differenza delle prime, sono caratterizzate dalla presenza di retroazioni, ovvero il segnale in uscita della rete viene fornito come un ulteriore ingresso dopo un ritardo temporale τ . In tal modo possono essere realizzati uno o più anelli di retrazione.

A titolo di esempio, entrambe le reti mostrate in Figura 4.5 sono *feed forward*, mentre in Figura 4.6 è riportata una rete con singolo anello di retroazione.

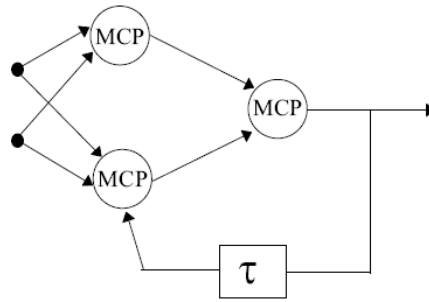


Figura 4.6: Esempio di rete retroazionata.

In tutta la trattazione seguente ci si riferirà soltanto alle reti *feed forward*.

4.1.2. Procedure di addestramento

Per addestramento si intende la procedura con la quale i pesi che caratterizzano le connessioni tra neuroni vengono aggiornati iterativamente (a partire da un valore arbitrario) in modo tale che la rete sia in grado di risolvere il problema in questione, ovvero sia in grado di fornire uscite adeguate a fronte di determinati ingressi.

La procedura di addestramento avviene, in generale, a partire da un set di esempi detto *training set*, e consiste nell'apprendimento di tale set da parte della rete. Da un punto di vista pratico, l'apprendimento si traduce nella modulazione dei pesi fino al raggiungimento di una configurazione ottimale.

A seconda della composizione del *training set* si possono distinguere due tipi di addestramento [55]:

- Addestramento supervisionato: il *training set* è composto da coppie ingresso/uscita. Fornendo alla rete una certa configurazione di ingressi e calcolando l'uscita che questa produce, è possibile valutare la differenza tra tale uscita e quella desiderata e, in base a tale errore, aggiornare opportunamente i pesi della rete.
- Addestramento non supervisionato: in questo caso non si ha disposizione un *training set* formato da coppie ingresso/uscita, ma si ha soltanto una serie di ingressi. La rete stessa deve estrarre, dai dati fornitigli in ingresso, caratteristiche o comportamenti secondo i quali classificarli.

Di seguito ci si riferirà soltanto ad addestramenti supervisionati, andando ad analizzare dapprima l'algoritmo di addestramento di un perceptrone semplice con funzione di attivazione "segno", per poi trattare l'algoritmo utilizzato nell'addestramento di un rete MLP, noto come "*back propagation*".

Addestramento di un perceptrone semplice

Si consideri un perceptrone semplice con funzione di attivazione “segno” e si ipotizzi che il problema che si vuole risolvere ammetta una soluzione lineare (condizione di separabilità lineare).

Si ricorda che l'uscita dell' i -esimo neurone è data dalla seguente equazione:

$$U_i^\mu = \text{sgn}[\sum_j w_{ij} q_j^\mu] \quad (4.7)$$

dove l'apice μ indica che si sta considerando la μ -esima configurazione di ingressi/uscite contenuta nel *training set*.

Poiché si vuole che le uscite delle neuroni coincidano con le uscite desiderate (indicate dal pedice d), si vuole minimizzare la seguente cifra di merito J :

$$J = \|U_{di}^\mu - U_i^\mu\| = \|U_{di}^\mu - \text{sgn}[\sum_j w_{ij} q_j^\mu]\| = \|\varepsilon_i^\mu\| \quad (4.8)$$

Dal momento che l'uscita della rete può valere soltanto ± 1 , la quantità ε può assumere soli i seguenti valori:

$$\varepsilon_i^\mu = \begin{cases} 0 & \text{se } U_{di}^\mu = U_i^\mu \\ \pm 2 & \text{se } U_{di}^\mu \neq U_i^\mu \end{cases} \quad (4.9)$$

La variazione dei pesi Δw_{ij} dovrà far sì che la quantità ε tenda a zero, dunque dovrà valere:

$$\Delta w_{ij}^\mu = \eta \varepsilon_i^\mu q_j^\mu \quad (4.10)$$

dove η rappresenta un fattore moltiplicativo > 0 detto *learning rate*, in quanto il suo valore modula la velocità con la quale, ad ogni interazione, la rete apprende l'informazione contenuta nel *training set*.

Infine, ad ogni iterazione (cioè per ogni μ), i pesi aggiornati vengono calcolati nel seguente modo:

$$w_{ij}^{\text{new}} = w_{ij}^{\text{old}} + \Delta w_{ij}^\mu \quad (4.11)$$

A titolo di esempio, si consideri il caso in cui si abbia $\varepsilon = + 2$, cioè quando l'uscita desiderata vale $+ 1$ mentre l'uscita della rete vale -1 . In tale condizione, la variazione dei pesi risulta positiva, il che tende a far crescere la somma pesata degli ingressi verso valori positivi, per i quali l'uscita della rete coinciderebbe con quella desiderata.

L'algoritmo di addestramento appena descritto prende il nome di “*delta rule*” ed è stato proposto da Widrow e Hoff nel 1960 [57].

Addestramento di un perceptrone multi strato

Come anticipato, per l'addestramento di un qualsiasi perceptrone multistrato si utilizza un algoritmo noto come “*back propagation*” [58]. Poiché tale algoritmo si basa sulla minimizzazione dell'errore tramite discesa del gradiente, è necessario ipotizzare che tutti i

neuroni siano caratterizzati da funzioni di attivazione continue e derivabili (ad esempio lineari o sigmoidei), escludendo quindi sia la funzione di Heaviside che la funzione segno.

Per prima cosa si definisca l'errore quadratico medio (MSE, *Mean Square Error*) nel seguente modo:

$$MSE = E(\mathbf{w}) = \frac{1}{P} \sum_{\mu=1}^P \sum_{i=1}^N (U_{di}^{\mu} - U_i^{\mu})^2 = \frac{1}{P} \sum_{\mu=1}^P \sum_{i=1}^N \left(U_{di}^{\mu} - g(\sum_j w_{ij} q_j) \right)^2 \quad (4.12)$$

dove P rappresenta il numero di configurazioni che costituiscono il *training set* e N il numero di uscite delle rete. Si osservi che l'MSE è funzione del vettore contenente i pesi della rete (\mathbf{w}), dunque applicare il metodo della discesa del gradiente significa calcolare la configurazione di pesi ottima (\mathbf{w}^o) che minimizzi l'MSE:

$$\mathbf{w}^o = \operatorname{argmin}_{\mathbf{w}}(E(\mathbf{w})) \quad (4.13)$$

Per fare ciò è necessario differenziare E rispetto al vettore dei pesi, ottenendo una correzione espressa nel seguente modo:

$$\Delta w_{ij}^{\mu} = -\eta \frac{\partial E_i^{\mu}}{\partial w_{ij}} = \eta (U_{di}^{\mu} - U_i^{\mu}) g'(\sum_j w_{ij} q_j) q_j^{\mu} \quad (4.14)$$

Si noti che la correzione dei pesi è modulata dal *learning rate* ed è proporzionale al gradiente di E .

Si consideri ora un generica rete *feed forward*, in cui gli ingressi, i pesi e le uscite dei neuroni siano identificati con la notazione riportata in Figura 4.7.

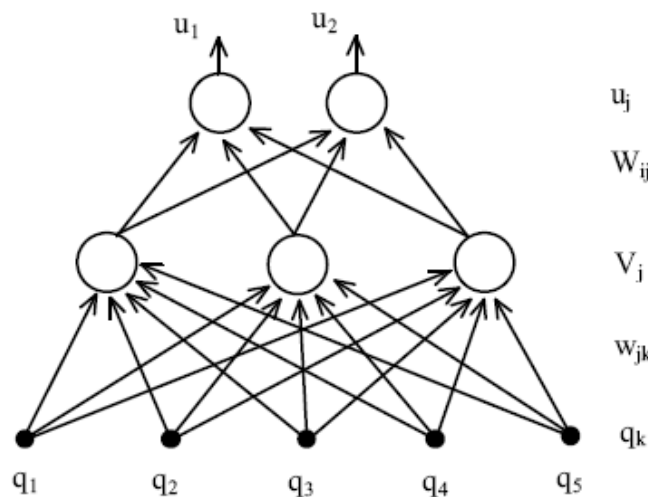


Figura 4.7: Generica rete multi strato *feed forward* con singolo strato nascosto.

Esattamente come visto nell'equazione 4.14, i pesi tra strato nascosto e strato d'uscita vengono calcolati come segue:

$$\Delta W_{ij} = \eta \sum_{\mu} (\delta_i^{\mu} V_j^{\mu}) \quad (4.15)$$

con:

$$\delta_i^\mu = (U_{di}^\mu - U_i^\mu)g'(\sum_j w_{ij}q_j) \quad (4.16)$$

Per quanto riguarda i pesi delle connessioni tra ingressi e strato nascosto, è necessario differenziare E rispetto ai pesi w_{jk} , ottenendo:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = \dots = \eta \sum_\mu (\delta_j^\mu q_k^\mu) \quad (4.17)$$

con:

$$\delta_j^\mu = g'(\sum_k w_{jk}q_k)(\sum_i w_{ij}\delta_i^\mu) \quad (4.18)$$

È ora possibile comprendere perché l'algoritmo appena descritto sia noto come “*back propagation*”. Si osservi, infatti, che l'errore sull'uscita della rete viene utilizzato per calcolare la quantità δ relativa allo strato d'uscita (equazione 4.16). Tale quantità, oltre che per aggiornare i pesi dello strato d'uscita (equazione 4.15), viene utilizzata per ricavare la δ relativa allo strato nascosto (equazione 4.18), grazie alla quale è possibile aggiornare i pesi di tale strato (equazione 4.17).

In generale, quindi, procedendo in questo modo fino allo strato direttamente collegato agli ingressi, si ottiene l'addestramento dell'intera rete.

Performance della rete

Dopo aver trattato gli algoritmi utilizzati per addestrare una rete neurale, si vuole discutere brevemente delle performance della rete e di come queste siano influenzate dalla procedura d'addestramento.

Da un punto di vista qualitativo la performance di una rete consiste nella capacità di quest'ultima di fornire delle uscite quanto più prossime possibili a quelle desiderate per l'ingresso fornito.

Quantitativamente, tale performance viene generalmente espressa in termini di *Root Mean Square Error* (RMS):

$$RMS = \sqrt{MSE} = \sqrt{\frac{1}{P} \sum_{\mu=1}^P \sum_{i=1}^N (U_{di}^\mu - U_i^\mu)^2} \quad (4.19)$$

Come l'MSE, anche l'RMS quantifica l'errore commesso dalla rete tenendo in considerazione tutte le sue uscite (1...N) e tutte le configurazioni del *training set* (1...P).

Da un punto di vista teorico, ad una rete addestrata si richiedono due capacità fondamentali: la prima consiste nell'aver appreso il *training set*, ovvero, per ogni configurazione di ingressi contenuta in tale set, la rete deve essere in grado di fornire un'uscita coincidente (o prossima) a quella desiderata. La seconda capacità, detta “capacità di generalizzazione”, consiste nel saper

fornire uscite adeguate anche a fronte di configurazioni di ingressi non contenute nel *training set* ed eventualmente molto differenti da quelle con le quali è stato condotto l'addestramento.

Si osserva, però, che un eccessivo sviluppo della prima capacità comporta una diminuzione della seconda, in quanto la rete, apprendendo con eccessiva precisione le configurazioni del *training set*, mostra errori elevati se in ingresso vengono fornite configurazioni mai incontrate prima.

Tale fenomeno, chiamato *overfitting*, tende a presentarsi quando alla rete viene proposto il *training set* per un numero eccessivo di iterazioni (dette "epoche").

Per valutare in termini di performance l'andamento dell'addestramento di una rete, si è soliti dividere i dati disponibili in tre set differenti:

- *Training set*: coppie ingresso/uscita proposte alla rete per l'aggiornamento dei pesi.
- *Validation set*: coppie ingresso/uscita utilizzate per valutare la performance della rete durante l'addestramento ma anche per l'aggiornamento dei pesi.
- *Testing set*: coppie ingresso/uscita che non vengono mai utilizzate per aggiornare i pesi, ma solo per valutare la performance.

Per ognuno dei tre set, si calcola l'andamento del RMS in funzione del numero di iterazioni e, in generale, si osserva il seguente andamento [Figura 4.8]:

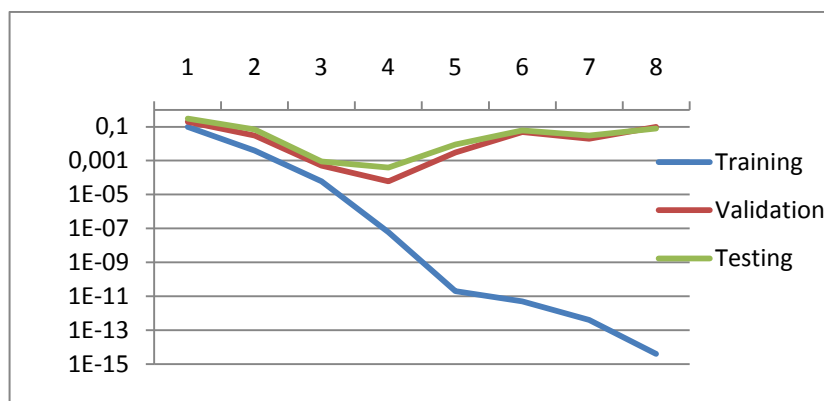


Figura 4.8: Andamento dell'RMS per il training, il *validation set* e il *testing set* [59].

Si nota che, all'aumentare delle iterazioni, l'RMS sul *training set* decresce continuamente, dato che la rete apprende tali configurazioni sempre più accuratamente. Al contrario, l'errore relativo ai set di validazione e *testing* decresce inizialmente per poi aumentare a causa dell'*overfitting*.

Dall'osservazione di quest'ultimo grafico è possibile stabilire il numero ottimale di epoche d'addestramento per ottenere la performance desiderata, che chiaramente dipende dal particolare problema che si vuole affrontare. Ad esempio, qualora si voglia ottenere una

performance che garantisca un buon compromesso tra apprendimento del *training set* e capacità di generalizzazione, allora il numero di epoche ottimale coincide con quello in cui si osserva il minimo dell'RMS per il *validation* e il *testing set*. Se, invece, la capacità di generalizzazione non è un requisito fondamentale per il problema da affrontare, questa può essere sacrificata per ottenere un apprendimento del *training set* più accurato: in tal caso è possibile stabilire il numero di epoche in modo che, al termine dell'addestramento, l'RMS sul *training set* risulti inferiore a ad un dato valore di soglia.

Condizioni di arresto dell'addestramento

Come si è appena accennato, le performance della rete dipendono fortemente dalle modalità con le quali questa viene addestrata ed, in particolare, dalle condizioni che vengono stabilite per terminare il ciclo di addestramento. Tali condizioni, in generale, possono riguardare sia il numero di interazioni eseguite sia la performance della rete e sono discusse in seguito.

- Numero di epoche. In questo caso è necessario stabilire un determinato numero di epoche, raggiunto il quale l'addestramento viene arrestato. Tale condizione ha il vantaggio di imporre un preciso termine all'addestramento ma, risultando indipendente dalla performance della rete, non garantisce che la rete addestrata sia in grado di affrontare efficacemente il problema in analisi.
- Performance. L'addestramento termina qualora la performance raggiunga un *goal* prefissato, come ad esempio un certo valore di RMS sul *training set* o sul *validation set*. Tale genere di condizioni ha il vantaggio di garantire che la rete, una volta addestrata, mostri le performance richieste, ma, non essendovi alcuna limitazione sul numero di interazioni da eseguire, la durata della procedura d'addestramento non è imposta a priori. Di conseguenza, se ad esempio non è possibile convergere alla performance imposta, l'algoritmo di addestramento procede all'infinito.

Da quanto appena esposto è evidente che l'utilizzo di condizioni d'arresto basate esclusivamente sul numero di interazioni o sulla performance comporta criticità spesso non accettabili. Per questo motivo, gli algoritmi di addestramento includono, in generale, più condizioni di arresto, sia basate sul numero di epoche che sulla performance, e terminano quando almeno una delle condizioni imposte viene raggiunta.

4.2. Modelli di stato

4.2.1. Introduzione

I modelli di stato hanno come obiettivo quello di fornire la rappresentazione matematica di un determinato sistema fisico tramite la descrizione dell'evoluzione temporale di quelle variabili, dette appunto "variabili di stato" che, istante per istante, caratterizzano la condizione interna di tale sistema.

La descrizione di un generico sistema dinamico S a tempo discreto tramite un modello a variabili di stato può essere espressa come segue:

$$S: \begin{cases} \mathbf{x}(t+1) = F\mathbf{x}(t) + G\mathbf{u}(t) \\ \mathbf{y}(t) = H\mathbf{x}(t) \end{cases} \quad (4.20)$$

dove $\mathbf{x}(t)$ rappresenta il vettore delle variabili di stato, $\mathbf{u}(t)$ il vettore degli ingressi del sistema e $\mathbf{y}(t)$ il vettore delle uscite. Le matrici F , G ed H contengono i parametri che governano l'evoluzione temporale del sistema S . In particolare F governa la dinamica delle variabili di stato, G modula gli ingressi e H determina la trasformazione d'uscita [Figura 4.9].

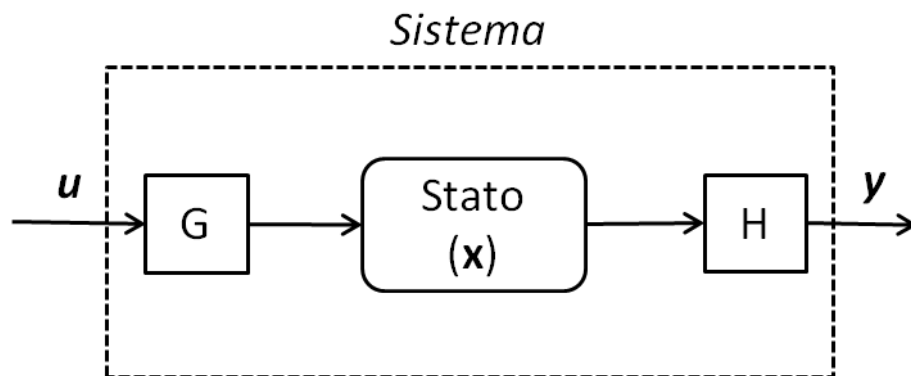


Figura 4.9: Modello a variabili di stato di un generico sistema.

Si noti che il sistema S , per come è stato definito, è lineare. L'estensione al caso non lineare è la seguente:

$$S: \begin{cases} \mathbf{x}(t+1) = f(\mathbf{x}(t)) + g(\mathbf{u}(t)) \\ \mathbf{y}(t) = h(\mathbf{x}(t)) \end{cases} \quad (4.21)$$

dove $f(\cdot)$, $g(\cdot)$, ed $h(\cdot)$ sono generiche funzioni che governano l'evoluzione del sistema in modo analogo alle matrici F , G , H presentate nell'equazione 4.20.

Da un punto di vista applicativo, i modelli a variabili di stato possono essere utilizzati per calcolare l'uscita di un sistema di interesse a fronte di ingressi noti. Si noti, però, che per fare questo è necessario disporre delle seguenti informazioni:

- Stato iniziale del sistema: $\mathbf{x}_0 = \mathbf{x}(t=0)$. Tale informazione consiste nella descrizione della condizione iniziale del sistema, ovvero il punto di partenza della sua evoluzione temporale. Nel caso in cui la variabile di stato non sia soggetta ad una evoluzione dinamica, ovvero se lo stato ad un determinato istante non dipende dallo stato ad istanti precedenti ($F = \mathbf{0}$), l'informazione sull'inizializzazione del sistema non è necessaria.
- Parametri caratteristici del sistema. Come già detto tali parametri sono contenuti nelle matrici F,G,H (riferendosi per semplicità ad un caso lineare) e, noto il dato in ingresso e lo stato interno del sistema, permettono di calcolarne l'uscita per ogni istante temporale.

Nel caso in cui si consideri un sistema fisico la cui modellizzazione matematica è nota a priori, allora i parametri caratteristici sono noti ed possibile utilizzare il modello, ad esempio, per simulare il comportamento del sistema fisico di interesse, ovvero per calcolarne le uscite al variare degli ingressi.

Se, invece, per il sistema fisico di interesse non si dispone di una modellizzazione matematica (problema a scatola nera), ovvero non si dispone dei suoi parametri caratteristici, è necessario, per prima cosa, effettuare una stima di quest'ultimi. A tal fine è possibile operare in modo concettualmente analogo a quanto è stato detto riguardo all'addestramento di una ANN (paragrafo 4.1.1), ovvero osservare il comportamento del sistema fisico raccogliendo coppie ingresso/uscita. A seguito di tali osservazioni è possibile ottenere una descrizione matematica della correlazione tra ingressi ed uscite [49,60], che può essere quindi utilizzata per ottenere l'uscita del sistema in funzione degli ingressi. Si noti che tale uscita, in quanto ottenuta a partire da parametri caratteristici stimati (e non noti a priori), non può essere considerata esatta bensì è da considerarsi come una stima.

In Figura 4.10 è proposto il confronto tra l'utilizzo di modelli a variabili di stato nel caso di sistemi noti a priori e nel caso di problemi a scatola nera.

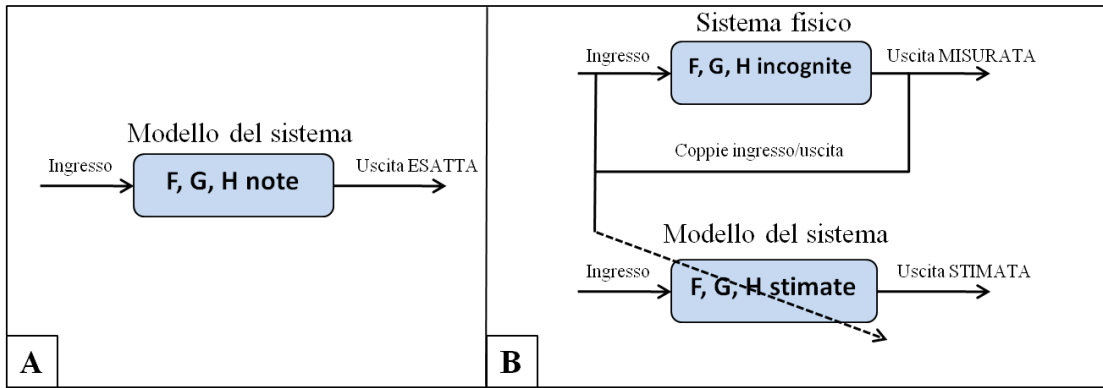


Figura 4.10: Pannello A: Il sistema fisico è noto a priori. Pannello B: Il sistema fisico è incognito (problema a scatola nera).

Nel paragrafo seguente viene brevemente descritto il metodo utilizzato per stimare i parametri del modello.

4.2.2. Stima dei parametri del modello

Il più diffuso metodo per effettuare la stima dei parametri di un modello sulla base di osservazioni di coppie ingresso/uscita relative al sistema fisico che si vuole modellizzare, consiste nel metodo dei minimi quadrati (LM, *Least Square*), che viene quindi richiamato sinteticamente [61].

Per prima cosa si definisca una cifra di merito J nel seguente modo:

$$J = \frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t))^2 \quad (4.22)$$

dove $y(t)$ rappresenta il vettore delle uscite misurate al t -esimo istante e $\hat{y}(t)$ la stima di tale vettore. Questa stima è calcolata come il prodotto tra un vettore di ingressi φ e un vettore θ formato dai parametri del modello che devono essere stimati. Dunque la cifra J può essere scritta come segue:

$$J = J(\theta) = \frac{1}{N} \sum_{t=1}^N (y(t) - \varphi^T(t)\theta)^2 \quad (4.23)$$

I parametri che meglio descrivono il sistema sono quelli che garantiscono che l'uscita stimata sia quanto più prossima possibile a quella misurata, ovvero quelli che minimizzano la cifra J :

$$\hat{\theta} = \operatorname{argmin}_{\theta} (J(\theta)) \quad (4.24)$$

Per calcolare tali parametri è necessario ricavare gli zeri della derivata prima di J rispetto a θ , imponendo:

$$\frac{dJ(\theta)}{d\theta} = 0 \quad (4.25)$$

e ottenendo:

$$\sum_t \varphi(t)y(t) = \left(\sum_t \varphi(t)\varphi(t)^T \right) \hat{\theta} \quad (4.26)$$

Dalla quale si ricava un sistema di equazioni dette “normali”:

$$\hat{\theta} = \left(\sum_t \varphi(t)\varphi(t)^T \right)^{-1} \left(\sum_t \varphi(t)y(t) \right) \quad (4.27)$$

Dunque, nell’ipotesi in cui la matrice $\sum_t \varphi(t)\varphi(t)^T$ sia invertibile, è possibile ottenere il vettore di parametri che minimizza l’errore di stima espresso dall’equazione 4.22. Valutando il segno della derivata seconda di J rispetto a θ , si dimostra che $\hat{\theta}$ costituisce effettivamente un punto di minimo:

$$\frac{d^2J}{d\theta^2} = \frac{2}{N} \left(\sum_t \varphi(t)\varphi(t)^T \right) \geq 0 \quad (4.28)$$

I parametri ottimi così calcolati consentono di ricavare le matrici F, G ed H e, di conseguenza, di ottenere un modello di un dato sistema fisico.

4.3. Database utilizzato

Al fine di sviluppare e validare il progetto descritto in questo elaborato, sono stati utilizzati i dati contenuti in un database inerente ai trattamenti effettuati con il sistema *Synchrony* presso il Georgetown University Medical Center (Washington, DC) [Tabella 4.1]

Database Georgetown University Medical Center	
Numero di pazienti	86
Numero totale di frazioni	340
Numero di collocazioni anatomiche considerate	25

Tabella 4.1: Database relativo ai trattamenti con *Synchrony* presso il Georgetown University Medical Center.

Per ogni paziente il database riporta, per tutta la durata del trattamento, le informazioni relative alla posizione dei marcatori esterni, alla posizione stimata del *target* e, in corrispondenza degli istanti in cui vengono acquisite le immagini di controllo, alla sua posizione effettiva.

È chiaro quindi che tale database è risultato essenziale nell’ambito del lavoro descritto, in quanto rende disponibile una grande quantità di informazioni dettagliate e quantitative sul funzionamento di un sistema di *tracking*. Una descrizione più accurata delle informazioni utilizzate per lo sviluppo del progetto proposto verrà proposta nel seguente capitolo.

5. Sviluppo del progetto

Il presente lavoro ha come scopo generale quello di realizzare un prototipo software per l'implementazione di un sistema di *tracking online* utilizzabile in adroterapia. Come già anticipato nel paragrafo 3.5, tale software deve essere in grado di apprendere la correlazione tra il movimento di marcatori posti sulla superficie corporea del paziente (segnali esterni) e quello della lesione tumorale (segnale interno). Tale correlazione può fornire informazioni in tempo reale sulla posizione della lesione in funzione della posizione dei marcatori esterni. Inoltre, una caratteristica peculiare del software deve essere quella di riuscire a garantire un adeguato aggiornamento dei parametri dei modelli qualora sia registrato uno scostamento rilevante tra la posizione vera del bersaglio (acquisita con immagini RX di controllo) e quella stimata.

Gli strumenti che verranno proposti per assolvere alle funzioni appena citate sono le reti neurali e i modelli a variabili di stato, la cui teoria è stata presentata rispettivamente nei paragrafi 4.1 e 4.2. Le motivazioni che hanno portato alla scelta di tali strumenti sono state dettate sia dalla natura del problema che si vuole affrontare che dalle specifiche del progetto che si vuole realizzare e possono essere riassunte nei seguenti punti:

- Capacità di affrontare problemi a scatola nera

La predizione del segnale interno in funzione dei segnali esterni è un problema che può essere affrontato efficacemente con un approccio di tipo *black box*. Il fenomeno dell'*organ motion*, infatti, è di natura estremamente complessa: si pensi che la traiettoria sulla quale si muove un bersaglio interno è fortemente dipendente sia dalla collocazione anatomica della lesione sia dalla condizione fisica del paziente, che può variare tra una frazione di trattamento e la successiva. È quindi chiaro che non può esistere una relazione matematica semplice valida per tutte le lesioni e per ogni condizione del paziente, che permetta di calcolare la posizione del bersaglio in funzione del movimento della superficie corporea. Inoltre, ricordando che una delle specifiche esposte nel paragrafo 3.5 richiede allo strumento sviluppato di eseguire le stime in tempo reale, l'approccio di tipo *black box* risulta vantaggioso anche in termini computazionali, in quanto la complessità che sarebbe richiesta ad un eventuale modello fisico per la soluzione del problema posto renderebbe estremamente difficoltoso soddisfare tale specifica.

Per quanto riguarda le reti neurali, la soluzione di problemi a scatola nera è sicuramente una delle loro applicazioni principali, in quanto il loro addestramento non richiede alcuna conoscenza a priori riguardo al sistema che si vuole studiare e/o simulare.

Anche i modelli a variabili di stato risultano adeguati ad affrontare problemi *black box*, in quanto i parametri caratteristici del sistema possono essere identificati a seguito dell'osservazione di coppie ingresso/uscita di quest'ultimo, anche in questo caso senza bisogno di alcuna conoscenza a priori.

- Disponibilità di dati riguardanti la correlazione interno/esterno

Come già scritto, entrambi gli strumenti proposti, per poter essere impiegati con successo, devono “apprendere” la relazione tra segnale interno e segnali esterni tramite l'osservazione di esempi. Tale apprendimento si traduce nella modulazione dei pesi nelle reti neurali e nell'identificazione dei parametri per quanto riguarda i modelli di stato.

Nell'ambito di questo lavoro, il database relativo ai trattamenti con il modulo *Synchrony* per *CyberKnife* raccolto presso il Georgetown University Medical Center (paragrafo 4.3) costituisce la fonte di esempi richiesta. Per ogni trattamento registrato, infatti, sono riportate, su una comune base temporale, le informazioni relative sia alla posizione spaziale di tre marcatori esterni sia alla predizione della posizione del bersaglio eseguita in tempo reale dal sistema. Inoltre, in corrispondenza degli istanti in cui vengono acquisite le immagini di controllo, sono rese disponibili le effettive coordinate della lesione.

In aggiunta ai dati reperiti del suddetto database, altri dati sono stati acquisiti attraverso un'attività sperimentale svolta utilizzando un fattoccio per la simulazione del movimento respiratorio, che sarà descritto nel paragrafo 5.3.

- Possibilità di implementare algoritmi adattativi

Poiché tra le specifiche del progetto vi è la possibilità di aggiornare la correlazione interno/esterno appresa dagli algoritmi implementati, è necessario che gli strumenti scelti permettano di soddisfare tale specifica. Per quanto riguarda le reti neurali, l'aggiornamento della correlazione può essere facilmente eseguito tramite un nuovo addestramento della rete nel quale si utilizza un *training set* che comprenda i nuovi esempi resi disponibili grazie all'acquisizione delle immagini di controllo. In modo analogo, anche la stima dei parametri del modello di stato può essere aggiornata considerando i nuovi dati disponibili.

- Capacità di stimare la posizione del bersaglio in tempo reale

Come già scritto nel paragrafo 3.5, un software di correlazione interno/esterno deve fornire la stima della posizione del bersaglio al sistema di somministrazione della dose necessariamente in tempo reale, dunque non può comprendere algoritmi la cui esecuzione richieda un tempo elevato, in quanto, in tal caso, la stima fornita si riferirebbe ad un istante temporale precedente, rispetto al quale il bersaglio si è sicuramente spostato. Si ricorda,

infatti, che i fenomeni di *organ motion* intra-frazione mostrano una ciclicità dell'ordine del secondo e quindi causano spostamenti molto rapidi.

Le reti neurali, nell'ipotesi di non implementare architetture eccessivamente complicate e di utilizzare calcolatori adeguati, soddisfano la condizione appena esposta. Lo stesso vale per i modelli a variabili di stato, con i quali, oltre al calcolo dell'uscita, anche l'apprendimento e l'aggiornamento della correlazione può avvenire quasi in tempo reale [60].

L'iter che è stato seguito per lo sviluppo del lavoro è mostrato schematicamente in Figura 5.1 e commentato di seguito.

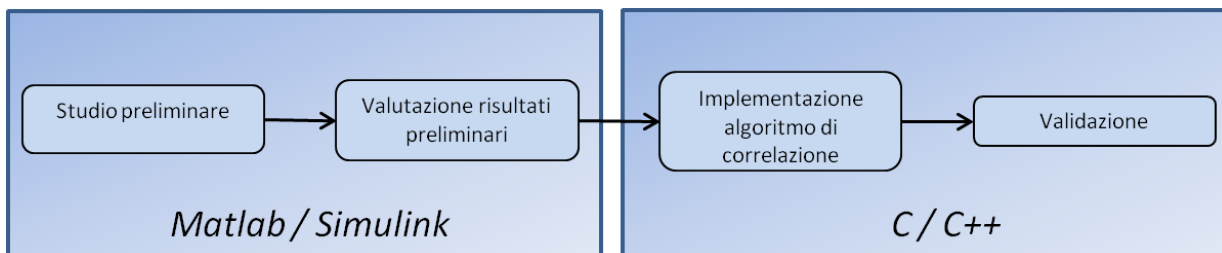


Figura 5.1: Schema a blocchi dell'iter seguito per lo svolgimento del progetto. È evidenziato l'ambiente di sviluppo utilizzato in ogni fase.

La prima fase del lavoro consiste in uno studio preliminare volto a valutare l'adeguatezza degli strumenti proposti per la soluzione del problema che si vuole affrontare. I risultati di tale studio sono stati sfruttati per indirizzare lo svolgimento delle fasi successive verso l'utilizzo dei metodi e degli strumenti la cui performance, in termini di capacità di apprendimento della correlazione interno/esterno, è risultata migliore. In queste fasi preliminari del progetto si è deciso di lavorare in ambiente Matlab, ricorrendo anche a Simulink, un'applicazione dedicata alla simulazione e allo sviluppo di sistemi dinamici. Si è scelto tale ambiente dal momento che esso dispone di strumenti tramite i quali l'implementazione di reti neurali e modelli di stato risulta notevolmente rapida e semplificata per l'utente, come è necessario che sia per lo svolgimento di uno studio preliminare.

Dopo aver individuato lo strumento migliore con il quale operare, si è proceduto all'effettiva implementazione del software di correlazione interno/esterno, utilizzando il linguaggio C/C++. Tale ambiente, seppure non disponga di strumenti dedicati allo sviluppo di reti neurali o modelli di stato, garantisce una rapida esecuzione degli algoritmi e permette di sviluppare codici che possono essere facilmente connessi con altri strumenti software. Entrambe queste caratteristiche risultano fondamentali per ottenere un'esecuzione in tempo reale e per fornire la possibilità d'integrazione all'interno di sistemi più complessi. Si ricorda, infatti, che l'algoritmo

proposto consiste nel prototipo di un modulo che necessariamente dovrà essere connesso con il software che gestisce il sistema di somministrazione della dose.

In ultimo, il lavoro si conclude con la validazione dello strumento sviluppato,

Lo descrizione dello studio preliminare e dell'implementazione del software saranno l'argomento dei paragrafi 5.1 e 5.2 rispettivamente, mentre la fase di validazione sarà oggetto del capitolo 6.

5.1. Studio preliminare

Come già accennato, lo scopo dello studio preliminare consiste nel valutare la capacità degli strumenti computazionali proposti (reti neurali e modelli a variabili di stato) di apprendere la correlazione tra segnale interno e segnali esterni e di utilizzare tale correlazione per predire il movimento della lesione in tempo reale. In aggiunta, i risultati di tale studio hanno permesso di indirizzare le fasi successive del lavoro verso lo sviluppo dello strumento che ha mostrato una migliore performance.

Sia per l'implementazione della rete neurale sia del modello di stato, nell'ambito di questo studio preliminare si è proceduto nel seguente modo:

- Il segnale interno che è stato considerato non consiste direttamente nella posizione del *target* individuata grazie alle acquisizione radiografiche, bensì nella predizione effettuata dell'apposito software del sistema *Synchrony*, chiamato *modeler*. Si è deciso di operare in questo modo dal momento che il segnale del *modeler* è calcolato in tempo reale e, di conseguenza, può essere associato istante per istante al segnale esterno (*markers* esterni). In tal modo è semplice ottenere un numero di coppie ingresso/uscita sufficienti all'apprendimento della correlazione. Al contrario, se si fosse considerato come segnale interno la posizione del *target* acquisita dal sistema di *imaging*, la frequenza di tale segnale (un dato ogni 2 minuti circa), supposto di operare in condizioni di stazionarietà, avrebbe reso indispensabile la stima di ulteriori coppie ingresso/uscita tramite interpolazione.

In aggiunta, l'utilizzo del *modeler* come segnale interno di riferimento permette di valutare la predizione non soltanto negli istanti in cui viene acquisita un'immagine radiografica.

- Per ogni frazione analizzata, non si è considerata l'intera durata del trattamento, bensì un intervallo costituito da 5000 campioni (corrispondenti a circa 200 secondi) del

segnale esterno, ai quali corrispondono altrettanti campioni del *modeler*. Tale intervallo è stato scelto con l'unico criterio di evitare le fasi iniziali del trattamento, nelle quali il software del sistema *Synchrony* costruisce il modello di correlazione e, di conseguenza, la predizione effettuata dal *modeler* non è ancora affidabile.

Tale numero di campioni risulta adeguato per valutare efficacemente e rapidamente la differenza tra la predizione del *modeler* e quella effettuata dalla rete o dal modello di stato.

- L'apprendimento della correlazione interno/esterno è stato effettuato considerando le prime 500 coppie *markers/modeler*, corrispondenti a circa 20 secondi di trattamento. Tali coppie costituiscono il *training set* per l'addestramento della rete neurale e i dati dai quali estrarre la matrice dei parametri per il modello di stato.
- La correlazione appresa su tali coppie non viene mai aggiornata. Ciò è giustificato, oltre che dalla natura preparatoria dello studio che si sta presentando, anche dalla durata dell'intervallo di trattamento considerato: 5000 campioni, infatti, corrispondono a circa 3 minuti del trattamento, tempo nel quale il modello di correlazione può considerarsi adeguato senza la necessità di essere aggiornato.

Nei prossimi paragrafi si descriverà il loro sviluppo per l'implementazione della rete neurale (paragrafo 5.1.1) e del modello di stato (paragrafo 5.1.2). Successivamente, nel paragrafo 5.1.3 verrà presentata una semplice interfaccia grafica che è stata sviluppata per semplificare l'acquisizione e l'analisi dei dati. In fine, il paragrafo 5.1.4 presenterà i risultati ottenuti dallo studio preliminare.

5.1.1. Rete neurale

La rete neurale è stata realizzata lavorando sia in ambiente Matlab sia in Simulink. La Figura 5.2 mostra uno schema a blocchi delle attività svolte per implementare la rete neurale ed è commentata di seguito.

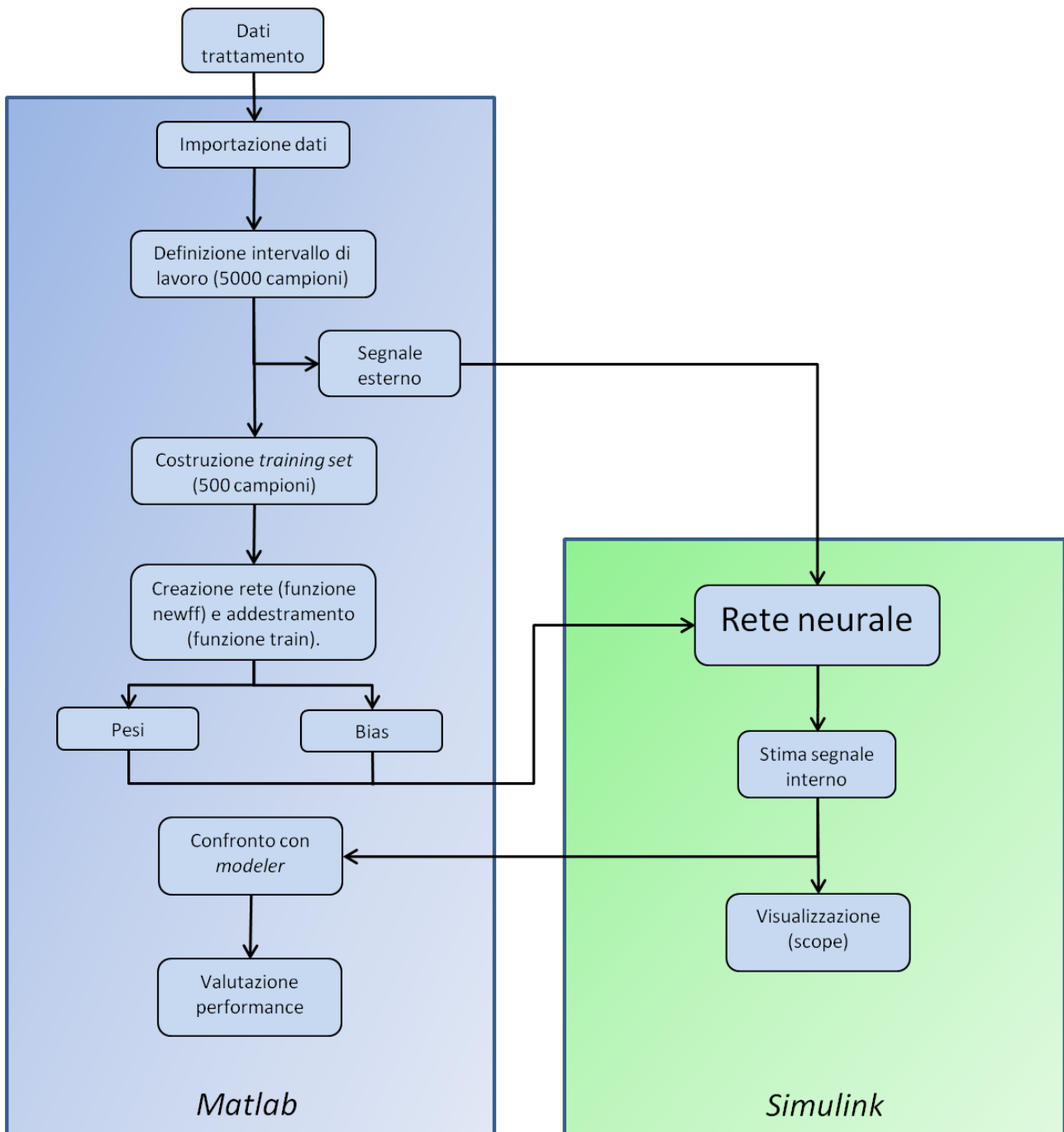


Figura 5.2: Schema a blocchi dell'implementazione della rete neurale in Matlab/Simulink.

I file contengono le informazioni relative al trattamento (in formato *.log*) vengono importati in Matlab attraverso un'apposita funzione che, leggendo tali file, costruisce delle strutture contenenti i dati di interesse. Nell'ambito di questo studio vengono considerate le seguenti due strutture:

- *Markers*: contiene un vettore di riferimenti temporali (chiamate *timestamps*) e tre matrici contenenti ognuna le coordinate tridimensionali di ciascuno dei 3 marcatori posti sulla superficie corporea del paziente, rilevate in corrispondenza di ogni *timestamp*. Da tale struttura verrà estratto il segnale esterno che, nell'ambito di questo studio preliminare consiste nella posizione spaziale del primo marcatore.
- *Modeler*: contiene un vettore di riferimenti temporali e una matrice formata dalle coordinate della posizione stimata del *target*, dalla quale si estrarrà il segnale interno.

Completata l'importazione dei dati, si è proceduto ad identificare l'intervallo di 5000 campioni sui quali operare. A tal fine si è tracciato l'andamento temporale del *modeler* e si è identificato un intervallo entro il quale tale segnale non presentasse particolari anomalie e non subisse interruzioni.

Successivamente, sfruttando la comune base temporale data dalle *timestamps*, sono stati identificati, i 5000 campioni del segnale esterno corrispondenti temporalmente a quelli appena selezionati per il *modeler*.

Al termine di questa operazione, si ha a disposizione una matrice contenente il segnale interno e una contenente il segnale esterno, mentre il *training set* è stato costruito semplicemente accoppiando i primi 500 campioni delle due suddette matrici. La Tabella 5.1 riassume le caratteristiche dei segnali appena descritti:

Nome	Descrizione	Dimensioni
Segnale esterno	Un marcatore esterno	5000x3
Segnale interno	Stima eseguita dal <i>modeler</i>	5000x3
Training set	Coppie ingresso/uscita desiderata	500x3 (in) 500x3 (out)

Tabella 5.1: Riassunto dei dati ricavati dalle informazioni sul trattamento.

Con tali dati a disposizione è possibile creare ed addestrare una rete neurale che apprenda la correlazione interno/esterno. A tal fine sono state utilizzate due funzioni del *Neural Network Toolbox* di Matlab, configurate nel seguente modo:

- `net = newff(ts_in,ts_out,20)`

Crea una rete neurale *feed forward* per l'apprendimento del *training set* descritto dalle matrici `ts_in` (ingressi) e `ts_out` (uscite desiderate). Utilizzando la sintassi riportata e senza specificare altri parametri, la funzione crea una rete neurale avente un numero di ingressi pari alle dimensioni di `ts_in` (3), un singolo strato nascosto formato da 20 neuroni aventi funzione di attivazione “*tansig*” (ovvero un'approssimazione della tangente iperbolica) e uno strato d'uscita formato da un numero di neuroni pari alle dimensioni di `ts_out` (3) con funzione d'attivazione “*purelin*” (cioè lineare).

La topologia della rete creata è riportata in Figura 5.3.

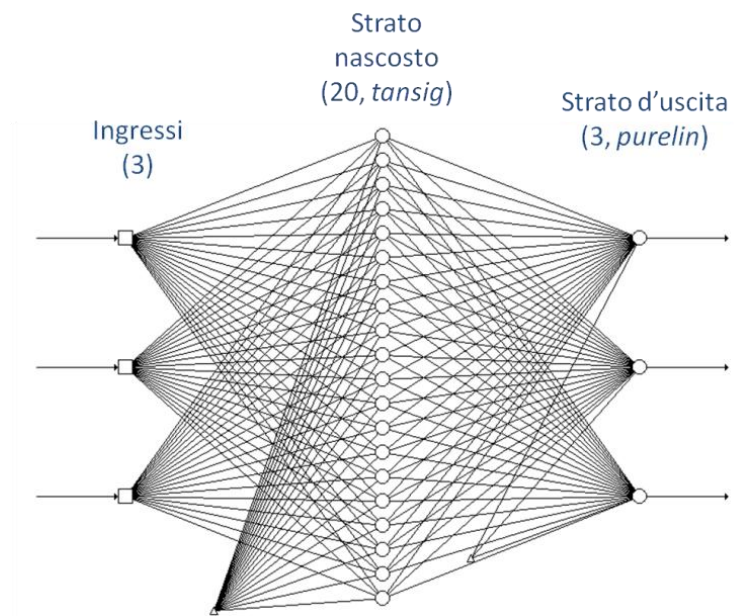


Figura 5.3: Topologia della rete neurale creata.

- `train(net,ts_in,ts_out)`

Esegue l'addestramento supervisionato della rete neurale precedentemente creata, utilizzando come *training set* le matrici specificate (`ts_in` e `ts_out`). L'addestramento avviene secondo i parametri impostati automaticamente quando è stata creata la rete. Tra questi, di particolare rilevanza è l'algoritmo di addestramento che, se non diversamente specificato, utilizza il metodo di Levenberg-Marquardt (*trainlm*). Tale metodo sfrutta le potenzialità sia dei metodi basati sulla discesa del gradiente sia dei metodi newtoniani e garantisce una rapida convergenza e ottime prestazioni, al costo, però, di un'ingente allocazione di memoria [63]. Il ciclo di addestramento termina nel momento in cui viene raggiunta la performance indicata (*goal* in termini di errore quadratico) o quando viene raggiunto il massimo numero di epoche consentite. Sia il

goal che il massimo numero di epoche consentite sono specificati alla creazione della rete e, se non vengono modificati dall'utente, valgono rispettivamente 0.001 e 1000.

Una volta eseguite tali funzione, i pesi e i bias ottenuti a seguito dell'addestramento sono stati salvati in quattro variabili descritte nella seguente tabella.

Variabile	Descrizione	Dimensioni
Weights_1	Pesi ingressi-strato nascosto	3x20
Weights_2	Pesi strato nascosto-strato d'uscita	20x3
Bias_1	Bias strato nascosto	20x1
Bias_2	Bias strato d'uscita	3x1

Tabella 5.2: Variabili contenenti pesi e bias della rete addestrata.

Come evidenziato in Figura 5.2, il calcolo dell'uscita della rete neurale è stato effettuato in Simulink, dove è stata costruita una rete analoga a quella creata in Matlab tramite la funzione newff [Figura 5.4].

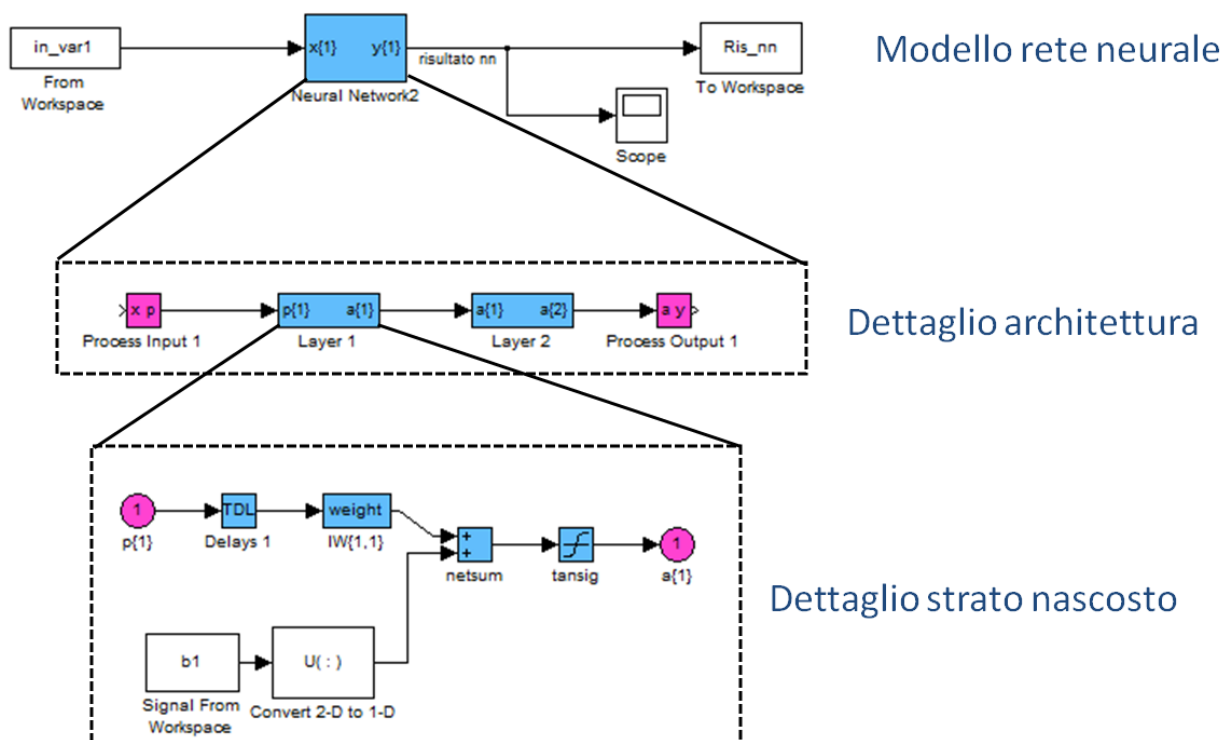


Figura 5.4: Implementazione della rete neurale in Simulink, con dettagli dell'architettura e dello strato nascosto.

Come si osserva in figura, il segnale esterno è stato importato in Simulink leggendo dal workspace di la variabile (in1) contenente il segnale relativo al marcatore esterno. In accordo con quanto detto precedentemente, la rete si compone di uno strato nascosto (Layer 1) e di uno strato d'uscita (Layer 2). A titolo di esempio è riportato il dettaglio del funzionamento dello

strato nascosto: il segnale in ingresso viene moltiplicato per i corrispondenti pesi (blocco “*weight*”), i quali vengono letti dal *workspace* (variabile *Weights_1*), successivamente il blocco “*netsum*” somma i bias, anch’essi letti dal *workspace* (variabile *Bias_1*). In ultimo viene applicata la funzione di trasferimento che, nel caso dello strato nascosto, consiste nella “*tansig*”. Il funzionamento dello strato d’uscita è ovviamente identico a quello appena descritto, fatta eccezione per la funzione d’attivazione che è lineare.

Le uscite della rete neurale vengono visualizzate in tempo reale tramite il blocco “*scope*” che, appunto, mostra la stima del segnale interno man mano che viene calcolato [Figura 5.5]. Le uscite vengono poi salvate nel *workspace* di Matlab, in modo tale da renderle disponibili per le successive analisi.

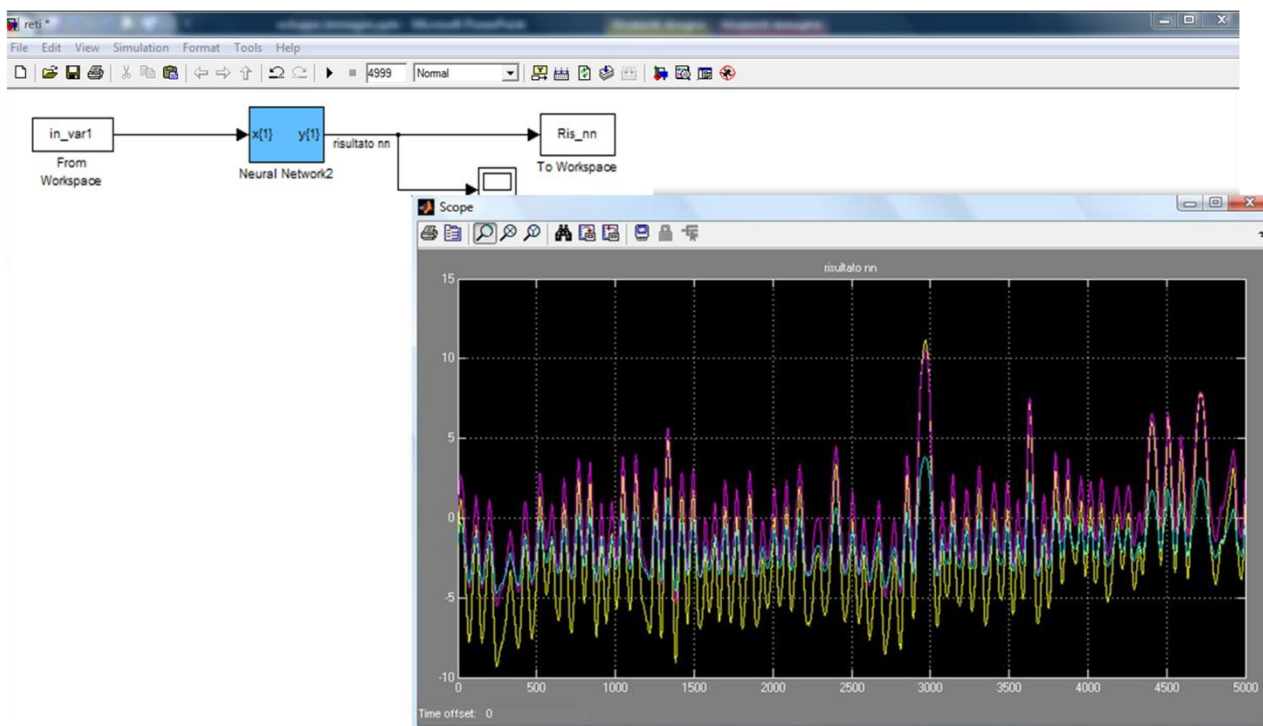


Figura 5.5: Schermata di Simulink al termine dalla simulazione. È visibile, nella finestra dello “*scope*”, la stima del segnale interno (3 coordinate).

Tali analisi consistono principalmente nel confronto con il segnale interno di riferimento, ovvero il *modeler*, e hanno lo scopo di valutare la performance della rete osservando, per ognuno dei 5000 campioni, lo scostamento tra le sue uscite e il *modeler*. I risultati di queste analisi verranno presentati nel paragrafo 5.1.4.

5.1.2. Modello a variabili di stato

Introduzione

Prima di descrivere l'implementazione in Matlab/Simulink del modello a variabili di stato, è opportuno chiarire, da un punto di vista teorico, come si è operato [60].

Per prima cosa si ricorda che lo scopo del modello consiste nello stimare la posizione interna del bersaglio in funzione di un segnale esterno, secondo l'equazione 5.1:

$$\hat{p}(r) = Af(r) \quad (5.1)$$

dove $\hat{p}(r)$ indica la posizione stimata del bersaglio, A è la matrice che contiene tutti i parametri del modello e $f(r)$ è un vettore funzione del surrogato esterno scalare r .

Per ottenere quest'ultimo segnale a partire dal segnale tridimensionale relativo alla posizione di un marcatore esterno, si è utilizzato il metodo dell'analisi delle componenti principali (PCA), che viene richiamato brevemente di seguito.

Si consideri una matrice X di dimensioni $n \times m$, con n numero di osservazioni e m numero di variabili. Nel caso specifico che si sta trattando, X è una matrice di dimensioni $n \times 3$, dove n sono le osservazioni di un marcatore esterno, ognuna delle quali consiste in 3 coordinate spaziali.

Come è noto, il metodo PCA prevede il calcolo di autovalori e autovettori della matrice di covarianza di X :

$$C = E[(X - \mu)(X - \mu)^T] \quad (5.2)$$

dove μ indica il valor medio.

C è una matrice $m \times m$, dunque si avranno m autovettori corrispondenti a m autovalori.

Successivamente, è necessario disporre in ordine decrescente gli autovalori ottenuti: la prima direzione principale sarà data dall'autovettore corrispondente al primo autovalore (il maggiore), la seconda dall'autovettore corrispondente al secondo autovalore (il secondo in ordine decrescente) e così via.

È quindi chiaro che, proiettando X su una direzione principale, si ottiene un segnale scalare ($n \times 1$). In questo specifico caso, r è stato ottenuto nel seguente modo:

$$r = Xv_1 \quad (5.3)$$

dove X è la matrice $n \times 3$ già citata e v_1 è il versore che identifica la prima direzione principale.

Tornando ad osservare l'equazione 5.1, il vettore $f(r)$ può essere descritto come segue, a seconda che si voglia realizzare un modello lineare o quadratico.

$$f_{lin} = \begin{bmatrix} r \\ 1 \end{bmatrix}; f_{quad} = \begin{bmatrix} r^2 \\ r \\ 1 \end{bmatrix} \quad (5.4)$$

Si anticipa che d'ora in avanti si tratterà soltanto di modelli quadratici, dato che quello che sarà implementato rientrerà in questa categoria.

Avendo a disposizione N osservazioni di r (indicate con $r_i, i = 1 \dots N$) e le corrispondenti osservazioni della posizione interna del bersaglio (indicate con $p_i, i = 1 \dots N$), è possibile stimare la matrice A utilizzando il metodo di minimi quadrati (paragrafo 4.2.2):

$$F = \begin{bmatrix} f(r_1)^T \\ \vdots \\ f(r_N)^T \end{bmatrix}; P = \begin{bmatrix} p_1^T \\ \vdots \\ p_N^T \end{bmatrix} \quad (5.5)$$

$$\hat{A} = P^T F (F^T F)^{-1} \quad (5.6)$$

In tal caso, la matrice di parametri A assume dimensioni 3×3 .

L'utilizzo di un modello a variabili di stato così come è stato appena descritto comporta, però, delle criticità legate all'isteresi che caratterizza la traiettoria di movimento del target interno [Figura 5.6], fenomeno che fa sì che ad un unico valore di r , possano corrispondere due valori di p .

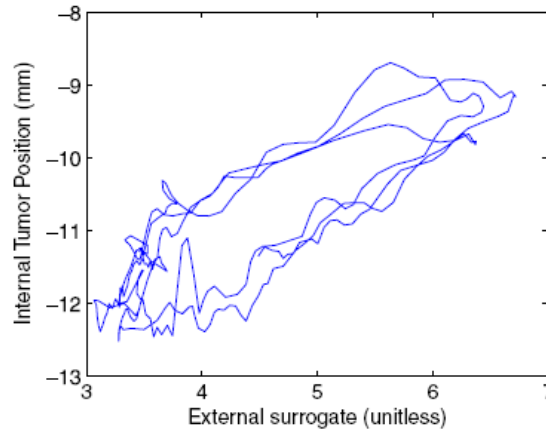


Figura 5.6: Posizione interna del tumore (p) in funzione del surrogato esterno (r) [60].

Al fine di risolvere tale ambiguità, una possibile soluzione consiste nell'aumentare lo stato del surrogato esterno includendo un campione ritardato di un certo numero di istanti (τ). Il surrogato esterno $r(n)$ viene quindi definito nel seguente modo:

$$r(n) \triangleq (r(n), r(n - \tau)) \quad (5.7)$$

Conseguentemente alla definizione appena data, il vettore $f(r)$ assume la seguente forma:

$$f(r) = \begin{bmatrix} r^2(n) \\ r^2(n - \tau) \\ r(n)r(n - \tau) \\ r(n) \\ r(n - \tau) \\ 1 \end{bmatrix} \quad (5.8)$$

Anche in questo caso la matrice A viene stimata esattamente come è riportato nell'equazione 5.6, con l'unica differenza che tale matrice avrà ora dimensioni 3×6 .

In conclusione, l'incremento dello stato del surrogato esterno permette di catturare nel modello una dinamica di primo ordine del sistema, infatti, la differenza $r(n) - r(n-\tau)$ può essere vista come una stima della velocità del bersaglio.

Di particolare rilevanza è la scelta del ritardo τ : utilizzando ritardi troppo lunghi si ottiene una ridotta informazione sulla dinamica locale del sistema. Al contrario, ritardi troppo brevi comportano un'eccessiva sensibilità al rumore. Nell'ambito di questo studio, il ritardo è stato posto pari a 40 campioni che, data la frequenza di acquisizione del sistema di circa 25 Hz, corrispondono a circa 1,6 secondi. Tale valore corrisponde a circa la metà della durata di un ciclo respiratorio medio.

Implementazione

Dopo aver descritto dal punto di vista teorico il modello a variabili di stato che si vuole realizzare, viene di seguito esposto il lavoro svolto per l'implementazione in Matlab/Simulink [Figura 5.7].

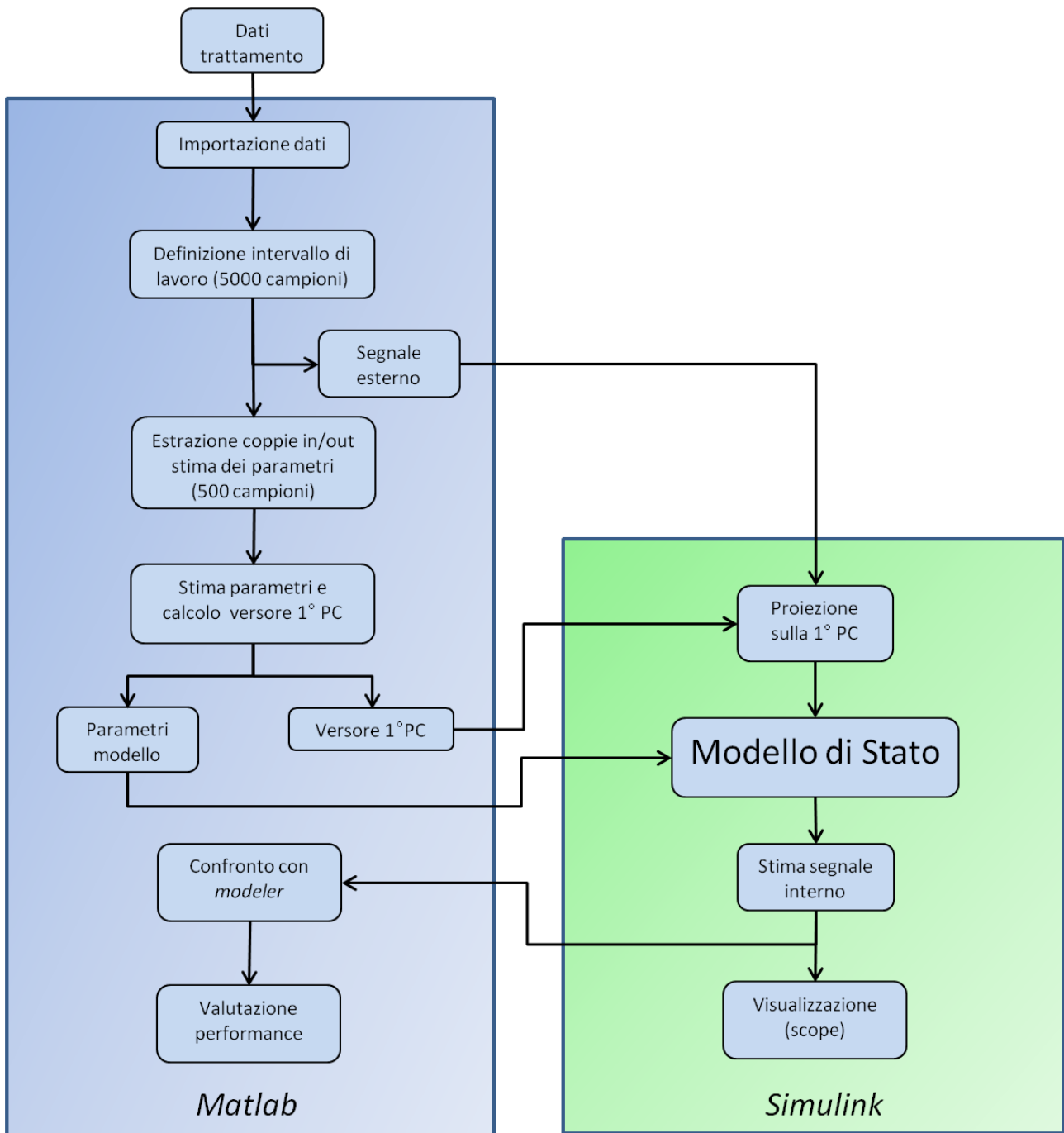


Figura 5.7: Schema a blocchi dell'implementazione del modello a variabili di stato in Matlab/Simulink.

Le operazioni svolte per ottenere le variabili contenenti il segnale esterno e il segnale interno, oltre alla 500 coppie ingresso/uscita necessarie per il calcolo della matrice dei parametri, sono identiche a quelle effettuate precedentemente all'implementazione della rete neurale.

Per calcolare i parametri del modello (matrice A) e il versore della prima componente principale (v_1), è stata scritta un'apposita funzione in Matlab (chiamata "state_model") la cui sintassi è:

$[A, v_1] = \text{state_model}(\text{marker}, \text{mod})$

Tale funzione riceve in ingresso due matrici di dimensione 500x3: la prima (“*marker*”) comprende le posizioni tridimensionali del marcatore esterno e la seconda (“*mod*”) le stime della posizione del bersaglio eseguite dal *modeler* negli istanti corrispondenti.

La prima operazione eseguita dalla “*state_model*” consiste nel chiamare la funzione “*princomp*”, la cui sintassi è riportata di seguito:

$[V, R] = \text{princomp}(\text{marker})$

Tale funzione riceve in ingresso la matrice “*marker*” e restituisce una matrice 3x3 (*V*) le cui colonne rappresentano i versori delle tre direzioni principali e una matrice *R* (500x3) che consiste nella rappresentazione di “*marker*” nello spazio della componente principale, essendo:

$$R = \text{marker} * V \quad (5.9)$$

Ora, la prima colonna della matrice *V* consiste nel versore della prima componente principale (v_1) e la prima colonna della matrice *R* consiste nel segnale scalare *r*.

Successivamente la funzione “*state_model*” calcola, per ogni campione, il vettore $f(r)$ esattamente come riportato nell’equazione 5.8 e, come mostrato nell’equazione 5.5, costruisce la matrice *F*.

La matrice *P*, contenente le stime della posizione del bersaglio eseguite dal *modeler*, coincide, per definizione, con la matrice “*mod*” fornita in ingresso.

A seguito di tali operazioni è quindi possibile calcolare la matrice *A*, tramite il metodo dei minimi quadrati, secondo l’equazione 5.6 che viene richiamata di seguito:

$$\hat{A} = P^T F (F^T F)^{-1} \quad (5.6)$$

Le uscite della funzione “*state_model*”, ovvero *A* e v_1 , vengono salvate nel *workspace* e di Matlab, in modo tale che possano essere richiamate durante l’esecuzione del modello in Simulink.

Tale modello è rappresentato in Figura 5.8 ed è commentato di seguito.

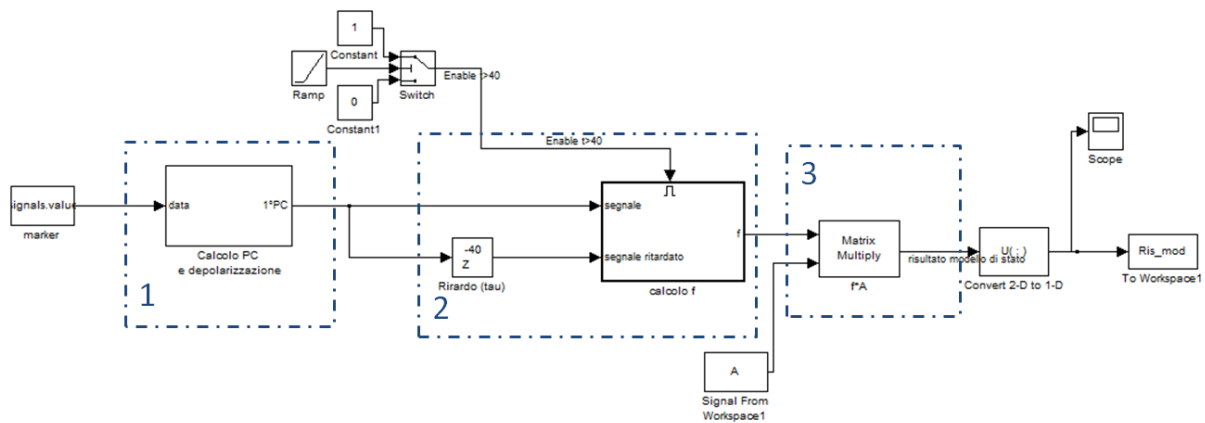


Figura 5.8: Modello a variabili di stato realizzato in Simulink. Sono evidenziati e numerati i principali blocchi funzionali.

Il modello importa dal *workspace* il segnale esterno (matrice “*marker*”), che viene elaborato dal blocco funzionale n°1 [Figura 5.9].

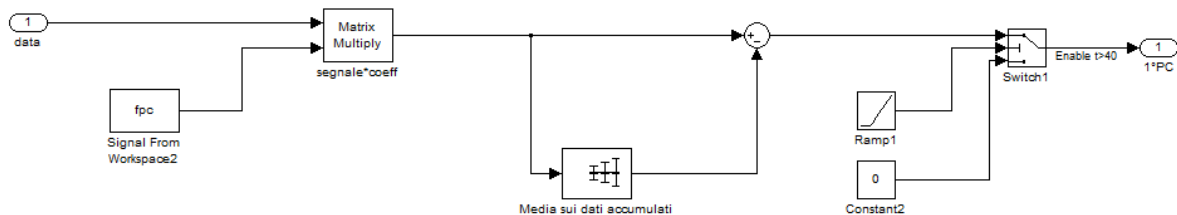


Figura 5.9: Blocco funzionale n°1.

La prima operazione eseguita da tale blocco consiste nell’importare dal *workspace* il versore della prima componente principale (v_1) e di moltiplicare il segnale esterno per questo vettore. Il segnale ottenuto viene poi depolarizzato sottraendovi la media calcolata su tutti i campioni accumulati.

L’uscita del blocco n°1 è abilitata soltanto dopo 40 istanti: in tal modo si garantisce che sia stato accumulato un numero di dati sufficiente affinché la media calcolata sia attendibile.

Successivamente il segnale viene elaborato dal blocco n°2 [Figura 5.10].

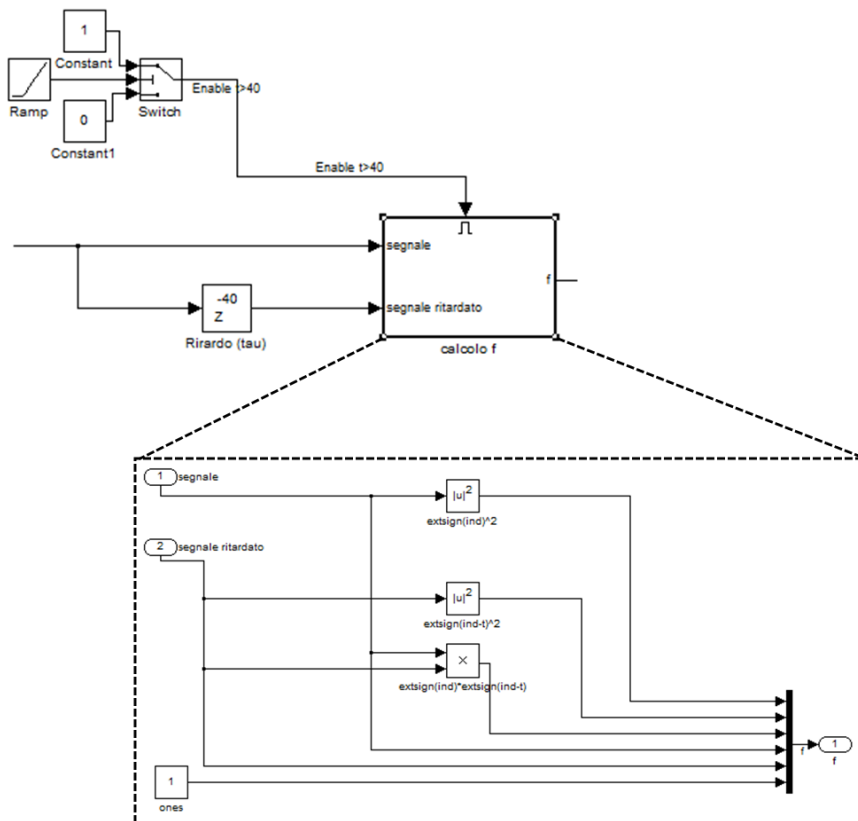


Figura 5.10: Blocco funzionale n°2.

Tale blocco ricava, per prima cosa, il campione ritardato $r(n-\tau)$, con $\tau = 40$ istanti. Successivamente calcola il vettore $f(r)$ come mostrato nel dettaglio riportato nella precedente figura.

Come si può notare, il blocco n°2 viene abilitato solo dopo 40 istanti, sia per attendere che l'uscita del blocco n°1 venga abilitata sia per consentire il calcolo del campione $r(n-\tau)$.

In ultimo, il blocco n°3 esegue semplicemente la moltiplicazione tra il segnale calcolato dal blocco n°2 e la matrice dei parametri A , importata dal *workspace*. Il risultato, che consiste nella stima del segnale interno, viene visualizzato tramite uno "scope" e salvato nel *workspace* per le successive analisi.

5.1.3. Interfaccia grafica

Come si è visto nei due paragrafi precedenti, per ottenere la stima del segnale interno e per valutarne la performance confrontandola con il *modeler*, è necessario eseguire, nel corretto ordine, delle operazioni preliminari in Matlab, solo a seguito delle quali il modello in Simulink può funzionare correttamente. Inoltre, per l'analisi dei risultati, è necessario ritornare a lavorare in ambiente Matlab.

Al fine di semplificare tale procedura, si è ritenuto opportuno sviluppare un'interfaccia grafica (GUI) che guidi l'utente dalla selezione della frazione di trattamento da analizzare fino alla presentazione dei risultati relativi alla performance sia della rete neurale che del modello di stato.

L'interfaccia sviluppata è mostrata in Figura 5.11.

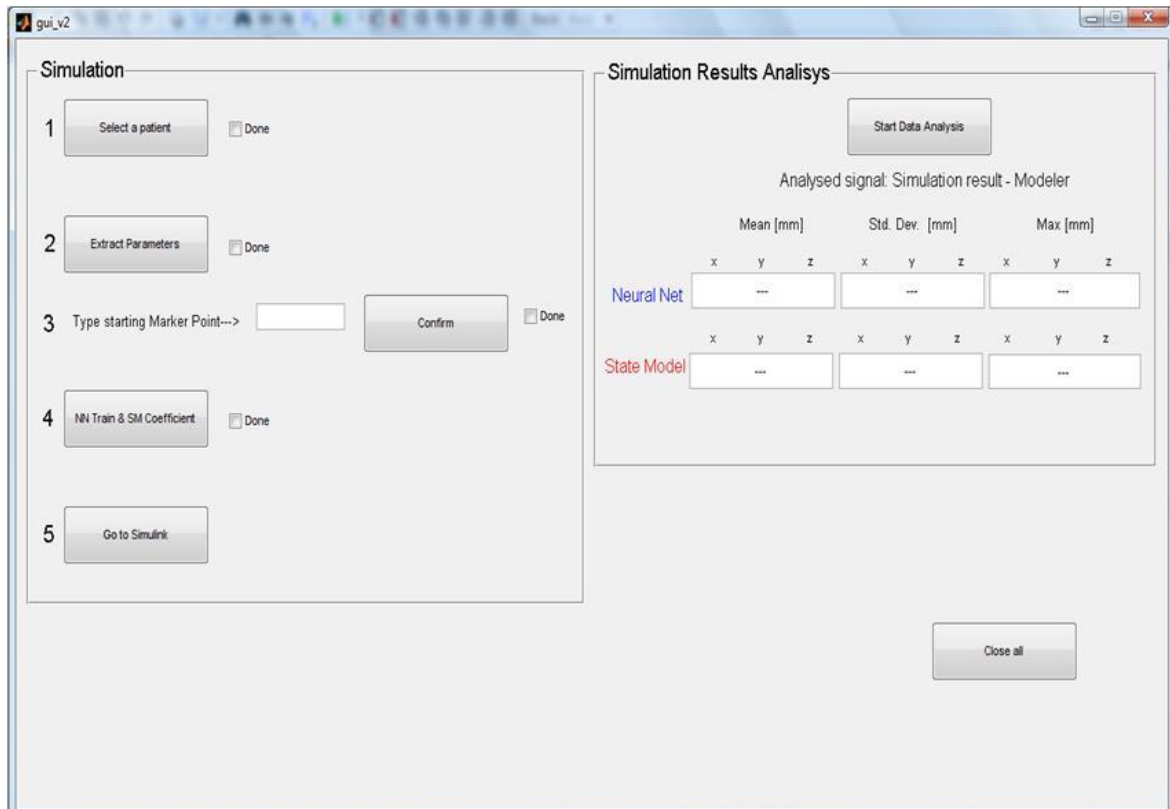


Figura 5.11: Interfaccia grafica per l'utente (GUI).

Come si può osservare, i comandi sul lato sinistro comprendono tutte le operazioni necessarie per eseguire la simulazione (in Simulink) del trattamento considerato, mentre il lato destro consente di analizzare i risultati ottenuti.

Di seguito vengono descritte le funzionalità associate ai pulsanti numerati da 1 a 5. Si noti che tali pulsanti sono numerati perché i comandi che questi richiamano devono necessariamente essere eseguiti nell'ordine indicato. Inoltre, alla destra dei primi quattro pulsanti, vi è un indicatore che comunica all'utente che l'operazione associata al corrispondente pulsante è stata completata ed è quindi possibile passare all'operazione successiva.

- Pulsante 1. Apre una finestra [Figura 5.12] nella quale l'utente è invitato a selezionare una cartella che contenga le informazioni relative ad un trattamento, sotto forma di 5 file *.log*.

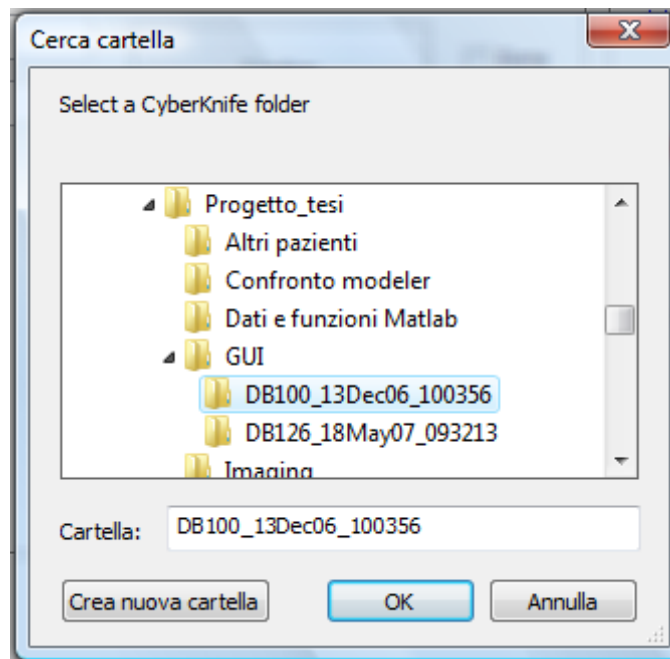


Figura 5.12: Finestra per la selezione del trattamento da importare.

- Pulsante 2. Esegue l'importazione in Matlab dei file *.log* contenuti nella cartella appena selezionata.
- Pulsante 3. L'utente deve compilare il campo posto alla sinistra del pulsante con l'indice del campione dei *Markers* dal quale iniziare l'analisi. In tal modo il codice può identificare i 5000 campioni sui quali operare. Utilizzando la corrispondenza in termini di *timestamp*, viene identificato il corrispondente campione del segnale relativo al *modeler*.

Una volta individuati gli indici iniziali, la pressione del pulsante estrae l'intervallo di lavoro (5000 campioni) e i dati per l'addestramento e l'identificazione dei parametri del modello di stato (500 campioni).

- Pulsante 4. Esegue le funzioni per la costruzione e l'addestramento della rete neurale ("*newff*" e "*train*") e la funzione "*state_model*" che, si ricorda, calcola i parametri del modello di stato e il versore della prima componente principale.
- Pulsante 5. Apre un modello di Simulink in cui sono presenti sia la rete neurale sia il modello a variabili di stato [Figura 5.13].

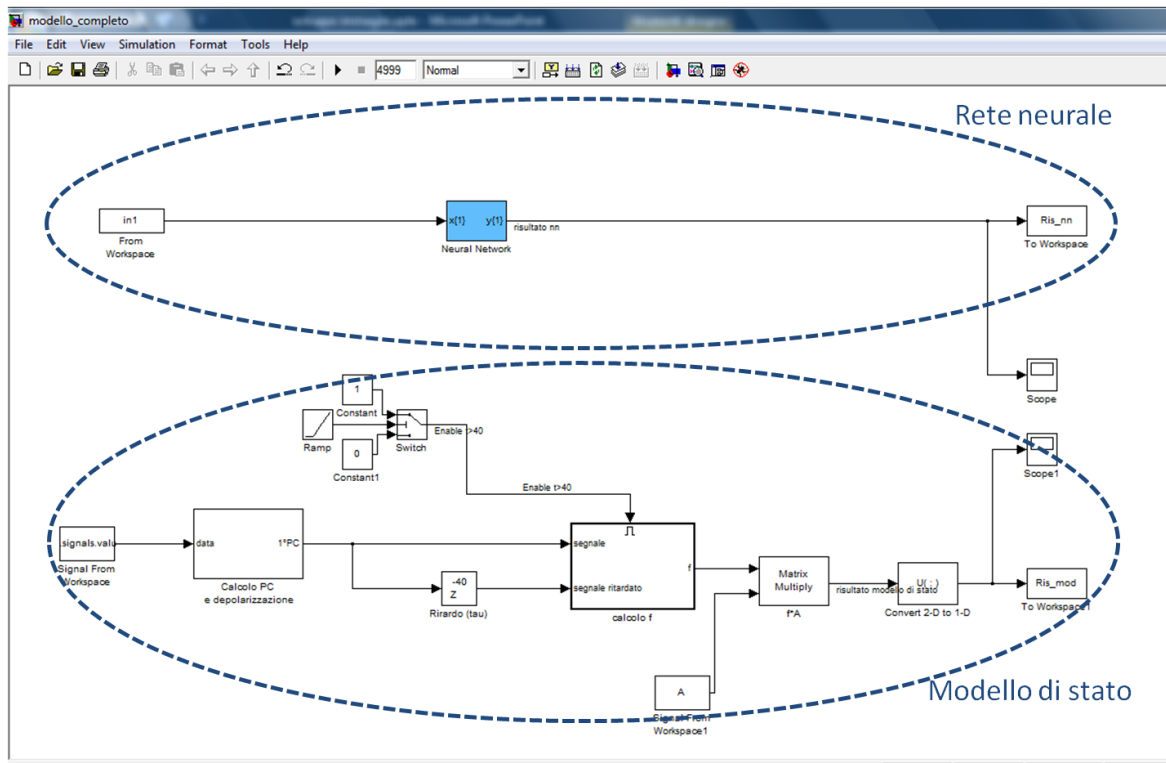


Figura 5.13: Modello in Simulink completo. Sono evidenziati i blocchi relativi alla rete neurale e quelli relativi al modello di stato.

Una volta che viene aperto questo modello, l'utente deve lanciare la simulazione in modo da ottenere le stime effettuate dalla rete neurale e dal modello di stato. Per valutarne la performance, le stime effettuate da questi due strumenti sono state confrontate con il segnale di riferimento, ovvero il *modeler*. Tale confronto è stato eseguito semplicemente calcolando, per ognuno dei 5000 campioni, la differenza (in valore assoluto) tra le stime effettuate dalla rete (indicata da $\hat{p}_{nn,i}$) e dal modello di stato (indicata da $\hat{p}_{sm,i}$) e quella effettuata dal *modeler* (indicata da p_i):

$$d_{nn,i} = |\hat{p}_{nn,i} - p_i|; d_{sm,i} = |\hat{p}_{sm,i} - p_i|; \forall i = 1 \dots 5000 \quad (5.10)$$

Risulta chiaro che i segnali differenza così ottenuti (d_{nn} e d_{sm}) conservano l'informazione relativa a ciascuna coordinata, dunque è possibile confrontare le performance valutando la singola coordinata. Questo non sarebbe stato possibile se si fosse calcolata una differenza 3D invece che una differenza componente per componente.

Per avviare le analisi appena descritte, l'utente deve premere il pulsante "Start Data Analysis", sul lato destro dell'interfaccia grafica. In tal modo i segnali d_{nn} e d_{sm} vengono calcolati e riportati in un grafico, mentre negli appositi campi sottostanti al pulsante vengono visualizzate, per ogni coordinata, la media la deviazione standard e il valore massimo di tali segnali. In tal modo, viene presentata all'utente una valutazione della performance sia in modo qualitativo

(attraverso i grafici) sia in modo quantitativo (attraverso il calcolo di media, deviazione standard e valore massimo). A titolo d'esempio, la Figura 5.14 riporta un'immagine della GUI al termine dell'analisi dei dati.

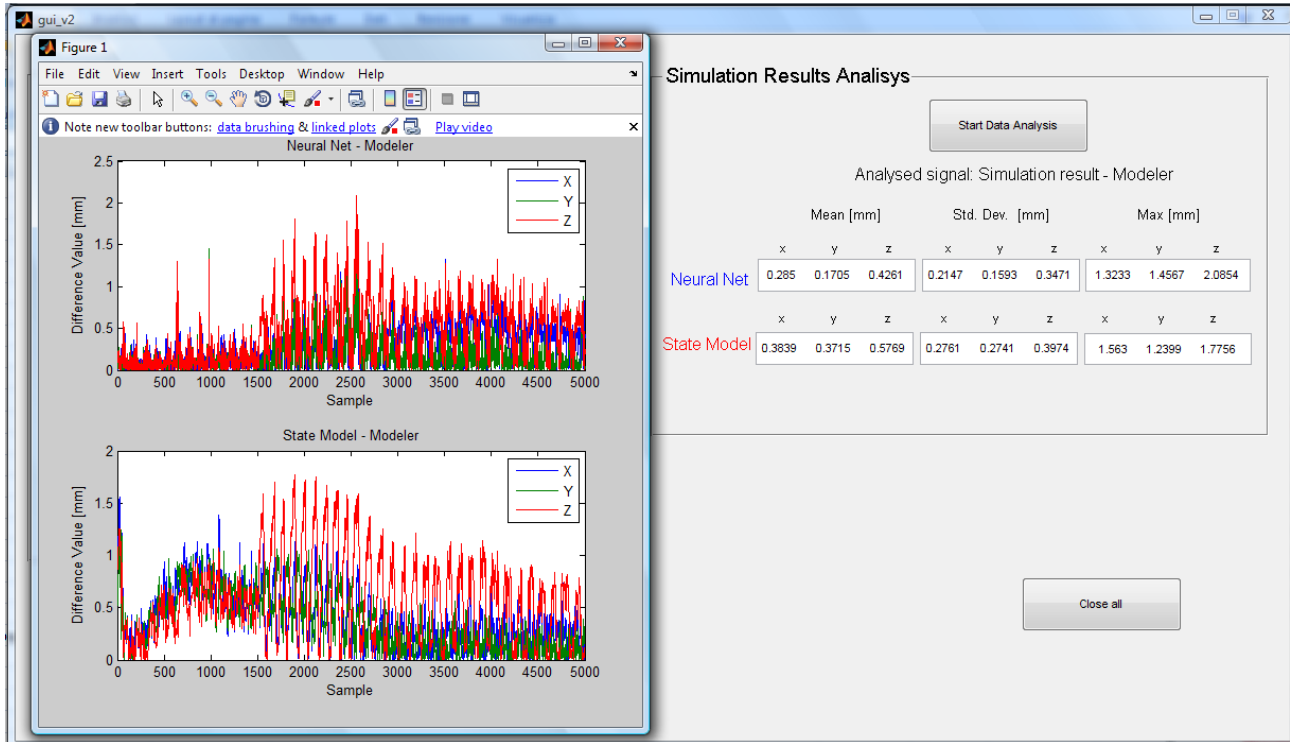


Figura 5.14: La GUI al termine dell'analisi dei risultati. A sinistra si notano i grafici dei segnali differenza d_{nm} (sopra) e d_{sm} (sotto), a destra i valori medi, le deviazioni standard e i valori massimi dei medesimi segnali.

In ultimo, il pulsante “Close all” chiude i grafici e l'interfaccia e i salva i risultati delle analisi in un file formato “.mat”.

5.1.4. Risultati

Come precedentemente descritto, lo scopo dello studio preliminare consiste nella valutazione della performance degli strumenti utilizzati (rete neurale e modello a variabili d stato) su un certo numero di trattamenti estratti dal database a disposizione. Nell'ambito di questo studio, tale valutazione è stata condotta considerando quattro pazienti scelti casualmente all'interno del database, per ognuno dei quali sono state analizzate due frazioni di trattamento.

Le modalità con le quali, per ogni trattamento analizzato, è stata quantificata la performance degli strumenti implementati sono esattamente quelle esposte nel paragrafo precedente (equazione 5.10).

Al fine di ottenere una valutazione complessiva dei segnali d_{nm} e d_{sm} sulle otto frazioni analizzate, sono stati calcolati, per ogni coordinata, i seguenti parametri:

$$\begin{aligned}
 \text{Media} &= \frac{1}{8} \sum_{k=1}^8 m_k \\
 \text{Deviazione standard} &= \frac{1}{8} \sum_{k=1}^8 \sigma_k \\
 \text{Massimo} &= \max (M_{k=1}, M_{k=2}, \dots, M_{k=8})
 \end{aligned}
 \tag{5.12}$$

dove m_k , σ_k e M_k indicano, per la k -esima frazione ($k = 1 \dots 8$), rispettivamente la media, la deviazione standard e il massimo dei suddetti segnali.

Nelle Tabelle 5.3 e 5.4 sono riportati i risultati ottenuti dalla rete neurale e dal modello a variabili di stato rispettivamente.

<i>Rete neurale</i>	X	Y	Z
Media [mm]	0,4885	0,4166	0,3455
Deviazione standard [mm]	0,4857	0,3671	0,3310
Massimo assoluto [mm]	9,3105	6,7544	4,6302

Tabella 5.3: Dati relativi alla performance della rete neurale.

<i>Modello di stato</i>	X	Y	Z
Media [mm]	0,9089	0,6993	0,4281
Deviazione standard [mm]	0,7922	0,6155	0,3744
Massimo assoluto [mm]	13,5610	9,2570	6,4840

Tabella 5.4: Dati relativi alla performance del modello di stato.

Dall'osservazione di tali risultati si possono trarre le seguenti conclusioni:

- Sia le reti neurali sia i modelli a variabili di stato risultano essere in grado di apprendere la correlazione interno/esterno e di utilizzare tale correlazione per stima della posizione intera del tumore.

Tale affermazione è sostenuta dal fatto che entrambi gli strumenti proposti forniscono una stima sicuramente confrontabile con quella eseguita dal *modeler*, in quanto, in media, lo scostamento tra le due non è mai superiore ad un millimetro.

Inoltre bisogna notare che i valori massimi dello scostamento, seppure siano di entità non trascurabile, rappresentano casi isolati, come confermato dai valori medi la cui entità è sempre di almeno un ordine di grandezza inferiore a quella dei valori massimi.

- Tutti i valori ottenuti, per ogni parametro analizzato e per ogni coordinata, sono inferiori nel caso della rete neurale rispetto al modello di stato. Si può quindi affermare che, in riferimento ai casi considerati in questo studio preliminare, la rete neurale sia in grado di apprendere la correlazione interno/esterno più efficacemente del modello di stato.

In conclusione, alla luce di tali risultati preliminari, si è deciso di implementare in ambiente C/C++ un software di correlazione interno/esterno basato su reti neurali, la cui realizzazione è descritta nel prossimo paragrafo.

5.2. Realizzazione del software di correlazione

In questo paragrafo verrà presentato il software di correlazione interno/esterno che è stato realizzato in ambiente C/C++ per incontrare tutte le specifiche esposte nel paragrafo 3.5.

Come anticipato nel paragrafo precedente, tale software sarà basato sull'implementazione di reti neurali. Di seguito vengono sintetizzate le modalità con le quali si è lavorato.

- Il segnale interno considerato consiste direttamente nella posizione 3D del *target* rilevata dalle acquisizioni radiografiche (punti di *imaging*) e non più nella stima effettuata dal *modeler*. In questo modo, la correlazione calcolata risulta indipendente dalle modalità con le quali opera il sistema *Synchrony*. Si ricorda che, a differenza della stima effettuata dal *modeler*, l'acquisizione di immagini RX di controllo non avviene in tempo reale, bensì ad intervalli di circa 2 minuti.

All'interno di un singolo trattamento il numero di acquisizioni radiografiche varia tra 48 e 159 (campione di 20 frazioni). È quindi chiaro che, per disporre di un numero di dati sufficiente all'addestramento della rete neurale, è necessario, ipotizzando la stazionarietà del segnale, interpolare i punti di *imaging* al fine di ottenere un segnale interno con una frequenza maggiore.

- La stima del segnale interno viene effettuata utilizzando tre distinte reti neurali, ognuna delle quali ha come ingresso il segnale proveniente da un marcatore, mentre, ovviamente, il segnale interno considerato per l'addestramento è il medesimo. Il risultato complessivo è ottenuto semplicemente mediando le uscite delle tre reti. La decisione di utilizzare tre reti e non una singola come è stato fatto nello studio preliminare è stata presa con l'intenzione di sfruttare completamente l'informazione relativa al movimento della superficie corporea del paziente, dunque considerando tutti i tre marcatori disponibili. Di conseguenza, il segnale esterno considerato nell'ambito dello sviluppo del software di correlazione consiste nella posizione spaziale di tali marcatori.
- Per ogni frazione analizzata, viene considerata la sua intera durata e non solo un intervallo specificato. In un campione di 20 frazioni, tale durata varia tra 48 e 130 minuti circa, corrispondenti a 71234 e 195857 campioni del segnale esterno (acquisizioni di marcatori esterni).
- L'apprendimento iniziale della correlazione interno/esterno è stato effettuato considerando gli stessi punti di *imaging* che il sistema *Synchrony* utilizza per la costruzione del modello di correlazione precedentemente all'irradiazione. Anche il

numero di tali punti presenta una notevole variabilità, risultando compreso tra un minimo di 4 e un massimo di 27. Indipendentemente dalla numerosità di questi campioni, *il training set* è stato sempre costruito con 500 coppie ingresso/uscita ottenute tramite l'interpolazione dei suddetti punti e, ovviamente, dei campioni del segnale esterno corrispondenti temporalmente.

- Al fine di ottenere tutte le funzionalità richieste, il software che sarà descritto in questo capitolo sarà in grado di aggiornare la correlazione qualora risulti necessario. Inoltre sarà predisposto per acquisire i dati direttamente da un sistema di acquisizione ottico e per effettuare la predizione del segnale interno in tempo reale.

L'algoritmo implementato è stato sviluppato in tre versioni successive, ognuna della quali differisce dalla precedente in quanto comprende una funzionalità aggiuntiva. Si ricorda che le principali funzionalità richieste sono:

1. Costruire ed addestrare (tramite algoritmo di *backpropagation*) tre reti neurali. A seguito dell'addestramento iniziale, stimare il segnale interno in corrispondenza di ogni campione del segnale esterno.
2. Valutare l'errore di stima commesso da ogni rete in corrispondenza di ogni punto di *imaging* successivo a quelli utilizzati per l'addestramento iniziale. In funzione di tale valutazione, stabilire se procedere ad un nuovo addestramento delle reti.
3. Acquisire campioni del segnale esterno *online* (direttamente da un sistema di acquisizione ottica o da un suo simulatore). Eseguire la stima del segnale interno ed eventuali riaddestramenti *online*.

Nei seguenti tre paragrafi sarà descritto il codice sviluppato per realizzare ciascuna delle funzionalità appena citate.

5.2.1. Algoritmo di addestramento delle reti neurali

L'algoritmo per l'addestramento delle reti neurali è descritto schematicamente nel seguente schema a blocchi [Figura 5.15] ed è commentato successivamente.

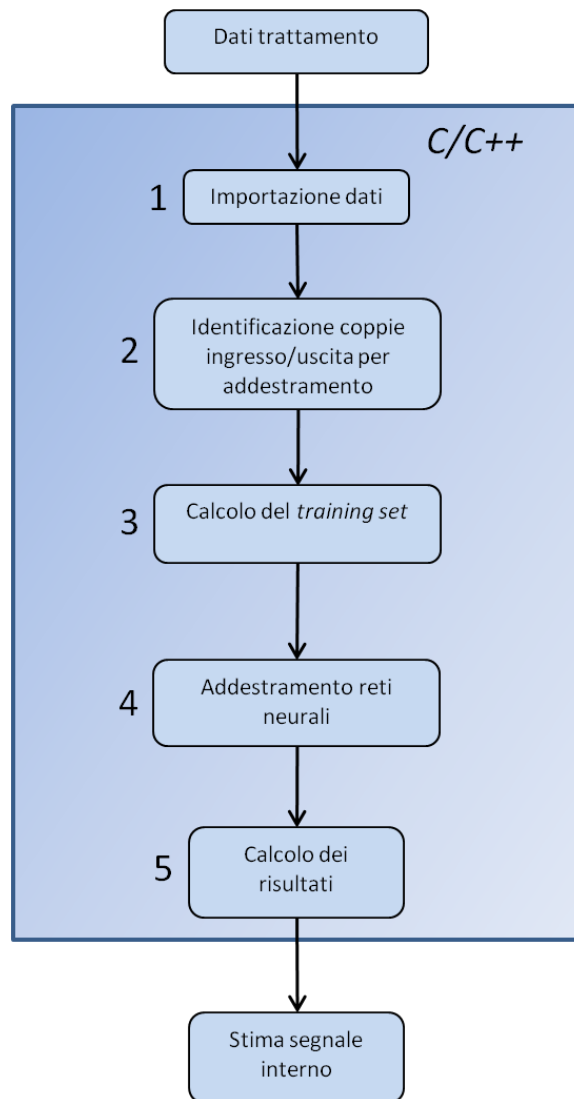


Figura 5.15: Schema a blocchi dell’algoritmo della prima versione del software di correlazione.

I dati in ingresso all’algoritmo sono stati ricavati dai file formato .log relativi al trattamento tramite *CyberKnife* e, operando in ambiente Matlab, sono stati raccolti all’interno di file formato testo (.txt). In particolare, tali file sono:

- *Markers.txt*: per ogni *timestamp* del trattamento, identifica la posizione tridimensionale dei tre marcatori esterni. Tali dati costituiscono il segnale esterno.
- *Imaging.txt*: in corrispondenza delle *timestamp* in cui viene acquisita un’immagine radiografica, identifica la posizione tridimensionale del bersaglio, dunque tali dati costituiscono il segnale interno utilizzato per l’addestramento delle reti neurali.

In aggiunta alle coordinate del *target* e alla corrispondente *timestamp*, tale file comprende, per ogni acquisizione, un indice di *status* che può assumere soltanto valori pari a 0 o 1. In particolare, *status*=0 significa che l’immagine corrispondente è stata utilizzata dal software del *CyberKnife* per la costruzione del modello di correlazione

precedentemente al trattamento, mentre *status*=1 indica che l'immagine corrispondente è stata utilizzata per l'aggiornamento del modello di correlazione durante il trattamento.

Fase 1: importazione dati

Come prima operazione, l'algoritmo importa i due file di testo appena descritti utilizzando la funzione "fscanf" della libreria "stdio.h" e salva i dati in essi contenuti nelle seguenti variabili di tipo *double* (64 bit) e di formato *array* (monodimensionale o bidimensionale):

- *Markers[num_mk][9]*: contiene le coordinate dei tre marcatori esterni per tutte le acquisizioni effettuate dal sistema di localizzazione ottico, il cui numero è identificato dalla costante *num_mk*.
- *Mk_tstamp[num_mk]*: contiene le *timestamp* corrispondenti a ciascuna acquisizione dei marcatori esterni.
- *Imaging[num_im][3]*: contiene le coordinate del bersaglio rilevate dal sistema di *imaging*. *Num_im* è un costante che identifica il numero di tali acquisizioni.
- *Im_tstamp[num_im]*: contiene le *timestamp* corrispondenti a ciascuna acquisizione radiografica.
- *Im_status[num_im]*: contiene lo *status* associato a ciascuna acquisizione radiografica.

A seguito dell'importazione di tali dati, l'algoritmo è in grado di svolgere tutte le successive operazioni senza la necessita di ulteriori informazioni provenienti da fonti esterne al codice.

Fase 2: identificazione coppie ingresso/uscita

Al fine di ottenere un *training set* con il quale addestrare le reti neurali, l'algoritmo deve identificare le coppie ingresso/uscita da utilizzare come esempi. In particolare tali coppie sono costituite da un campione del segnale interno (3 coordinate) e dal corrispondente campione (in termini di *timestamp*) del segnale esterno (9 coordinate, 3 per ogni marcatore).

Come già anticipato, in questa prima versione del software, viene eseguito soltanto l'addestramento iniziale delle reti neurali, ovvero quello effettuato considerando i campioni del segnale interno aventi *status*=0.

Le coppie sono state ottenute considerando ogni campione della variabile *Imaging* e, qualora lo *status* fosse pari a 0, individuando nella variabile *Mk_tstamp* la *timestamp* coincidente a quella del campione di *Imaging*. In tal modo, si ottiene l'indice del corrispondente campione all'interno variabile *Markers*. La Figura 5.16 mostra schematicamente come viene eseguita tale associazione.

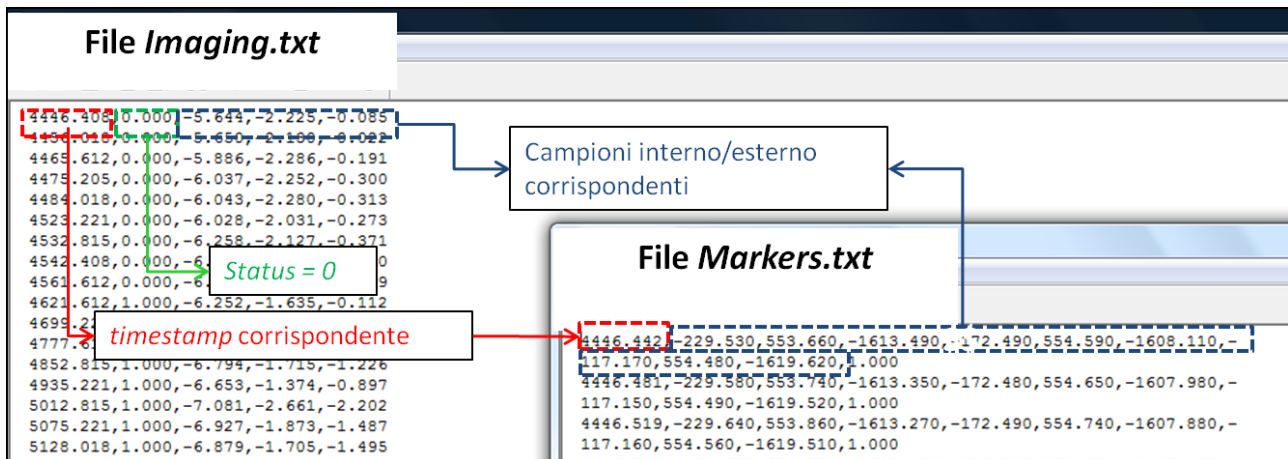


Figura 5.16: Procedura per l'identificazione di una coppia ingresso/uscita.

Al termine di tale procedura si hanno a disposizione due variabili, “im_train[num_train_iniziale][3]” e “mk_train[num_train_iniziale][9]”, dove num_train_iniziale è un costante pari al numero di campioni di Imaging aventi *status*=0.

Fase 3: calcolo del training set

Come già anticipato, le coppie ingresso/uscita così ottenute non possono essere direttamente utilizzate come *training set* per l'addestramento delle reti neurali, dal momento che il loro numero è troppo ridotto. È quindi necessario stimare ulteriori esempi tramite l'interpolazione dei campioni contenuti nelle variabili “im_train” e “mk_train”. Si ricorda che per fare questo è necessario ipotizzare che i segnali considerati siano stazionari, ovvero che le loro caratteristiche statistiche siano invariante rispetto al tempo.

Per eseguire l'interpolazione (di tipo *spline*) sono state utilizzate due apposite funzioni:

- void spline(...): riceve in ingresso i dati non interpolati sotto forma di vettore e calcola i tre coefficienti che definiscono la funzione *spline* interpolante.
- void seval(...): riceve in ingresso i tre coefficienti appena calcolati ed esegue la valutazione della funzione interpolante con la frequenza prestabilita. Più precisamente, tale frequenza viene modulata in modo tale da ottenere sempre 500 valutazioni indipendentemente dal numero iniziale di campioni. A tal fine il periodo di campionamento della funzione interpolante (τ) viene calcolato nel seguente modo:

$$\tau = (\text{num_train_iniziale} - 1)/500 \quad (5.13)$$

Applicando le due funzioni appena descritte a ciascuna colonna delle variabili “im_train” e “mk_train”, si ottengono due nuove variabili “im_train_interp[500][3]” e “mk_train_interp[500][9]”, contenenti appunto 500 esempi delle uscite e degli ingressi rispettivamente.

Si è scelto di ottenere proprio 500 coppie dal momento che nello studio preliminare descritto nel capitolo 5.1 si è osservato che tale numerosità è sufficiente per l'apprendimento della correlazione interno/esterno da parte di una rete neurale, consentendo, allo stesso tempo, di ottenere tempi di addestramento accettabili.

Prima di procedere all'addestramento, è necessario comprimere la dinamica dei dati contenuti nelle variabili "im_train_interp" e "mk_train_interp", in modo tale che queste assumano valori compresi tra -1 e +1. A tal scopo, per ogni colonna di tali variabili è stato calcolato il valore minimo e il valore massimo. Successivamente è stata applicata la trasformazione riportata nell'equazione 5.14.

$$x' = -1 + \frac{x - \min}{(\max - \min)/2} \quad (5.14)$$

dove, x indica il dato originale e x' il dato con dinamica compresa tra -1 e +1.

In ultimo, ricordando che si vogliono addestrare tre distinte reti neurali, ognuna avente in ingresso il segnale proveniente da un solo marcatore esterno, la variabile "mk_train_interp" è stata suddivisa in tre variabili separate di dimensione 500x3, chiamate "mk1_train_interp", "mk2_train_interp" ed "mk3_train_interp".

Fase 4: addestramento reti neurali

Le tre variabili appena citate, assieme alla variabile "im_train_interp", costituiscono un *training set* con il quale è possibile addestrare una singola rete neurale.

Come imposto dal problema che si vuole affrontare, le reti hanno tre ingressi (le coordinate di un marcatore esterno) e tre uscite (le coordinate stimate del bersaglio).

Riguardo all'architettura, il software sviluppato prevede un certo grado di flessibilità. In particolare l'utente può agire sull'architettura delle reti modificando il valore di due costanti chiamate "strato_uscita" e "num_hidden". La Tabella 5.5 riporta l'architettura ottenuta in funzione del valore di tali costanti.

strato_uscita	num_hidden	Strato nascosto	Strato d'uscita
0	0	Assente	3 neuroni, f.d.t lineare
0	$n > 0$	n neuroni, f.d.t tanh	3 neuroni, f.d.t lineare
1	0	Assente	3 neuroni, f.d.t tanh
1	$n > 0$	n neuroni, f.d.t tanh	3 neuroni, f.d.t tanh

Tabella 5.5: Possibili architetture delle reti neurali in funzione delle costanti definite dall'utente.

Come si osserva, l'utente può escludere lo stato nascosto ponendo num_hidden=0 mentre, qualora si decida di includerlo, questo può essere composto da un numero arbitrario di neuroni

(n) la cui funzione d'attivazione è sempre la tangente iperbolica (\tanh). In aggiunta è possibile definire la funzione d'attivazione dei neuroni dello strato d'uscita: se si impone `strato_uscita=0` tale funzione risulta essere lineare, altrimenti se si impone `strato_uscita=1` allora questa risulta essere la tangente iperbolica.

Una volta stabilita l'architettura, questa viene applicata a tutte le tre reti neurali, che risultano quindi identiche.

Seppure non sia stato condotto uno studio volto ad ottimizzare la configurazione delle reti, lo studio preliminare ha dimostrato che una rete *feed-forward* con singolo strato nascosto è in grado di apprendere la correlazione interno/esterno, dunque non è stata inclusa la possibilità di realizzare reti neurali aventi maggiore complessità.

Per eseguire l'addestramento delle reti sono state utilizzate le seguenti funzioni:

- `void init_weights()`: esegue un'inizializzazione random dei pesi che caratterizzano le interconnessioni tra neuroni.
- `void calcnet(...)`: riceve in ingresso un campione del segnale esterno (le 3 coordinate di un marcatore) e le matrici contenenti i pesi, quindi calcola l'uscita della rete in accordo la sua architettura. Quando viene utilizzata in fase di addestramento, fornisce anche la differenza tra il valore calcolato e l'uscita desiderata.
- `void WeightChanges(...)`: riceve in ingresso la differenza tra uscita calcolata e uscita desiderata e, in base a tale dato, modula i valori dei pesi. L'algoritmo utilizzato a tal fine è quello della *back propagation*, applicato come è stato descritto nel paragrafo 4.1.2.
- `void calcOverallError(...)`: al termine di ogni epoca di addestramento, calcola l'errore complessivo (RMS) su tutti gli esempi contenuti nel *training set*.

L'algoritmo utilizzato per l'addestramento di una singola rete neurale è descritto nel seguente schema a blocchi [Figura 5.17] e commentato di seguito.

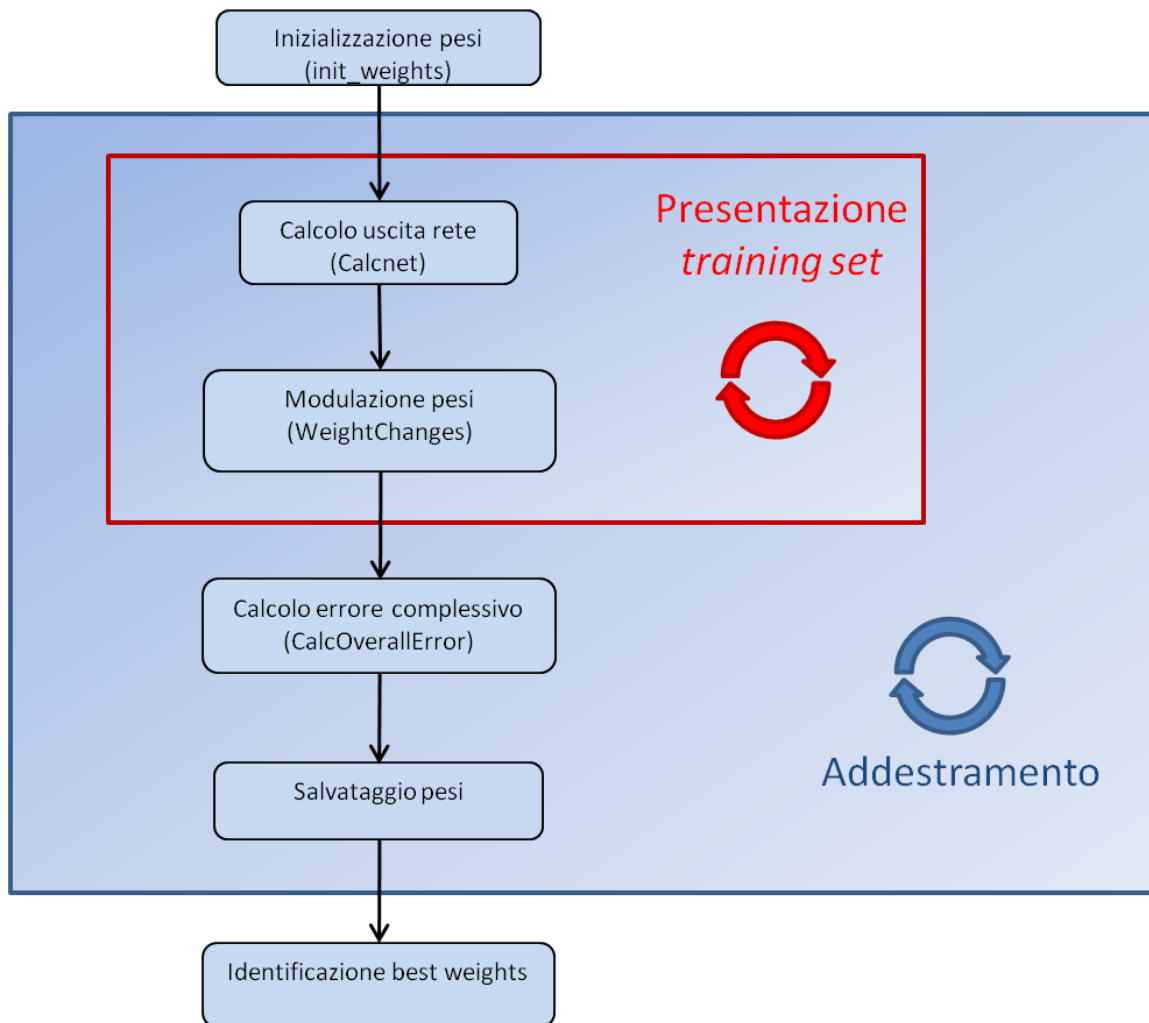


Figura 5.17: Algoritmo di addestramento di una singola rete neurale.

La prima operazione eseguita dall'algoritmo consiste nell'inizializzazione random dei pesi tramite la funzione "init_weights".

Viene quindi avviato il ciclo di addestramento (riquadro blu), le cui successive iterazioni corrispondono a successive epoche. All'interno di tale ciclo, vi è un secondo ciclo (riquadro rosso) che, per ogni epoca, presenta alla rete tutti gli esempi contenuti nel *training set*. Dunque, ad ogni iterazione di quest'ultimo ciclo, alla rete viene presentata una singola coppia ingresso/uscita, in base alla quale vengono modulati i pesi grazie alle funzioni "Calcnet" e "WeightChanges". Terminato tale ciclo, la funzione "calcOverallError" calcola l'errore complessivo (RMS), in termini di differenza tra uscita calcolata e uscita desiderata, su tutti gli esempi del *training set*, mentre i pesi ottenuti vengono salvati in una variabile.

Il ciclo di addestramento termina quando viene raggiunto il numero di epoche specificato dall'utente oppure quando l'RMS associato all'ultima epoca risulta inferiore ad un valore di soglia, anche quest'ultimo stabilito dall'utente.

Al termine di quest'ultimo ciclo, l'algoritmo identifica l'epoca nella quale si è registrato il minimo valore di RMS, ovvero l'epoca in cui la rete ha fornito la performance migliore. I pesi relativi a tale epoca (*best weights*) costituiscono l'uscita dell'intera procedura di addestramento. Si ricorda che tale algoritmo viene ripetuto per ciascuna delle tre reti neurali.

Fase 5: calcolo dei risultati

A seguito dell'addestramento, il software è in grado di stimare il segnale interno in funzione di un dato segnale esterno.

Per eseguire tale calcolo, la funzione "Calcnet" utilizza i *best weights* appena ottenuti per stimare il segnale interno in corrispondenza di ogni acquisizione di un marcatore esterno.

Il calcolo viene eseguito per tutte le tre reti neurali, ognuna della quali stima la posizione del *target* in funzione della posizione di un marcatore indipendentemente dalle altre due reti. Successivamente le tre uscite, la cui dinamica è compresa tra -1 e +1 (y'), vengono trasformate in modo da essere ricondotte alla dinamica originale (y), come mostrato in equazione 5.15:

$$y = \min + (y' + 1)[(\max - \min)/2] \quad (5.15)$$

dove i valori *min* e *max* sono relativi alle uscite desiderate del *training set*.

In ultimo, al fine di ottenere un'unica stima del segnale interno, le uscite delle tre reti vengono mediate tra di loro. Il valore così ottenuto viene scritto su un file di testo, grazie alla funzione "fprintf".

A titolo di esempio, la Figura 5.18 mostra la stima effettuata dal software della coordinata x (direzione cranio caudale, CC) del *target* in corrispondenza di ogni acquisizione dei marcatori esterni.

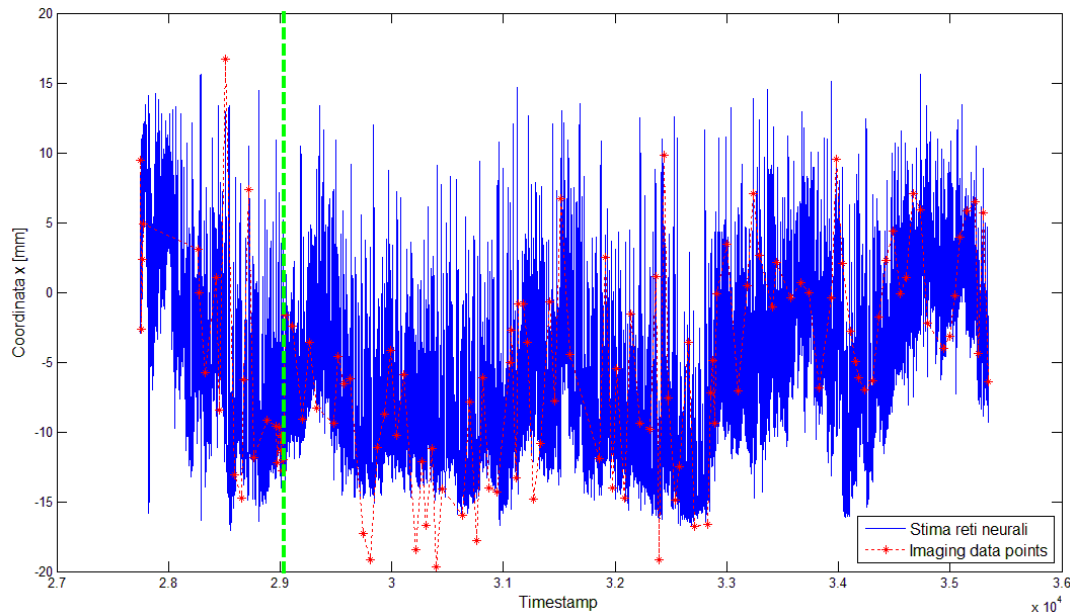


Figura 5.18: In blu è riportata la stima della coordinata x del bersaglio eseguita dal software. I punti rossi indicano le effettive coordinate x del bersaglio in corrispondenza delle acquisizioni radiografiche. I campioni di *imaging* a sinistra della linea verde sono quelli utilizzati per l'addestramento.

Si può notare che il software si dimostra in grado di stimare la posizione del bersaglio a seguito dell'addestramento eseguito soltanto sui campioni di Imaging aventi *status*=0 (a sinistra della linea verde). Si osserva, infatti, che la stima eseguita ha un andamento confrontabile con quello mostrato dalla reale coordinata x del *target*.

Si nota anche che, in presenza di campioni di Imaging la cui coordinata x sia esterna al *range* di variabilità dei campioni con i quali si è eseguito l'addestramento, la stima risulta poco accurata. In conclusione, i risultati inerenti alle stime effettuate utilizzando l'algoritmo appena descritto evidenziano che lo strumento sviluppato è adeguato al problema che si vuole affrontare ma, al fine di ottenere una stima più accurata in corrispondenza dei campioni successivi a quelli utilizzati per l'addestramento iniziale, risulta necessario includere la possibilità di riaddestrare le reti in corrispondenza delle acquisizioni radiografiche di controllo che vengono eseguite durante il trattamento.

5.2.2. Valutazione dell'errore di stima

La seconda funzionalità richiesta all'algoritmo, che si ricorda consistere nella capacità di valutare l'errore di stima ed eventualmente di riaddestrare le reti neurali, è stata implementata ampliando l'algoritmo descritto nel paragrafo precedente. Di conseguenza, il nuovo codice comprende l'importazione dei dati, la costruzione del *training set*, l'addestramento iniziale e il calcolo dei risultati a seguito di quest'ultimo, esattamente come precedentemente descritto.

A seguito di questi passaggi, il software include un ciclo “for” di simulazione [Figura 5.19], le cui principali fasi vengono descritte in seguito. Si precisa che tale ciclo viene definito “di simulazione” poiché ogni sua iterazione fornisce un nuovo campione in ingresso alle reti neurali, esattamente come se questi provenissero da un sistema di localizzazione ottica.

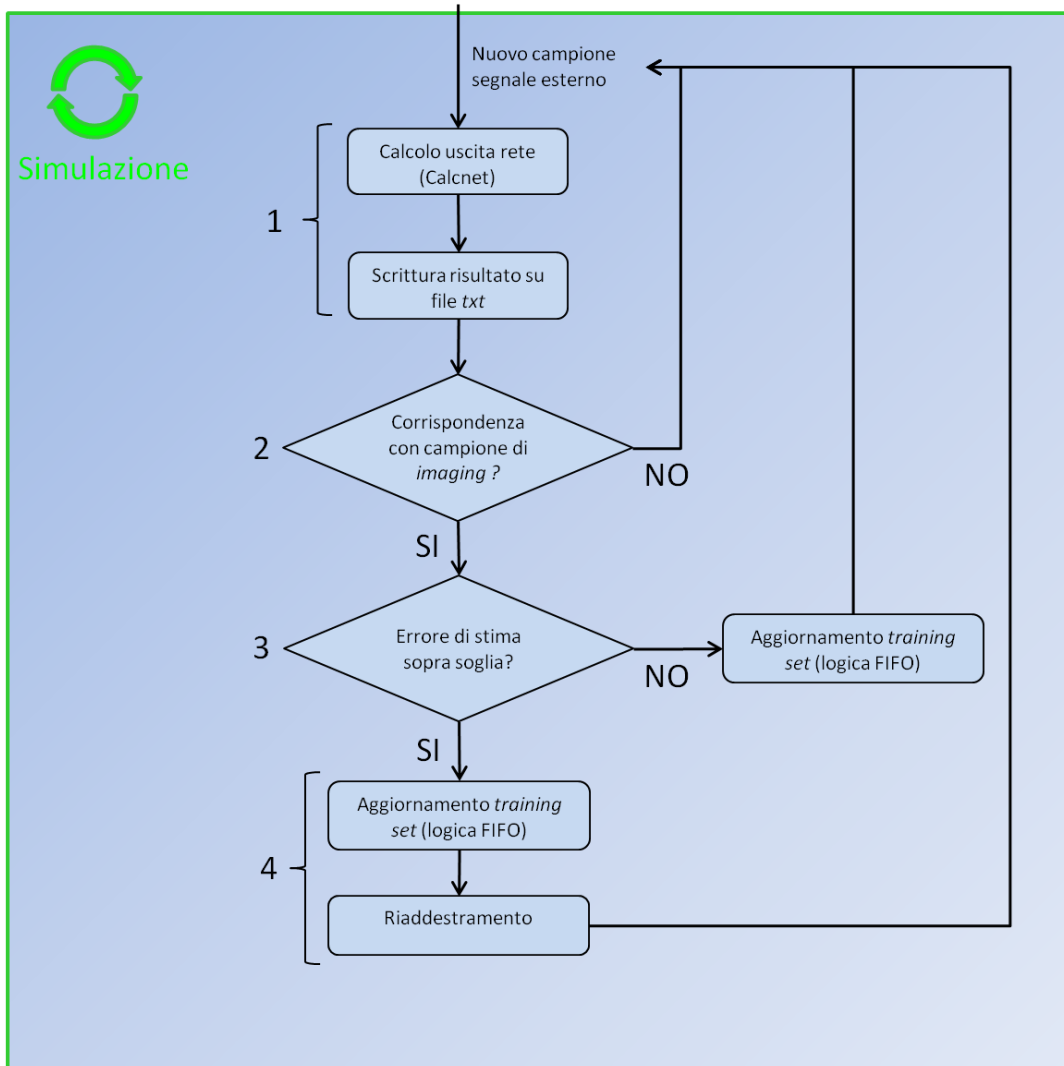


Figura 5.19: Diagramma di flusso del ciclo di simulazione.

Fase 1: inizializzazione e calcolo iniziale della stima

Il ciclo di simulazione, ad ogni sua iterazione, fornisce in ingresso alle reti neurali un nuovo campione del segnale esterno, dunque esegue tante iterazioni quanti sono i campioni disponibili di tale segnale, relativamente all’intera durata del trattamento considerato.

Ad ogni nuovo campione, ogni rete neurale calcola la stima del segnale interno tramite la funzione “Calcnet”. Le uscite delle tre reti vengono riportate alla loro dinamica originale (equazione 5.15) e successivamente mediate tra di loro per ottenere un’unica stima. Quest’ultima viene registrata in un file formato testo.

Fase 2: verifica corrispondenza con campione di *imaging*

In questa fase, il codice verifica se, in corrispondenza del campione corrente del segnale esterno, vi è una acquisizione radiografica di controllo, ovvero un'immagine acquisita dal sistema *CyberKnife* durante l'irradiazione allo scopo di verificare l'accuratezza delle stime effettuate e, se si dimostra necessario, per aggiornare il modello di correlazione. Si ricorda che tali immagini vengono acquisite successivamente a quelle utilizzate per la costruzione del modello e sono identificate dall'indice di *status* pari a 1.

Anche in questo caso, la corrispondenza viene rilevata sulla base delle *timestamps* che, come già detto, costituiscono una comune base temporale sia per il segnale esterno che per il segnale interno.

L'algoritmo, ad ogni nuovo campione del segnale esterno, legge la *timestamp* corrispondente e verifica se questa è presente anche nella variabile "Im_tstamp", ovvero tra le *timestamps* associate ai campioni di *imaging*. Al fine di migliorare l'efficienza dell'algoritmo, la verifica della corrispondenza in quest'ultima variabile inizia dall'indice successivo all'ultima immagine individuata.

Nel caso in cui venga individuata la corrispondenza, il codice salva in una variabile la coppia interno/esterno costituita dalle coordinate correnti dei tre marcatori esterni e dalle coordinate del *target* rilevate nell'immagine di controllo appena individuata. Inoltre, all'utente viene notificata la corrispondenza riportando a video gli indici del campione del segnale esterno e il campione di *imaging*.

Al contrario, se non venisse rilevata alcuna corrispondenza non vengono eseguite ulteriori operazioni nella corrente iterazione.

Fase 3: valutazione errore di stima

Nel caso in cui sia stata rilevata una corrispondenza, è possibile quantificare l'errore di stima commesso dalle reti, dal momento che si ha a disposizione la posizione vera del bersaglio rilevata dal sistema di *imaging*.

Poiché la stima della posizione del *target* viene effettuata parallelamente da tre distinte reti, l'errore di stima viene valutato separatamente per ognuna di esse. Di seguito è esposto il criterio di valutazione, che ovviamente, è applicato alle tre reti in modo identico.

Sia S la stima effettuata da una rete neurale della posizione del bersaglio e sia P la posizione effettiva di quest'ultimo. Con tali dati a disposizione viene calcolata la cifra D (equazione 5.16), la quale viene confrontata con due valori di soglia imposti dall'utente (α e β).

$$\mathbf{S} = [S_x \ S_y \ S_z]; \mathbf{P} = [P_x \ P_y \ P_z]$$

$$D = \max(|S_x - P_x|, |S_y - P_y|, |S_z - P_z|) \quad (5.16)$$

$$D > \alpha; D > \beta; \text{ con } \beta \geq \alpha$$

Come si osserva, la cifra D viene ottenuta considerando il massimo tra le differenze in valore assoluto eseguite, coordinata per coordinata, tra la stima effettuata dalla rete e la posizione vera del bersaglio.

La soglia α è utilizzata per valutare se sia necessario o meno eseguire un nuovo addestramento della rete: nel caso in cui si abbia $D > \alpha$, la rete viene riaddestrata.

La soglia β deve essere posta maggiore o uguale ad α in quanto viene utilizzata per valutare se riaddestrare la rete a partire dai pesi precedentemente calcolati o se riaddestrare a partire da nuovi pesi inizializzati casualmente. In quest'ultimo caso ($D > \beta$) viene completamente annullata l'esperienza acquisita dalla rete negli addestramenti precedenti ed è quindi chiaro che tale procedura vada applicata soltanto quando l'errore di stima risulta particolarmente elevato. Nonostante questa sia una eventualità poco frequente e che si vuole limitare, è possibile che questo si verifichi in corrispondenza di rapide variazioni della correlazione interno/esterno, dovute ad esempio ad artefatti quali il movimento dei marcatori esterni relativamente alla superficie corporea del paziente.

Il diagramma di flusso in Figura 5.20 riassume quanto è stato appena esposto.

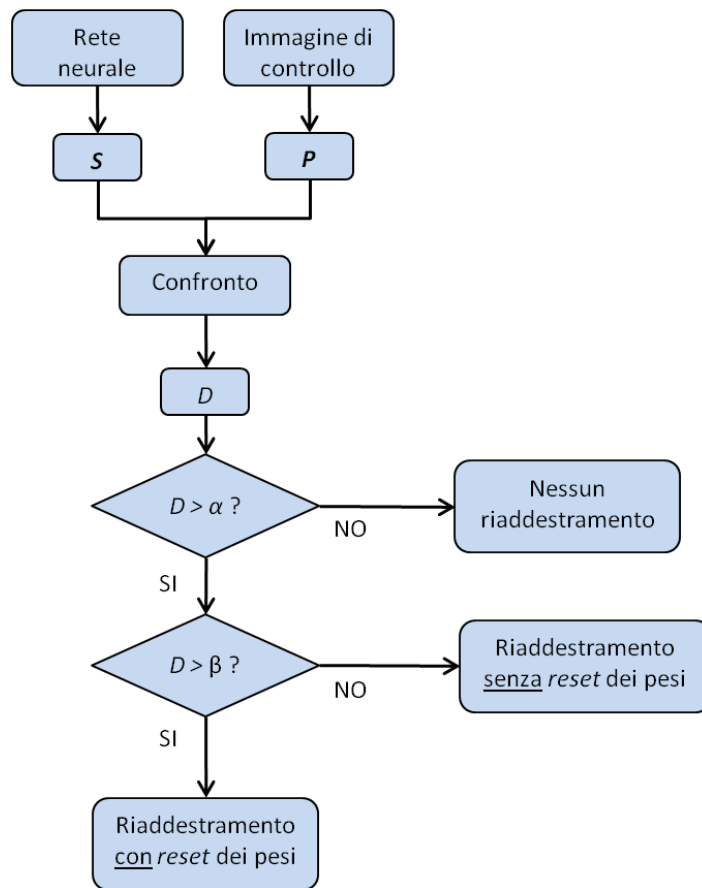


Figura 5.20: Diagramma di flusso della valutazione dell'errore di stima.

Nelle prove effettuate su 20 trattamenti, la soglia α è stata posta pari a 1 mm mentre la soglia β è stata posta a 2 mm. Tali valori sono stati scelti in quanto, nei test svolti, garantiscono un numero di riaddestramenti sufficiente a mantenere aggiornata la correlazione interno/esterno senza causare *overfitting*.

Nel caso particolare in cui la correlazione interno/esterno risultasse particolarmente instabile nel tempo, è possibile porre $\alpha = \beta$, in modo tale che ad ogni riaddestramento i pesi sia sempre inizializzati casualmente.

Fase 4: aggiornamento *training set* e riaddestramento

Al termine della valutazione dell'errore di stima, indipendentemente dalla decisione presa relativamente al riaddestramento, il codice aggiorna il *training set* utilizzando una logica *First In First Out* (FIFO). Ciò significa che ad ogni nuovo campione di *imaging* incontrato nel corso del ciclo di simulazione, la nuova coppia interno/esterno viene inserita nel *training set*, mentre la coppia meno recente (in termini di *timestamps*) viene esclusa. In tal modo la numerosità delle coppie utilizzate per gli addestramenti successivi a quello iniziale rimane invariata nel corso della simulazione. Tale numerosità può essere impostata dall'utente in modo che coincida con il numero di coppie impiegate nell'addestramento iniziale o che risulti inferiore a tale numero.

Si è deciso di includere questa flessibilità dal momento che può risultare utile effettuare addestramenti successivi con un numero di coppie inferiore a quello utilizzato nell'addestramento iniziale, così che la nuova informazione appresa dalla rete risulti più localizzata temporalmente. Questo permette di cogliere in maggiore dettaglio l'andamento locale della correlazione interno/esterno.

È chiaro che l'aggiornamento del *training set* deve essere necessariamente eseguito sia nel caso in cui si decida di riaddestrare la rete sia nel caso in cui si decida di continuare la simulazione senza riaddestrare. In quest'ultimo caso, infatti, se non si aggiornasse comunque il *training set*, eventuali riaddestramenti eseguiti successivamente nella simulazione si baserebbero *training set* in cui solo l'ultima coppia è recente mentre le altre non sono immediatamente precedenti.

Eseguito l'aggiornamento del *training set*, l'algoritmo procede all'eventuale riaddestramento delle reti neurali. Tale procedura avviene esattamente come descritto per l'addestramento iniziale (paragrafo 5.2.1) e può interessare soltanto una rete, due reti su tre o tutte le tre reti, a seconda dell'errore di stima commesso da ciascuna di esse. Nel caso in cui sia previsto il *reset* dei pesi per una determinata rete, l'algoritmo elimina i pesi precedentemente calcolati e li sostituisce con valori casuali.

Terminati gli eventuali riaddestramenti, la simulazione viene riavviata con in ingresso il successivo campione del segnale esterno.

Allo scopo di fornire una dimostrazione qualitativa dell'utilità del riaddestramento delle reti neurali, è riportata la Figura 5.21.

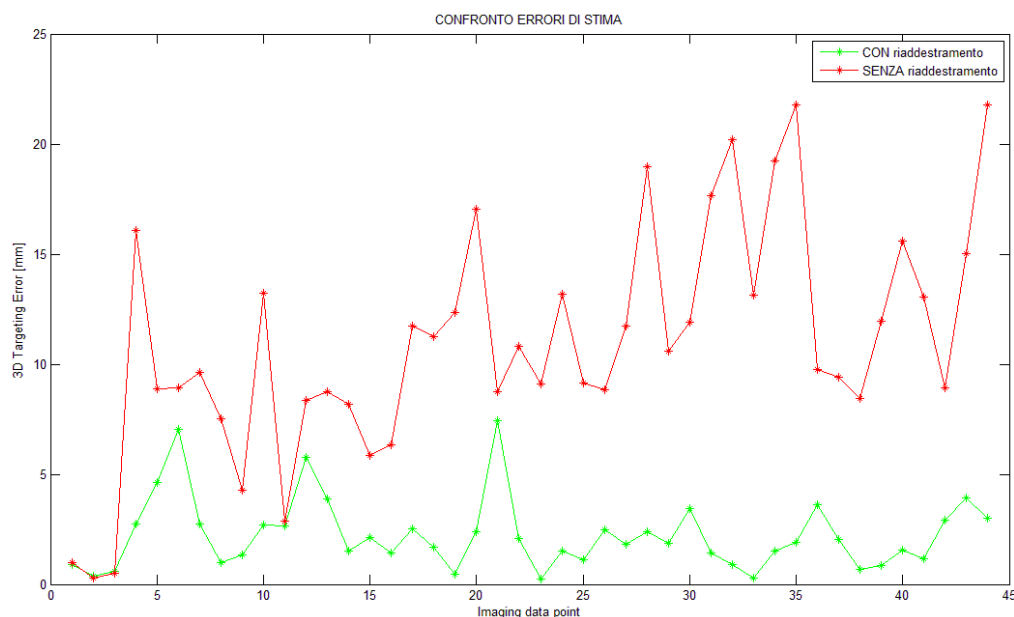


Figura 5.21: Confronto tra l'errore di stima, espresso come errore di puntamento 3D, commesso dalle reti neurali. Il tracciato rosso si riferisce alle stime eseguite con reti che non vengono mai riaddestrate, il tracciato verde, invece, si riferisce a reti che possono essere riaddestrate nel corso della simulazione.

In tale figura si nota come i riaddestramenti successivi a quello iniziale abbiano l'effetto di diminuire visibilmente lo scostamento tra la posizione stimata del bersaglio e quella effettiva, come testimoniato dall'andamento dell'errore di stima 3D relativo ai calcoli effettuati senza riaddestramenti delle reti neurali successivi a quelli iniziali, che risulta sempre maggiore di quello relativo alle stime ottenute consentendo il riaddestramento delle reti.

5.2.3. Acquisizione dei dati e calcolo della stima *online*

Le funzionalità fin ad ora descritte sono state integrate in un nuovo algoritmo volto a soddisfare la terza specifica richiesta, ovvero al capacità di acquisire i dati e di effettuare le stime *online*.

Prima di descrivere come tale nuovo algoritmo è stato realizzato è opportuno chiarire come si è proceduto.

- I dati in ingresso non vengono più importati da file di testo all'inizio dell'esecuzione dell'algoritmo. Tale procedura, infatti, risultava adeguata nelle precedenti versioni del software, il cui obiettivo principale era l'implementazione delle reti neurali. In questa terza versione, invece, l'obiettivo principale è quello di realizzare uno strumento che sia in grado di ricevere dati direttamente da un sistema di acquisizione ottica (o da un suo simulatore) e, in funzione dell'ingresso corrente, di stimare *online* la posizione del *target*.
- Conseguentemente a quanto detto nel punto precedente, i dati relativi ai trattamenti con *CyberKnife* non vengono più utilizzati, dal momento che, ovviamente, non possono essere acquisiti *online*.
- Nell'ambito di questo studio, l'esigenza di acquisire i dati *online* può essere soddisfatta solamente per un segnale proveniente da un sistema di localizzazione ottica, mentre non è soddisfabile per il segnale interno, la cui acquisizione richiederebbe l'utilizzo di un sistema di *imaging* radiografico. Per questo motivo il segnale interno è stato simulato tramite l'acquisizione di un ulteriore marcatore esterno, la cui posizione viene registrata dall'algoritmo ad una frequenza imposta dall'utente.

La Figura 5.22 riporta un'illustrazione schematica del sistema realizzato.

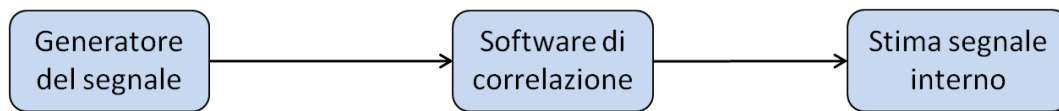


Figura 5.22: Schema del sistema impiegato per la stima *online* del segnale interno.

Il generatore del segnale può essere costituito da un sistema di localizzazione ottica o da un suo simulatore. In particolare, nell'ambito di questo studio, sono stati utilizzati i seguenti due sistemi:

- Sistema di localizzazione ottica SMART-D (BTS S.p.A.) [64]. Consiste in un sistema di localizzazione ad alta risoluzione, basato su telecamere con sensori CCD ad elevata sensibilità. Tale sistema è stato selezionato per la realizzazione dell'*Optical Tracking System* (OTS) da installare nelle sale di trattamento del CNAO.
- SMART *Virtual Machine*. Consiste in un software che simula una acquisizione tramite SMART-D leggendo un file (formato .tdf) precedentemente registrato. Tale file contiene, per ogni campione, l'indice corrispondente e la posizione spaziale di tutti i marcatori acquisiti. La frequenza di lettura coincide esattamente con la frequenza di acquisizione del sistema SMART (60 Hz).

I dati trasmessi dal generatore consistono nelle posizioni spaziali di tre marcatori (segnale esterno) più la posizione di un ulteriore marcatore che, come già anticipato, simula il segnale interno.

Per quanto riguarda la trasmissione dei dati dal generatore al software di correlazione, questa avviene in modo differente a seconda del generatore utilizzato.

- La SMART *Virtual Machine* comunica con il software di correlazione attraverso un protocollo UDP (*User Datagram Protocol*). Tale protocollo, collocato al livello di trasporto nel modello ISO/OSI, consente lo scambio di dati tra due algoritmi che vengono eseguiti contemporaneamente. In particolare, i dati provenienti dal generatore del segnale vengono scritti su una porta UDP appositamente aperta. Il software di correlazione è in grado di connettersi a tale porta e di leggere da quest'ultima i dati.
- Nel caso in cui si utilizzi direttamente il sistema di localizzazione ottica SMART-D, esso dispone di un software in linguaggio C che acquisisce la posizione di tutti i marcatori individuati dalle telecamere e, *frame per frame*, la salva in apposite variabili. È quindi chiaro che, in questo caso, non è necessario utilizzare alcuna porta UDP, in quanto il software di correlazione può accedere direttamente alle suddette variabili.

In entrambi i casi, ogni dato disponibile viene inviato in tempo reale all'algoritmo di correlazione, che lo elabora al fine di ottenere la stima del segnale interno. La posizione di tutti i quattro marcatori (tre per il segnale esterno e uno rappresentante il segnale interno) viene acquisita e trasmessa al software di correlazione alla stessa frequenza. Quest'ultimo legge i dati relativi al segnale esterno esattamente alla stessa frequenza con cui vengono trasmessi, mentre la posizione del quarto marcatore viene forzatamente registrata ad una frequenza inferiore. Di seguito verrà descritto l'algoritmo del software di correlazione, riportato nel diagramma di flusso mostrato in Figura 5.23.

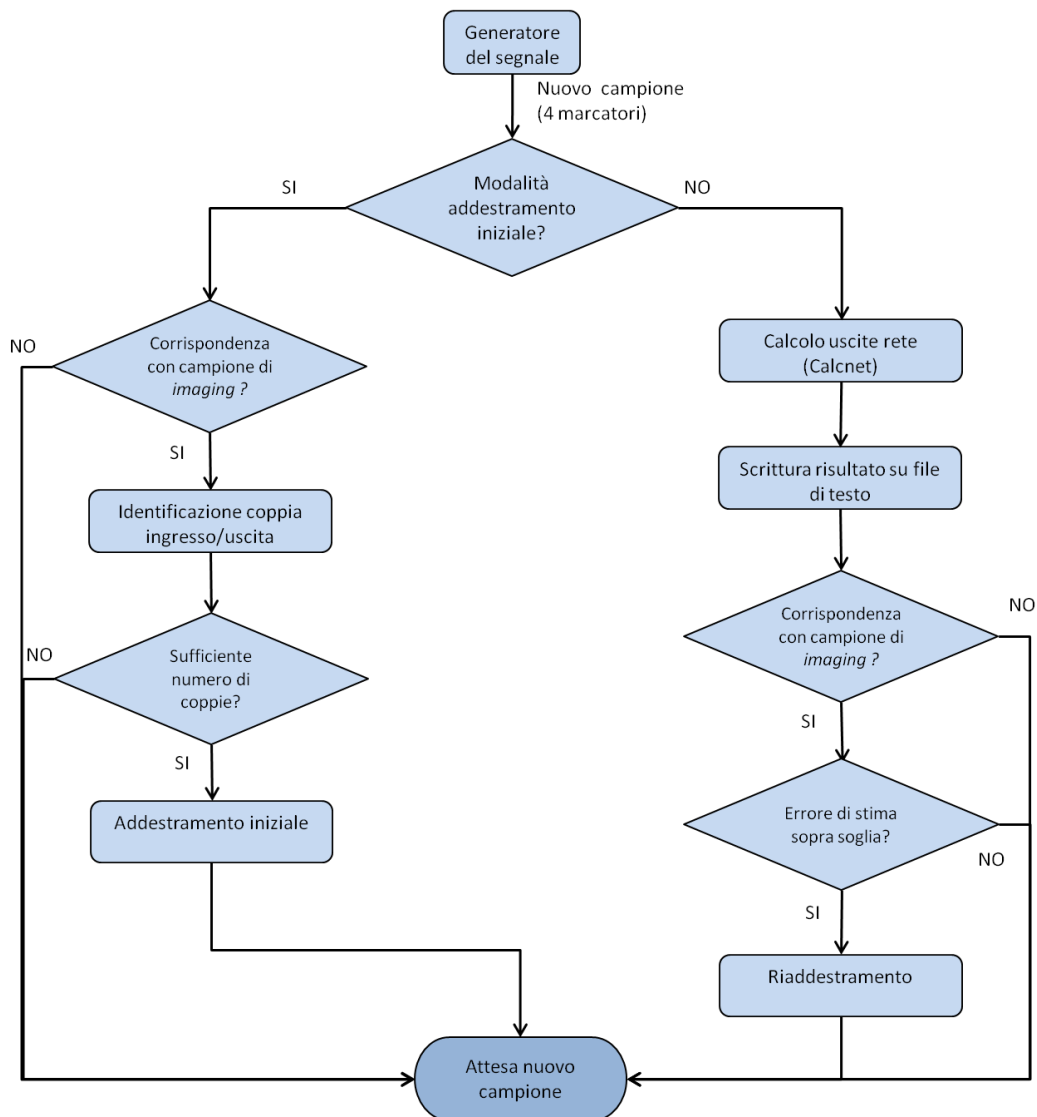


Figura 5.23: Diagramma di flusso dell'algoritmo del software di correlazione.

L'algoritmo riceve in ingresso i dati relativi alla posizione di quattro marcatori e può procedere in due modalità differenti: modalità di addestramento iniziale (ramo di sinistra del diagramma di flusso) o in modalità di esecuzione normale (ramo di destra del diagramma di flusso).

Modalità di addestramento iniziale

Questa modalità dell'algoritmo viene avviata in corrispondenza del primo campione ricevuto e ha come scopo quello di raccogliere un sufficiente numero di coppie interno/esterno sulle quali eseguire l'addestramento iniziale delle tre reti. In questa fase non viene calcolata alcuna stima del segnale interno.

La prima operazione effettuata in questa fase consiste nell'identificazione della corrispondenza temporale tra segnale esterno e segnale interno (quarto marcatore). A tal fine l'utente stabilisce la frequenza con la quale acquisire quest'ultimo marcatore, la cui posizione viene quindi registrata ad intervalli di tempo regolari. Di conseguenza, la corrispondenza si ha nell'istante in cui, oltre ai tre marcatori che compongono il segnale esterno, viene registrata anche la posizione del quarto marcatore.

Nel caso in cui la corrispondenza non venisse rilevata, l'algoritmo termina e viene riavviato solo in presenza di un nuovo campione. Se, invece, la corrispondenza viene rilevata, allora la coppia ingresso/uscita viene inserita nel *training set* iniziale.

Successivamente il codice verifica il numero di coppie raccolte e, se questo corrisponde al numero stabilito dall'utente, viene eseguito l'addestramento iniziale delle tre reti, esattamente come descritto nel paragrafo 5.2.1. Se, invece, il numero di coppie raccolte è inferiore a quello stabilito, l'algoritmo termina e attende un nuovo campione in ingresso.

Al termine dell'addestramento iniziale, una variabile di tipo logico (*bool*) viene impostata a *true*, in modo tale che, dalla successiva iterazione (ovvero a partire dal successivo campione in ingresso) l'algoritmo venga eseguito in modalità di esecuzione di normale e non più in modalità addestramento iniziale.

Modalità di esecuzione normale

Tale modalità procede in modo analogo al ciclo di simulazione realizzato nella seconda versione del software (paragrafo 5.2.2), con l'unica differenza che le successive iterazioni non vengono imposte da ciclo di tipo "for" bensì si ha una nuova iterazione in corrispondenza di ogni nuovo campione in ingresso, in funzione del quale viene stimato *online* il segnale interno.

L'algoritmo, come prima operazione, stima il segnale interno in funzione del segnale esterno utilizzando le reti precedentemente addestrate. Il risultato viene trascritto su un file formato testo.

Successivamente viene rilevata l'eventuale corrispondenza con un campione di *imaging*, esattamente come viene effettuato in modalità di addestramento iniziale.

Anche in questo caso, la frequenza con cui viene registrata la posizione del quarto marcatore viene impostata dall'utente e può essere diversa da quella utilizzata in modalità di addestramento iniziale. In tal modo è possibile simulare realisticamente le modalità con le quali il sistema *CyberKnife* acquisisce le immagini radiografiche. La frequenza di acquisizione di tali immagini, infatti, risulta maggiore nella fase iniziale del frazione di trattamento, ovvero quando il sistema deve costruire il modello di correlazione, mentre diminuisce quando le immagini vengono acquisite durante l'irradiazione a scopo di controllo.

Nel caso in cui non venisse riscontrata la corrispondenza con un campione di *imaging*, l'iterazione corrente termina e l'algoritmo viene riavviato solo nel momento in cui si rende disponibile un nuovo campione del segnale esterno. Se, invece, la corrispondenza viene rilevata, l'algoritmo calcola l'errore di stima e valuta se riaddestrare una o più reti neurali, a seconda dell'errore commesso da ciascuna di esse. Anche in questo il riaddestramento può avvenire a partire dai pesi precedentemente calcolati o a partire da pesi inizializzati casualmente.

Terminate tali operazioni, l'algoritmo termina e viene riavviato nel momento in cui si rende disponibile il campione successivo.

Come imposto dalla specifica di eseguire una stima *real time* del segnale interno, tale calcolo viene eseguito in un tempo inferiore al periodo con cui vengono forniti in ingresso i campioni. In tal modo è possibile eseguire una stima in corrispondenza di ogni ingresso. Nel caso in cui, però, l'algoritmo sia costretto a riaddestrare una o più reti neurali, il tempo di esecuzione risulta aumentato. Ciò comporta che alcuni campioni vengano resi disponibili quando l'algoritmo non è ancora terminato: in corrispondenza di tali campioni non è possibile avere una stima del segnale interno.

Nei test effettuati, si è osservato che, al massimo, vengono persi 50 campioni in corrispondenza dell'addestramento di tutte le tre reti neurali, corrispondenti a circa 0.8 secondi di acquisizione. Tali valori risultano ridotti qualora vengano addestrate soltanto due o soltanto una rete neurale. A titolo di esempio si riporta in Figura 5.24 l'andamento della stima del segnale interno effettuata con l'algoritmo appena descritto.

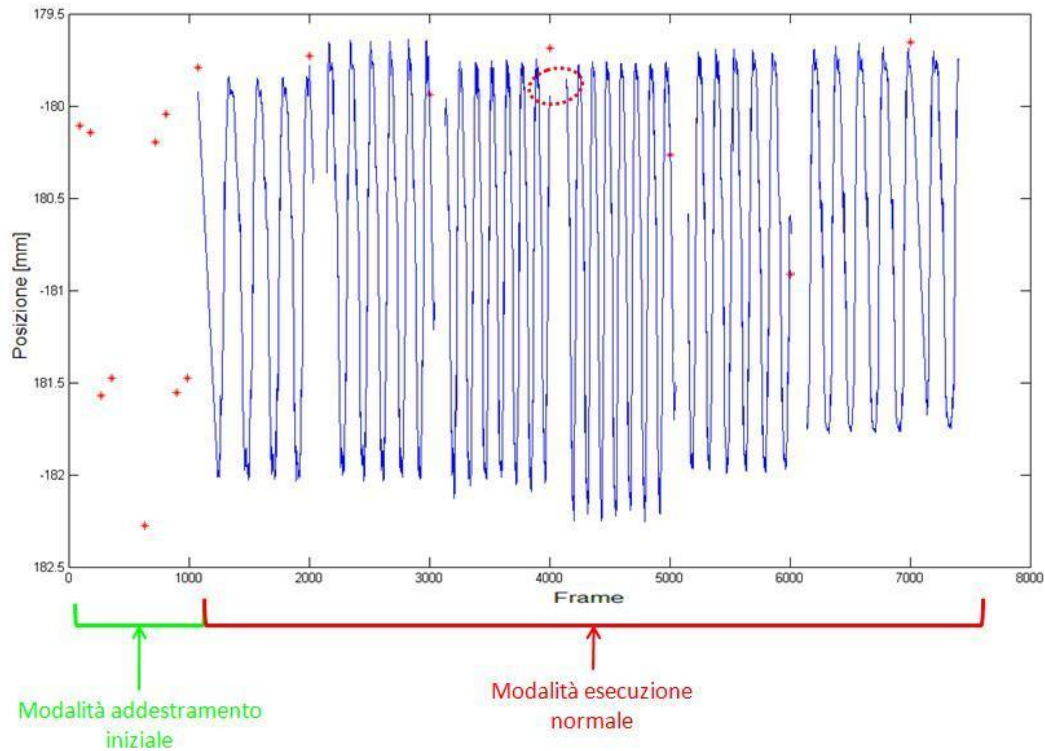


Figura 5.24: In blu è mostrata la stima della coordinata x del segnale interno, mentre i punti rossi mostrano l'effettiva coordinata x del quarto marcatore negli istanti in cui viene registrata.

Come si osserva, in modalità addestramento iniziale non è disponibile la stima del segnale interno e le acquisizioni del quarto marcatore avvengono con una frequenza maggiore di quella che sia ha in modalità di esecuzione normale. L'intervallo evidenziato dall'ovale rosso in figura 5.24, invece, evidenzia una parte dell'esecuzione normale in cui l'algoritmo, dovendo riaddestrare le reti neurali, non è in grado di fornire una stima del segnale interno.

5.3. Fantoccio per la simulazione del movimento respiratorio

Al fine di validare lo strumento software sviluppato, soprattutto per quanto riguarda la capacità di acquisire in tempo reale la posizione di marcatori esterni e di effettuare *online* la stima del segnale interno, è stato utilizzato un fantoccio volto a simulare il movimento dovuto alla respirazione sia della superficie corporea sia della lesione interna. In tale paragrafo si descriverà la struttura e il funzionamento del fantoccio appena citato [65].

Struttura

Il fantoccio è stato realizzato prevalentemente in legno e i suoi principali componenti sono visibili in Figura 5.25

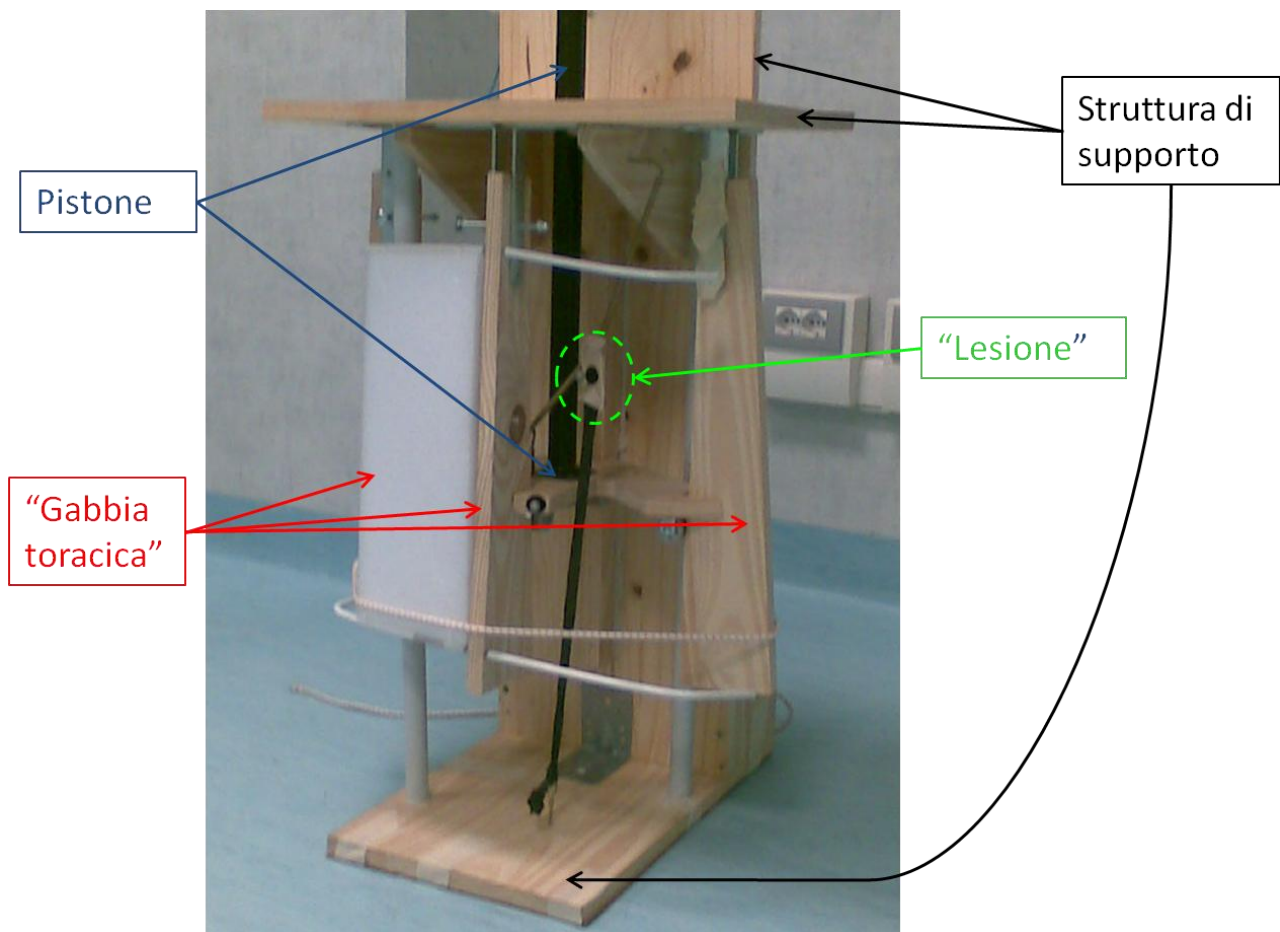


Figura 5.25: Immagine del fantoccio in cui sono evidenziati i componenti descritti in seguito.

- Struttura di supporto

Consiste in un basamento rettangolare sul quale è fissata perpendicolarmente un'asse verticale. A quest'ultima è vincolato un pianale orizzontale, distante circa 80 cm dal basamento e avente nel centro un foro passante.

La struttura di supporto ha lo scopo di sostenere le parti mobili del fantoccio e di vincolarne il movimento.

- “Gabbia toracica”

La struttura della gabbia toracica viene simulata attraverso l'utilizzo di tre elementi, due laterali e uno centrale, vincolati superiormente al pianale della struttura di supporto attraverso cerniere. Essendo l'estremità inferiore di tali elementi non vincolata, essi possono ruotare attorno alle suddette cerniere.

Sulla “gabbia toracica” possono essere fissati tre o più marcatori riflettenti, localizzabili tramite un sistema optoelettronico.

- Pistone

Il pistone è costituito da un'asta cilindrica in metallo alla cui estremità è fissata un'asse di legno sagomata e perpendicolare all'asta. La struttura è montata in modo che l'asta attraversi il foro presente sul pianale superiore e l'asse di legno sia posta internamente alla “gabbia toracica”. Tale asse è sagomata in modo da creare tre sedi per gli elementi della “gabbia toracica”. Questi ultimi vengono mantenuti all'interno delle loro sedi attraverso l'utilizzo di un maglia o di una corda elastica avvolta esternamente. Il pistone è in grado di eseguire soltanto movimenti traslatori in direzione verticale, dal momento che il foro passante funge da manicotto mentre gli elementi della “gabbia toracica” inseriti nelle apposite sedi impediscono la rotazione attorno all'asse verticale.

- “Lesione”

La lesione interna viene simulata tramite un piccolo supporto sostenuto da tre elementi elastici: il primo vincolato al pianale superiore, il secondo al pistone e il terzo al basamento.

Sul supporto viene fissato un marcatore riflettente, che rappresenta la lesione interna da irradiare.

Funzionamento

Come già anticipato, il fantoccio deve simulare il movimento respiratorio della superficie corporea e della lesione interna.

Tale movimento avviene azionando il pistone in uno dei seguenti modi:

- Motore elettrico

Un motore elettrico a corrente continua è fissato sull'estremità superiore dell'asse verticale della struttura di supporto. Una manovella è solidale al rotore e, tramite una

biella vincolata la pistone, il movimento rotatorio viene trasformato nel movimento traslatorio verticale di quest'ultimo.

L'utilizzo del motore elettrico permette di variare la frequenza del movimento agendo sulla tensione di alimentazione, ma non permette di variare l'ampiezza del movimento mentre questo viene eseguito, dato questo richiederebbe di agire sulla lunghezza della manovella.

- Puleggia a sezione variabile

Si è deciso di sviluppare un sistema d'azionamento diverso dal motore elettrico allo scopo di introdurre la possibilità di variare, oltre alla frequenza, anche l'ampiezza del movimento.

A tal fine è stato sviluppato un sistema in cui il pistone viene sollevato attraverso un tirante collegato ad una puleggia azionata manualmente. Tale puleggia è stata realizzata in modo da avere una sezione ellittica variabile in funzione della lunghezza [Figura 5.26].

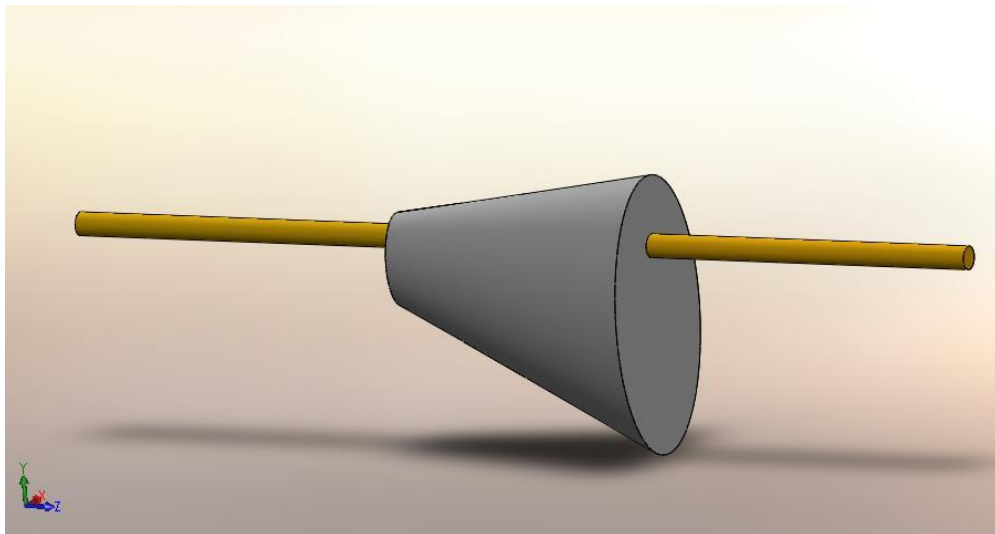


Figura 5.26: Disegno schematico della puleggia a sezione ellittica variabile.

Di conseguenza, l'ampiezza del movimento verticale trasferito al pistone risulta dipendente dalla posizione sulla puleggia dove è collocato il tirante, dal momento che a differenti posizioni corrispondono differenti sezioni. In particolare, l'ampiezza del movimento verticale è di circa 5 cm se il tirante agisce dove la sezione è minore, mentre è di circa 11 cm se il tirante è collocato presso l'estremità della puleggia con sezione maggiore. L'operatore può quindi modificare l'ampiezza del movimento verticale, mentre questo viene eseguito, tra 5 e 11 cm, con passo di un centimetro, semplicemente

spostando il tirante in una delle guide che sono state opportunamente collocate sulla superficie laterale della puleggia.

La frequenza del movimento ottenuto, invece, dipende solamente dalla frequenza con la quale l'operatore mette in rotazione la puleggia, ed quindi variabile a seconda delle esigenze.

La Figura 5.27 mostra il set-up sperimentale utilizzato nel caso in cui il fantoccio venga azionato manualmente tramite la puleggia appena descritta.

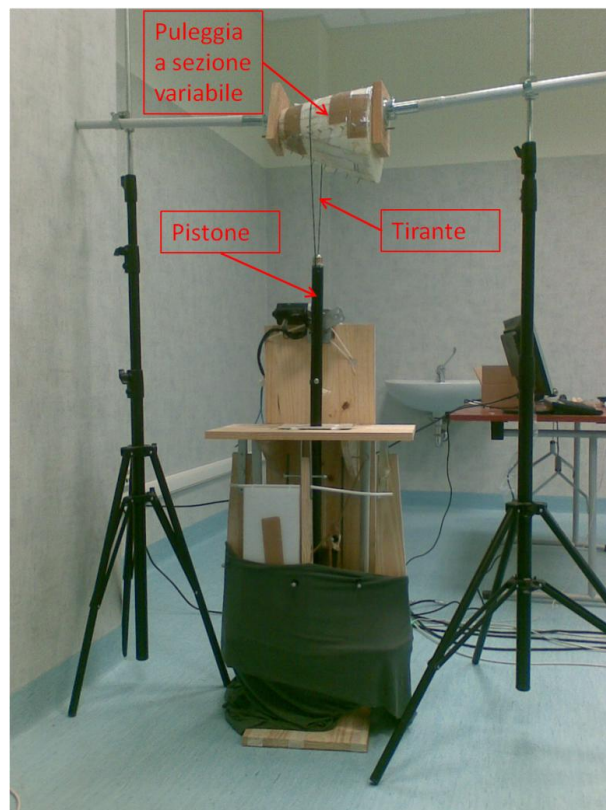


Figura 5.27: Set-up sperimentale: sono evidenziati la puleggia a sezione variabile, il tirante e il pistone.

Il movimento traslatorio del pistone viene trasferito agli elementi della “gabbia toracica”. In particolare, la discesa del pistone dilata tali elementi, simulando l’inspirazione, mentre quando il pistone si muove dal basso verso l’alto, la gabbia toracica viene contratta dagli elementi elastici che la avvolgono, simulando l’espirazione [Figura 5.28].

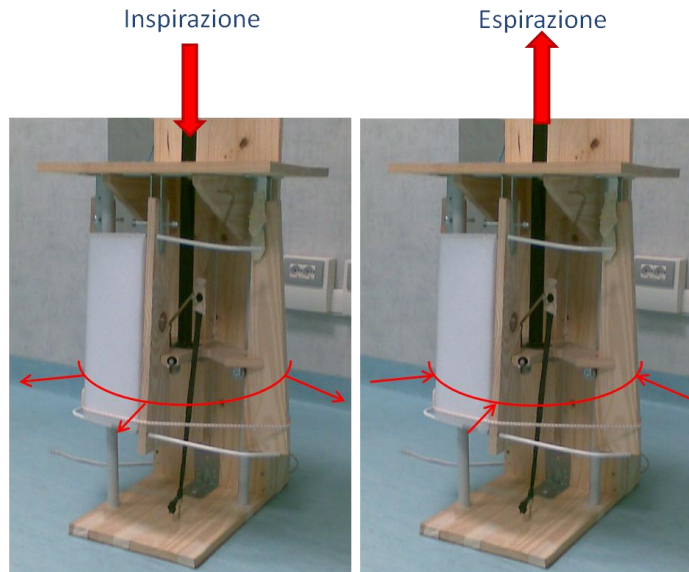


Figura 5.28: Schematizzazione dei movimenti eseguiti dal fantoccio per la simulazione della fase inspiratoria ed espiratoria.

Poiché la “lesione” è collegata al pistone, anch’essa viene mossa da quest’ultimo. Tramite la regolazione dell’elasticità degli elementi che la sostengono, è possibile agire sulla traiettoria compiuta da quest’ultima. A titolo d’esempio, la Figura 5.29 mostra la traiettoria della lesione ottenuta azionando il fantoccio con il motore elettrico alimentato a differenti tensioni. Si nota che è possibile riprodurre la spiccata isteresi tipica della reale trattoria interna delle lesioni tumorali.

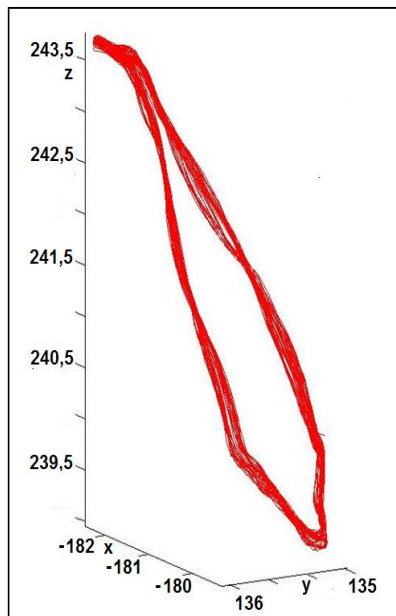


Figura 5.29: Isteresi della traiettoria della “lesione interna” ottenuta a seguito di una acquisizione eseguita alimentando il motore a differenti tensioni [65]. Unità di misura in [mm].

Si discuteranno di seguito le caratteristiche del movimento simulato che si ottiene con il fantoccio appena descritto. A tal fine, la Tabella 5.6 propone un confronto tra i *range* di movimento (nelle tre direzioni anatomiche e in 3D) osservati in pazienti trattati con il sistema *CyberKnife* e nelle simulazioni eseguite su fantoccio.

Il campione relativo ai trattamenti con *CyberKnife* consiste in 20 differenti frazioni estratte, con modalità che saranno descritte nel dettaglio nel paragrafo 6.1, dal database raccolto presso il Georgetown University Medical Center. Riguardo alle simulazioni su fantoccio, invece, i dati che saranno presentati si riferiscono a 7 differenti acquisizioni.

	Cranio-caudale (CC) [mm]		Latero-laterale (LL) [mm]		Antero-posteriore (AP) [mm]		3D [mm]	
	Min	Max	Min	Max	Min	Max	Min	Max
Trattamenti								
<i>CyberKnife</i>	2,00	70,50	1,80	75,70	1,80	40,70	4,42	108,25
Simulazioni								
su fantoccio	2,89	19,24	3,76	23,59	4,58	9,96	8,07	32,03

Tabella 5.6: Confronto tra i valori minimi e massimi dei *range* di movimento del bersaglio osservati in 20 frazioni di trattamento con *CyberKnife* e in 7 simulazioni su fantoccio.

Si noti che i *range* di movimento relativi ai trattamenti con *CyberKnife* risultano particolarmente elevati in quanto comprendono anche eventuali *shift* di posizione del paziente, dunque sono solo in parte riconducibili ad effettivi fenomeni di *organ motion*. Nonostante ciò, tali spostamenti vengo ugualmente compensati dal modello sviluppato, tramite il riaddestramento delle reti neurali.

Conseguentemente a quanto appena osservato, si può affermare che i *range* di movimento ottenuti nelle prove su fantoccio, pur risultando inferiori a quelli presentati per i trattamenti con *CyberKnife*, siano allo stesso modo compatibili con movimenti respiratori fisiologici. In ultimo, relativamente al periodo dei movimenti respiratori simulati, nelle 7 acquisizioni effettuate sono stati osservati valori compresi tra circa 3 e 7 secondi. Considerando il periodo medio di un respiro fisiologico pari a 3,2 secondi [60], tali valori risultano compatibili ed eventualmente superiori a quest'ultimo. Questo scostamento delle condizioni di respiro fisiologico è imputabile all'azionamento manuale. In particolare, dato il carico associato al sollevamento del pistone, risulta difficoltoso effettuare movimenti ad elevata frequenza per tempi prolungati.

6. Validazione

Al termine dello sviluppo del progetto descritto nel capitolo precedente, si è proceduto a validare lo strumento realizzato. In particolare, tale validazione è stata articolata in due fasi successive. La prima ha come obiettivo quello di verificare la performance degli algoritmi di correlazione implementati, in termini di capacità delle reti neurali di stimare accuratamente il segnale interno, utilizzando i dati provenienti dal database relativo ai trattamenti con *CyberKnife* e confrontando la performance del modello implementato con quella del *modeler* (paragrafo 5.2.2). La seconda fase, invece, ha lo scopo di verificare la capacità di stimare *on-line* il segnale interno a partire da un flusso di dati *real-time* (paragrafo 5.2.3). A tal fine sono stati utilizzati dati acquisiti tramite il fantoccio descritto nel paragrafo 5.3.

Tali fasi saranno descritte nei due paragrafi seguenti.

6.1. Confronto con *CyberKnife*

Come anticipato, in questa prima fase di validazione si è interessati a valutare la capacità delle reti neurali di stimare accuratamente il segnale interno in funzione del segnale esterno. Di conseguenza si valutano le loro capacità di apprendere la correlazione interno/esterno che, come specificato nel paragrafo 3.5, consiste nella principale specifica del progetto.

Tale valutazione è stata ottenuta elaborando, in ambiente Matlab, la stima del segnale interno eseguita dell'algoritmo e quella eseguita dal *modeler*, effettuando un confronto su un campione di 20 frazioni di trattamento.

I dati utilizzati sono stati prelevati dal database dei trattamenti eseguiti tramite *CyberKnife* presso il Georgetown University Medical Center. La popolazione considerata consiste in 20 distinte frazioni di trattamento, divise in due *subset* di 10 frazioni ciascuno, chiamati “*worst*” e “*control*” [66]. Il *subset worst* comprende le frazioni nelle quali si osserva il maggiore errore di puntamento medio, ovvero i casi nei quali si suppone che l'apprendimento della correlazione interno/esterno risulti più difficoltoso. Il *subset control*, invece, è costituito da frazioni scelte casualmente all'interno del database, ovviamente ad esclusione dei casi *worst*. Operando il confronto su entrambi i *subset*, si ottiene quindi una valutazione della performance dell'algoritmo implementato, in quanto si lavora su casi dalle caratteristiche estremamente diverse tra di loro.

Prima di presentare i risultati ottenuti, è opportuno specificare l'architettura delle reti neurali utilizzate e alcuni parametri con i quali sono state addestrate [Tabella 6.1].

Caratteristica	Impostazione
Strato nascosto:	10 neuroni, f.d.t = tanh
Strato d'uscita:	3 neuroni, f.d.t = tanh
Learning Rate:	0,07
Massimo numero di epoche d'addestramento:	100

Tabella 6.1: Caratteristiche delle reti neurali implementate nell'algoritmo di correlazione.

Come si osserva, le reti neurali utilizzate sono caratterizzate da un unico strato nascosto composto da 10 neuroni la cui funzione di attivazione è la tangente iperbolica. Anche lo strato d'uscita mostra la medesima funzione d'attivazione. Si è deciso di imporre a 100 il massimo numero di epoche d'addestramento, con l'obiettivo sia di preservare la capacità di generalizzazione della rete evitando il fenomeno dell'*overfitting* che di limitare il tempo di addestramento nel caso in cui questo non venga terminato per il raggiungimento del *performance goal* ($RMS < 0.01$).

La Figura 6.1 mostra schematicamente la procedura utilizzata per l'analisi dei risultati.

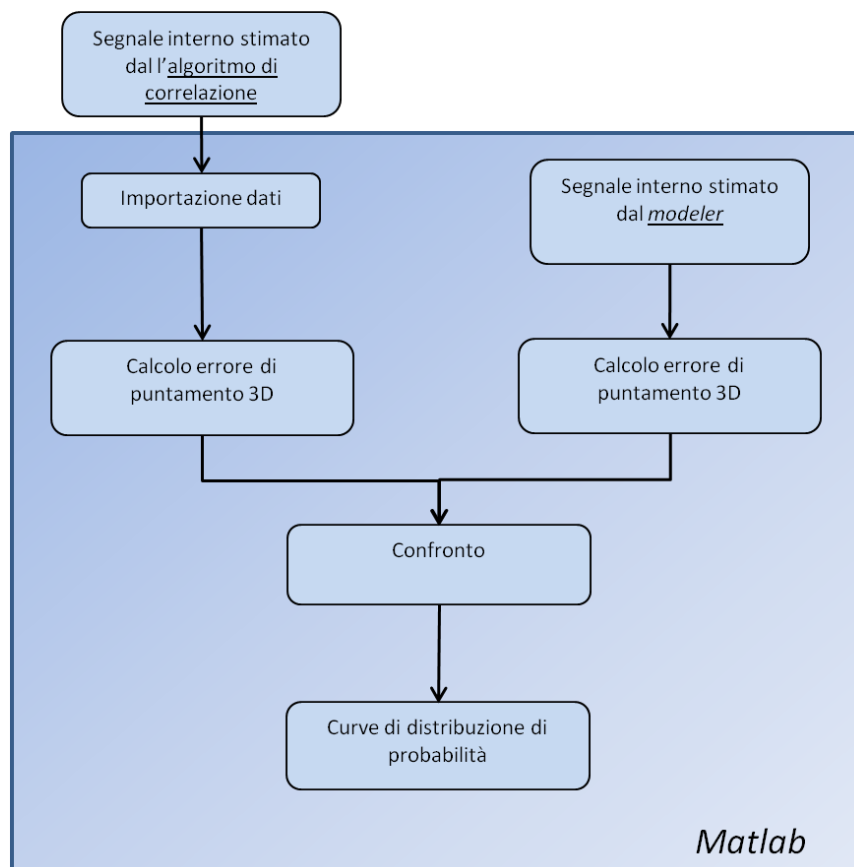


Figura 6.1: Schema della procedura utilizzata nella prima fase di validazione dell'algoritmo sviluppato.

La prima operazione eseguita consiste nell'importazione in Matlab delle stime effettuate dall'algoritmo di correlazione che, si ricorda, vengono salvate in un file formato testo. Le stime effettuate dal *modeler*, invece, sono già disponibili in Matlab nel momento in cui vengono importati i dati relativi al trattamento in analisi.

Il parametro utilizzato per confrontare le due stime consiste nell'errore di puntamento 3D, ovvero la differenza tridimensionale tra la posizione del bersaglio stimata e quella effettiva misurata in corrispondenza dell'acquisizione delle immagini radiografiche. Definita S la stima effettuata e P la posizione vera del *target*, gli errori di puntamento in corrispondenza di ogni campione di P vengono calcolati come segue:

$$S = [S_x \ S_y \ S_z]; \quad P = [P_x \ P_y \ P_z]$$

$$targeting\ error\ (TE) = \sqrt{(S_x - P_x)^2 + (S_y - P_y)^2 + (S_z - P_z)^2} \quad (6.1)$$

A titolo d'esempio, la Figura 6.2 mostra l'andamento del segnale TE in uno specifico trattamento.

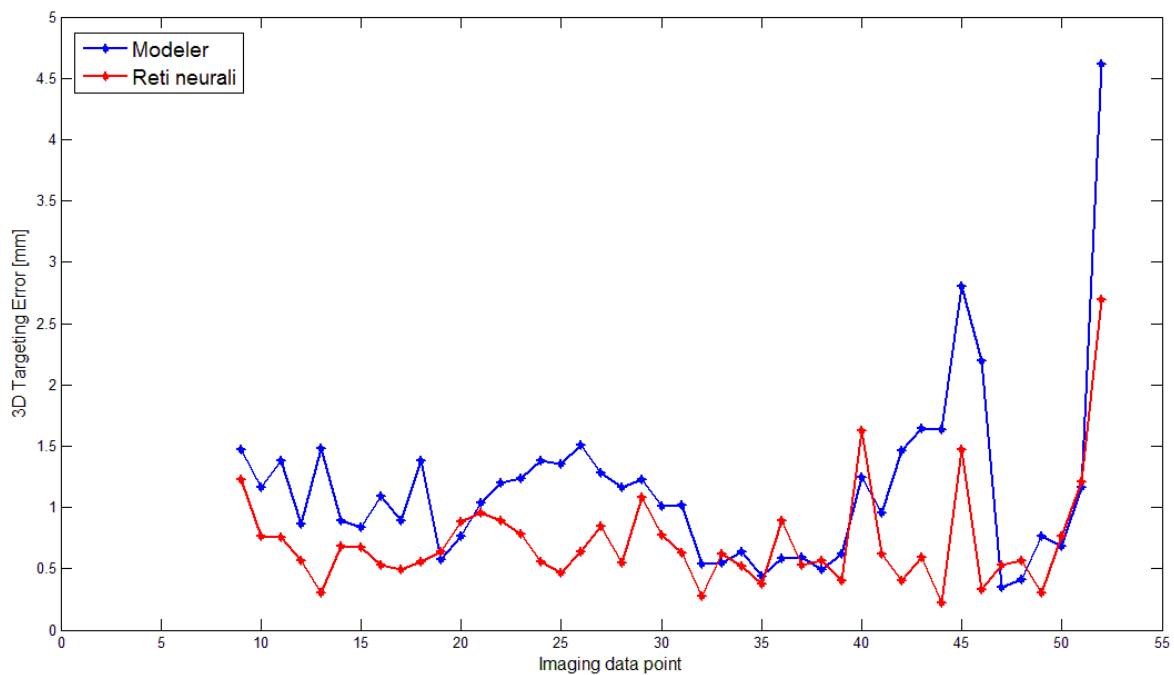


Figura 6.2: Errore di puntamento (TE) rilevato in corrispondenza di ogni immagine radiografica di controllo acquisita nel corso del trattamento.

Si precisa che il TE è stato calcolato considerando soltanto i punti di *imaging* di controllo, escludendo quelli che il sistema *CyberKnife* acquisisce nelle prime fasi del trattamento per la costruzione del modello di correlazione, in quanto, in corrispondenza di tali punti, la stima effettuata dal *modeler* si basa su modello non ancora affidabile.

Dopo aver calcolato il parametro TE per ogni acquisizione radiografica di controllo, le due stime sono state confrontate sia calcolando la media, la deviazione standard e valore massimo del segnale TE (relativo alle reti e al *modeler*) che tracciando la sua curva di distribuzione di probabilità cumulativa, relativamente a tutte le misure effettate nel *subset* considerato.

I valori medi complessivi dei *targeting error* calcolati per il *modeler* e per le reti neurali sono riportati nella Tabella 6.2.

<i>Subset</i>	Media TE <i>modeler</i> [mm]	Media TE reti neurali [mm]
<i>control</i>	1,6675	1,3870
<i>worst</i>	6,6086	4,8084

Tabella 6.2: Valori medi del *targeting error* per *modeler* e per le reti neurali.

Per prima cosa si può notare che la media del TE del *modeler* relativa ai casi *worst* risulta quasi 5 mm superiore a quella misurata nel caso del *subset control*. Si ricorda, infatti, che il *subset worst* è costituito dalle frazioni che, all'interno del database considerato, mostrano i più elevati errori di puntamento commessi dal *modeler*.

Inoltre si osserva che la differenza tra i valori medi ottenuti dal *modeler* e dalle reti neurali è di circa 0,3 mm per i trattamenti *control* e di 1,8 mm per i trattamenti *worst*.

Tale risultato viene confermato dal confronto tra le curve di distribuzione di probabilità riferite al TE del *modeler* e del modello sviluppato, sia per il *subset control* che per il *subset worst* [Figure 6.3 e 6.4].

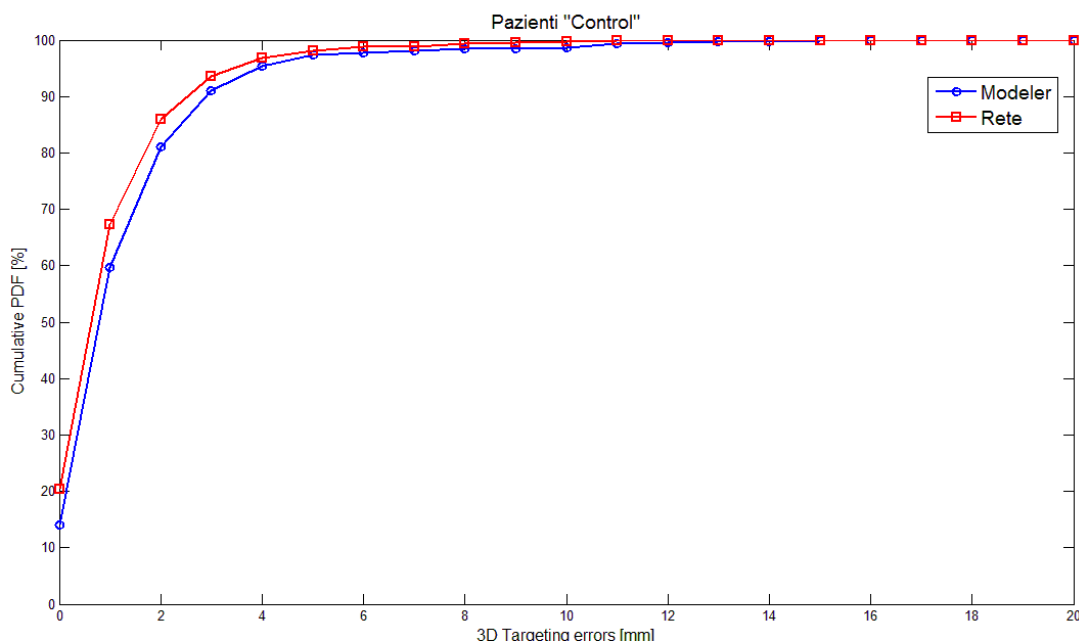


Figura 6.3: Distribuzione di probabilità cumulativa del TE per i casi *control*. In blu la curva relativa al *modeler*, in rosso la curva relativa all' algoritmo di correlazione sviluppato (reti neurali).

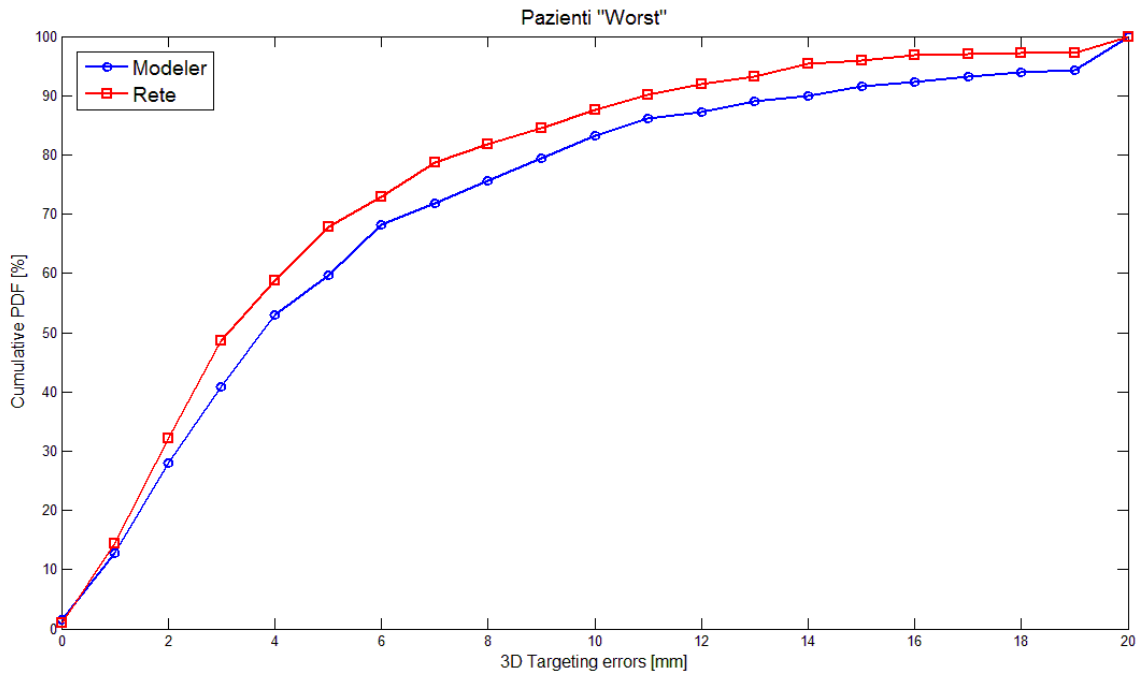


Figura 6.4: Distribuzione di probabilità cumulativa del *TE* per i casi *worst*. In blu la curva relativa al *modeler*, in rosso la curva relativa all’algoritmo di correlazione sviluppato (reti neurali).

Le distribuzioni appena riportate sono state confrontate con un test statistico non parametrico (test di Wilcoxon), il quale fornisce un *p-value* relativo al *subset* considerato. I valori ottenuti sono riportati nella seguente tabella [Tabella 6.3]:

<i>Subset</i>	<i>p-value</i>
<i>control</i>	$2,1066 \times 10^{-5}$
<i>worst</i>	0,0012

Tabella 6.3: *p-value* relativi al confronto tra le distribuzioni dell’errore di puntamento 3D osservate per le stime eseguite dalle reti neurali e dal *modeler*, sia per il *subset control* che per il *subset worst*.

Per entrambi i *subset* considerati, si osserva quindi una differenza statisticamente significativa tra gli errori di puntamento commessi dal modello sviluppato e quelli commessi dal *modeler*. In conclusione, verranno presentati due esempi che mostrano l’andamento della stime effettuate con le reti neurali implementate.

La Figura 6.5 mostra un estratto dell’andamento della stima della coordinata *x* (direzione carnio-caudale, CC) del segnale interno (curva in blu) in un trattamento del *subset “control”*. Tale stima può essere confrontata, *timestamp* per *timestamp*, con quella eseguita dal *modeler* (curva in verde). In corrispondenza dell’acquisizione delle immagini radiografiche di controllo, si può anche osservare l’effettiva coordinata *x* del *target* (punti in rosso).

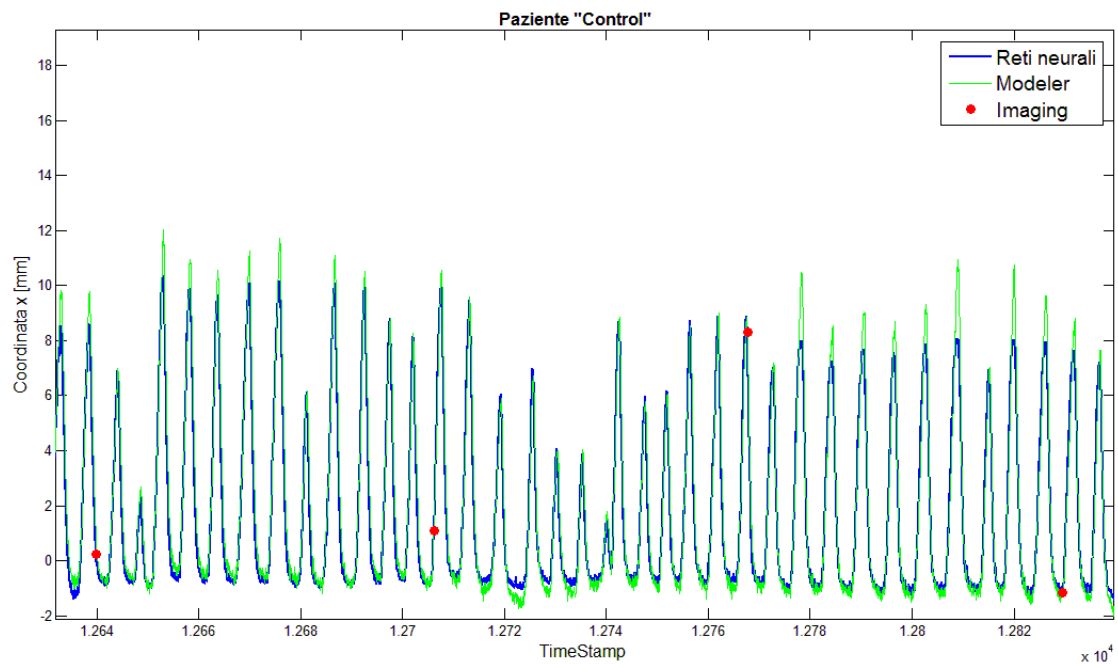


Figura 6.5: Estratto della simulazione di un trattamento “control”. In blu la stima della coordinata x (CC) del segnale interno eseguita dalle reti neurali, in verde quella eseguita dal *modeler*. I punti rossi evidenziano l’effettiva coordinata x del bersaglio rilevata tramite immagini radiografiche.

La Figura 6.6, invece, riporta l’estratto dell’andamento della stima in un trattamento del *subset* “worst”. Analogamente alla figura precedente, si possono notare anche la stima eseguita dal *modeler* e i punti di *imaging*.

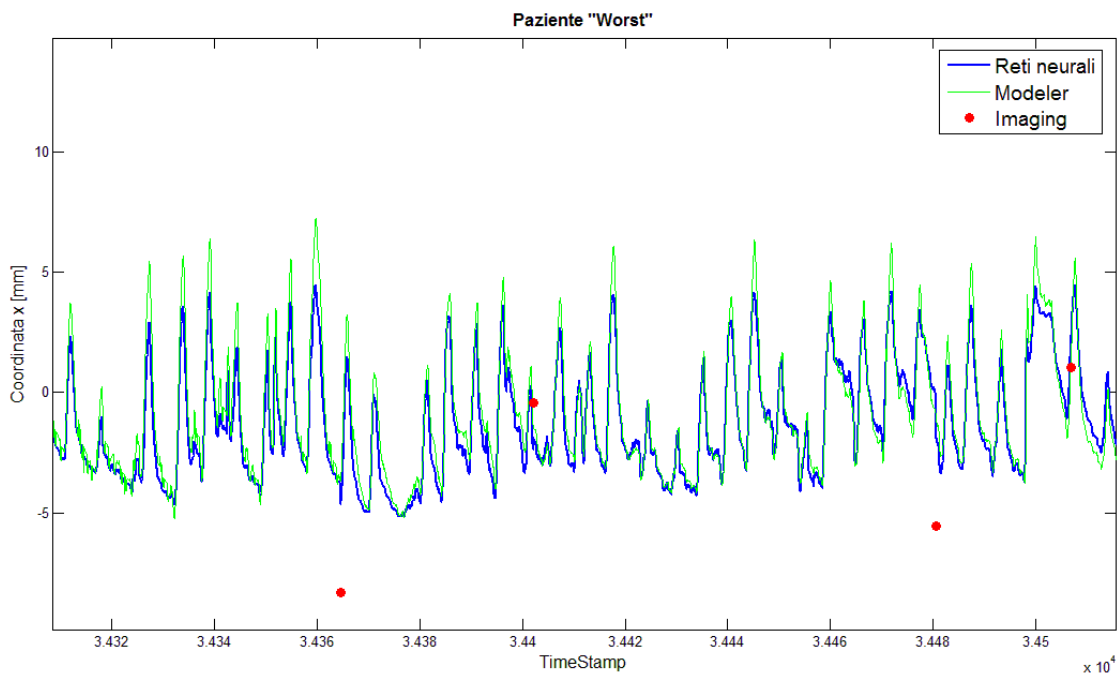


Figura 6.6: Estratto della simulazione di un trattamento “worst”. La traccia blu indica la stima della coordinata x (CC) del segnale interno eseguita dalle reti neurali, la traccia verde indica quella eseguita dal *modeler*. I punti rossi evidenziano l’effettiva coordinata x del bersaglio rilevata tramite immagini radiografiche.

6.2. Acquisizioni su fantoccio per la simulazione del movimento respiratorio

In questa seconda fase di validazione sono state svolte acquisizioni utilizzando il fantoccio per la simulazione del movimento respiratorio descritto nel paragrafo 5.3.

I primi test sono stati effettuati *offline*, successivamente sono state effettuate acquisizioni e predizioni *online*. Nel primo caso, le informazioni relative al movimento dei quattro marcatori vengono acquisite dal sistema SMART-D e poi salvate in un file formato .tdf. Terminata l'acquisizione, tale file viene letto utilizzando la *Smart Virtual Machine*, che si ricorda, è un'applicazione che legge un file formato .tdf esattamente alla frequenza con la quale sono stati acquisiti i dati e li rende disponibili scrivendoli su una porta UDP. Tale porta viene letta dall'algoritmo di correlazione che è stato descritto nel paragrafo 5.2.3.

Nel caso in cui si operi *online*, invece, i dati relativi alla posizione dei quattro marcatori vengono resi disponibili dal sistema SMART-D direttamente all'algoritmo di correlazione *frame per frame*.

In entrambi i casi, il funzionamento dell'algoritmo sviluppato è il medesimo, dunque sia le prove *offline* che le prove *online* sono volte a valutare la capacità dell'algoritmo di acquisire dati da un flusso *real time* (dalla *Smart Virtual Machine* o dallo SMART-D) e, in corrispondenza di ogni *frame*, di stimare il segnale interno, ovvero la posizione del quarto marcatore acquisito. Il set-up sperimentale è mostrato in Figura 6.7:

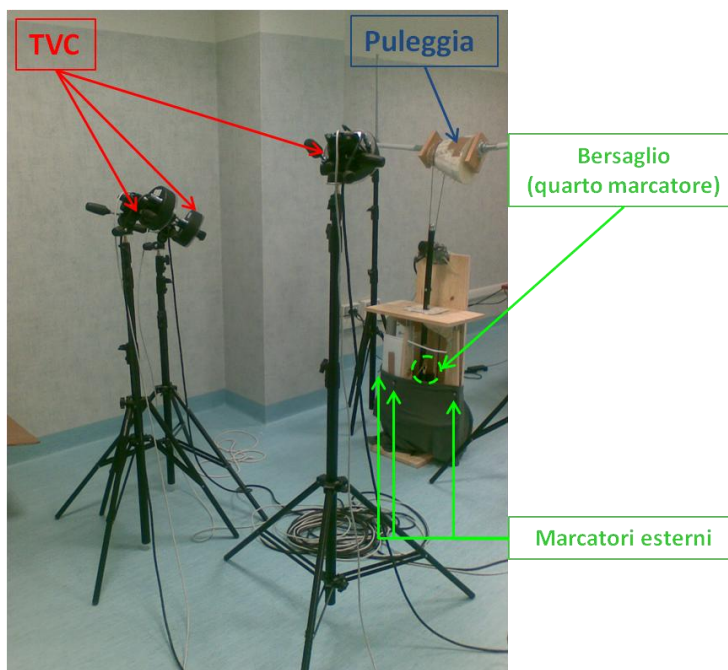


Figura 6.7: Set-up sperimentale utilizzato per le acquisizioni su fantoccio.

Si può notare il fantoccio che viene azionato, tramite il tirante, dalla puleggia a sezione ellittica collocata sopra di esso. Nella parte sinistra dell'immagine, invece, sono visibili le tre telecamere ad infrarossi del sistema SMART-D. Tali sistema è stato calibrato precedentemente alle acquisizioni, ottenendo errori di ricostruzione 3D aventi una media di 0,298 mm e una deviazione standard di 0,236 mm.

Sia per le prove *offline* che per le prove *online* è stato utilizzato il protocollo sperimentale descritto di seguito:

1. All'inizio dell'acquisizione, il fantoccio viene mosso in modo da percorrere tutte le posizioni intermedie tra la posizione di massima espirazione e massima inspirazione. In corrispondenza di questo movimento la rete neurale acquisisce gli esempi per l'addestramento iniziale.
2. Successivamente, il fantoccio viene azionato attraverso la puleggia, eseguendo alcuni "respiri" ad una data ampiezza e con un periodo di circa 5 secondi.
3. Durante l'acquisizione, al fine di testare il comportamento del modello implementato in corrispondenza di eventuali variazioni di ampiezza e/o frequenza della respirazione, viene variata l'ampiezza del movimento modificando la posizione del tirante sulla puleggia, mentre la frequenza viene modulata agendo sulla velocità di rotazione della puleggia.

La Figura 6.8 mostra l'andamento della coordinata x (direzione antero-posteriore, AP) del terzo marcatore in una delle acquisizioni eseguite:

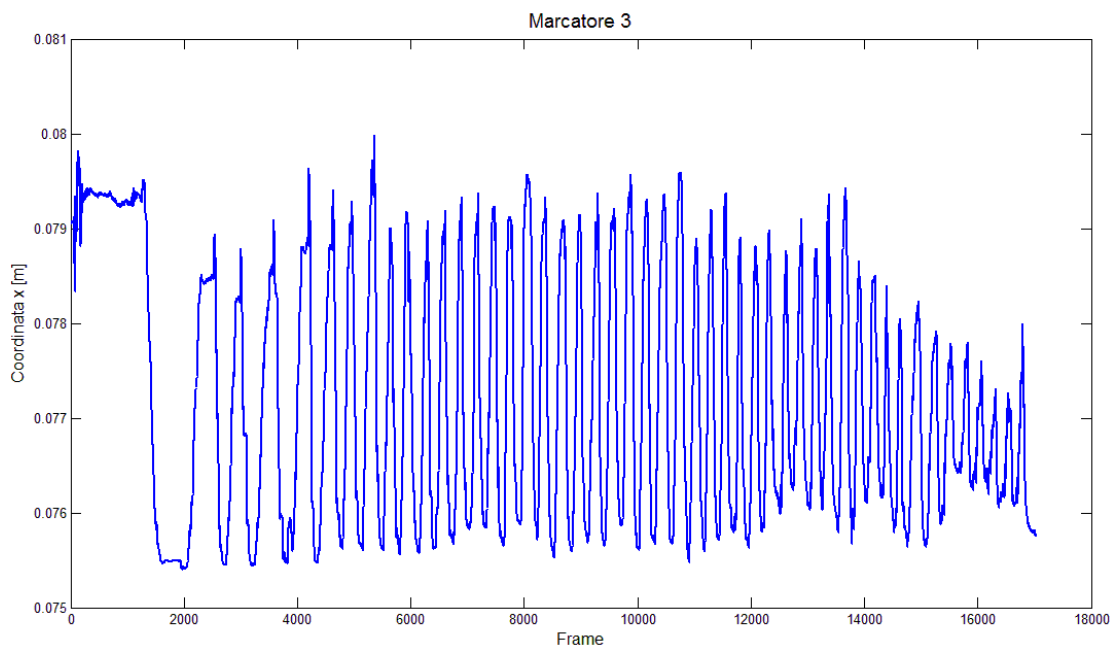


Figura 6.8: Andamento della coordinata x (AP) del marcatore numero 3.

In corrispondenza dei *frame* precedenti al numero 2000, si può osservare la manovra eseguita in accordo al primo punto del protocollo utilizzato. Nei frame successivi, invece, si osservano chiaramente le variazioni di ampiezza del movimento respiratorio che viene simulato.

La traiettoria tridimensionale del quarto marcatore che, si ricorda, simula il bersaglio da inseguire, è stata valutata in una apposita acquisizione ed è riportata in Figura 6.9.

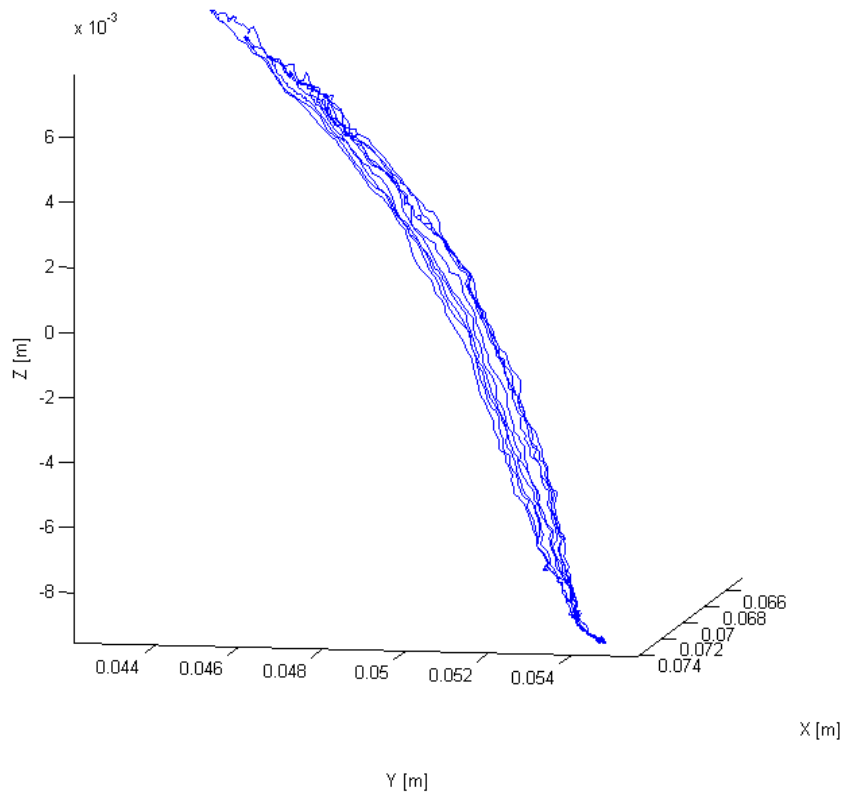


Figura 6.9: Traiettoria tridimensionale del bersaglio. L'asse X coincide con la direzione antero-posteriore (AP), l'asse Y con la direzione cranio-caudale (CC) e l'asse Z con la direzione latero-laterale (LL).

È possibile osservare come sia caratterizzata dall'isteresi tipica del movimento interno causato dall'*organ motion*[60].

Analisi offline delle acquisizioni

Lo schema a blocchi riportato in Figura 6.10 mostra le attività svolte per l'analisi *offline* delle acquisizioni.

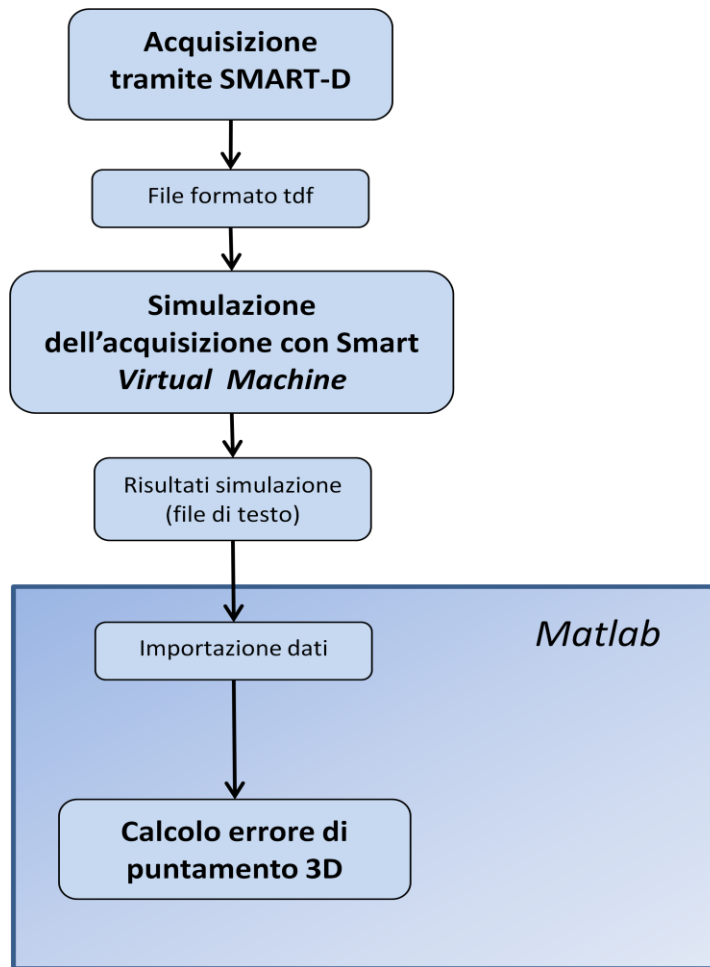


Figura 6.10: Schema a blocchi delle attività svolte per l'analisi *offline* delle acquisizioni.

Come già descritto nel paragrafo 5.2.3, l'algoritmo, tramite la lettura della porta UDP, riceve i dati relativi alla posizione spaziale di quattro marcatori, di cui i primi tre vengono considerati segnale esterno, mentre il quarto simula il segnale interno, dunque la posizione della lesione dal irradiare.

Le stime effettuate vengono salvate in un file formato testo. Allo stesso modo vengono registrate, *frame* per *frame*, le coordinate del quarto marcatore, nonché la posizione di tale marcatore in corrispondenza dei *frame* in cui la sua posizione viene letta per l'addestramento iniziale delle reti neurali o per la valutazione dell'errore di stima e l'eventuale riaddestramento. È quindi chiaro che, nell'ambito di questo studio, la stima effettuata dal modello sviluppato può essere valutata non solo in corrispondenza dell'acquisizione delle immagini di controllo, bensì istante per istante, consentendo di valutare la performance delle reti neurali in corrispondenza di ogni *frame* acquisito.

Anche in questo caso, l'architettura delle reti neurali utilizzate prevede un singolo strato nascosto con funzione d'attivazione a tangente iperbolica e uno strato d'uscita con la medesima funzione d'attivazione. La frequenza di acquisizione dei punti di controllo è stata impostata in

modo da acquisirne uno ogni 1000 *frame*, ovvero ogni 17 secondi circa. Tale frequenza risulta decisamente maggiore di quella osservata nei trattamenti eseguiti con *CyberKnife* (dell'ordine dei minuti). A tal proposito si ricorda che lo scopo principale della acquisizioni su fantoccio consiste nel valutare la capacità di operare in tempo reale, scopo per il quale sono sufficienti acquisizioni della durata di pochi minuti. È quindi chiaro che, per poter disporre di un numero significativo di punti di controllo, è stato necessario aumentare significativamente la frequenza con cui questi vengono acquisiti.

Per prima cosa è stata valutata la performance dell'algoritmo in termini di capacità di operare in *real time*. Si è osservato che il modello di correlazione sviluppato è in grado di eseguire la stima del segnale interno in funzione del segnale esterno in un tempo inferiore al periodo compreso tra due *frame* successivi acquisiti ad una frequenza di 60 Hz, che risulta essere pari a 0,016 secondi circa. Nel caso in cui, però, sia necessario riaddestrare le reti neurali, questo comporta, in media, la perdita di circa 48 *frame*, corrispondenti a circa 0,80 secondi.

La performance in termini di accuratezza della stima, invece, è stata quantificata in termini di errore di puntamento 3D, come descritto nel paragrafo precedente.

La Tabella 6.4 mostra, per le quattro acquisizioni effettuate, la media, la deviazione standard e il valore massimo dell'errore di puntamento 3D calcolato per ogni *frame* ed espresso in percentuale del massimo *range* di movimento 3D del bersaglio.

Acquisizione	Numero di Frame	Massimo range di movimento 3D [mm]	Media TE[%]	Deviazione standard TE[%]	Massimo TE[%]
1	4020	8,1081	17,17	7,42	35,92
2	9363	8,7390	8,99	4,33	24,11
3	12884	10,0010	6,89	4,82	27,49
4	4774	8,0749	6,34	4,14	27,69
Valori medi:			9,85	5,18	28,80

Tabella 6.4: Dati relativi alle analisi effettuate su quattro acquisizioni *offline*.

Si osservi che i *range* di movimento 3D del bersaglio ottenuti risultano di entità ridotta. Questo comporta una maggiore difficoltà, per il prototipo sviluppato, nell'apprendimento della correlazione interno/esterno, dunque, relativamente a tale aspetto, il prototipo è stato testato in condizioni meno favorevoli di quelle che potrebbero essere incontrate in un eventuale applicazione clinica. Nonostante ciò, gli errori osservati in termini assoluti risultano, in media, pari a 0,8499 mm, quindi di entità sub-millimetrica ed accettabili da un punto di vista clinico.

A titolo di esempio, le Figure 6.11, 6.12 e 6.13, mostrano l'andamento della stima del segnale interno (curva in blu) nelle tre coordinate spaziali nel caso della quarta acquisizione effettuata. Tale segnale può essere confrontato con quello relativo all'effettiva posizione del *target* (curva in rosso). Si noti che, a seguito dei punti di *imaging* (mostrati in verde) in corrispondenza dei quali è stato effettuato un riaddestramento, non si ha a disposizione né la stima del segnale interno né la registrazione del suo effettivo valore in quanto l'algoritmo non è in grado di eseguire il riaddestramento nel periodo di tempo compreso tra l'acquisizione di un *frame* ed il successivo (0,016 secondi). In Figura 6.11, tale fenomeno è evidenziato, soltanto in corrispondenza del primo punto di controllo, da un'ellisse tratteggiata, ma può essere notato anche nelle Figure 6.12 e 6.13 di seguito all'acquisizione di ogni punto di controllo.

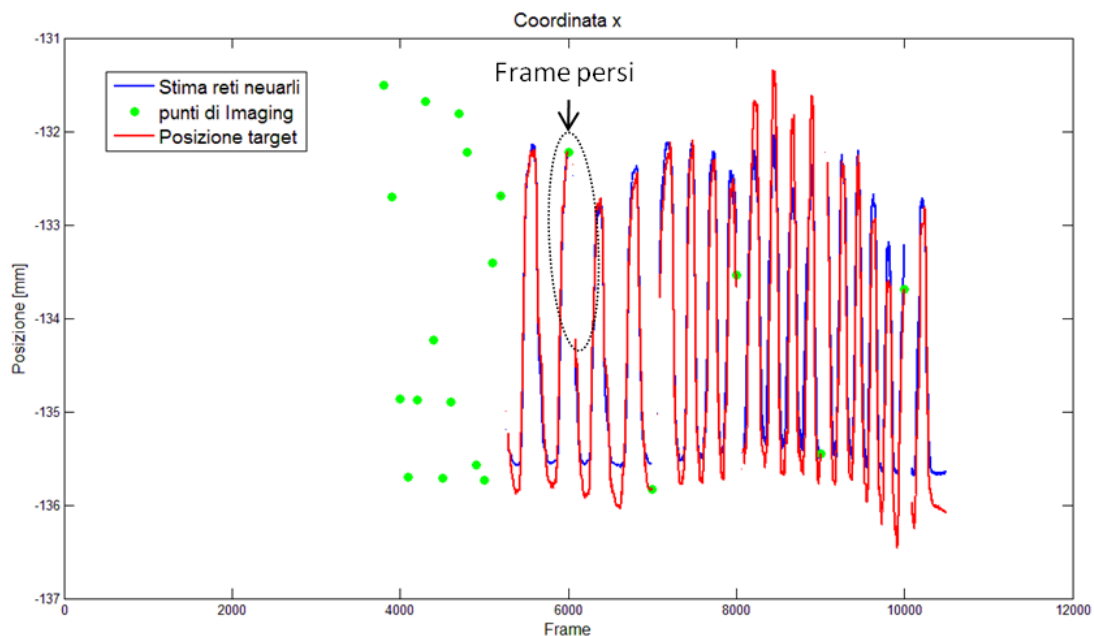


Figura 6.11: Coordinata x (AP). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di *imaging* acquisiti. L'ellisse nera tratteggiata evidenzia i *frame* persi a seguito del primo riaddestramento.

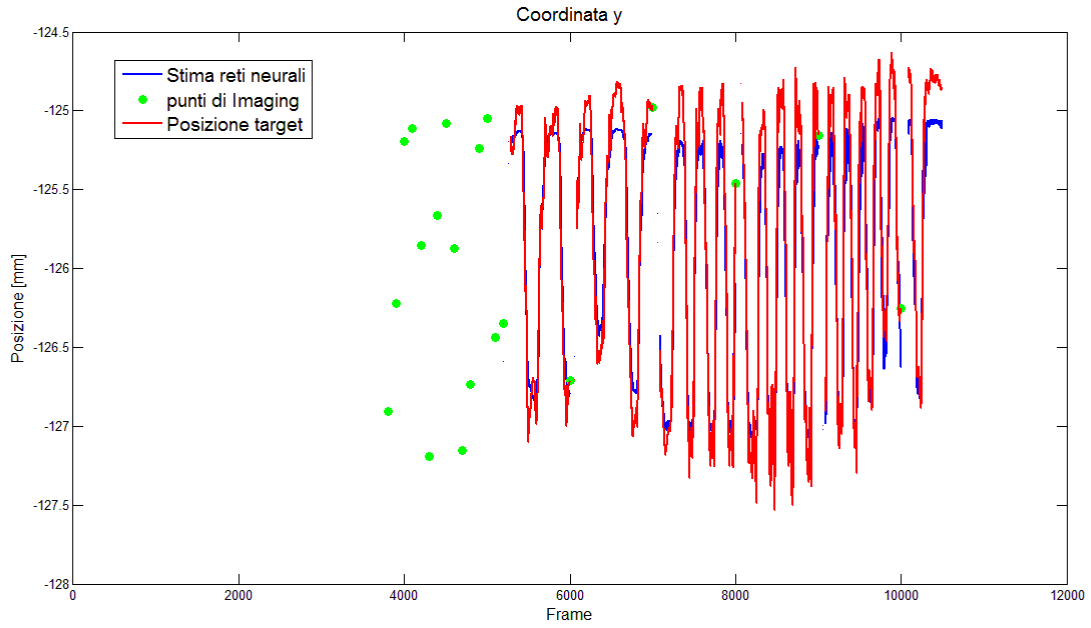


Figura 6.12: Coordinata y (CC). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di *imaging* acquisiti.

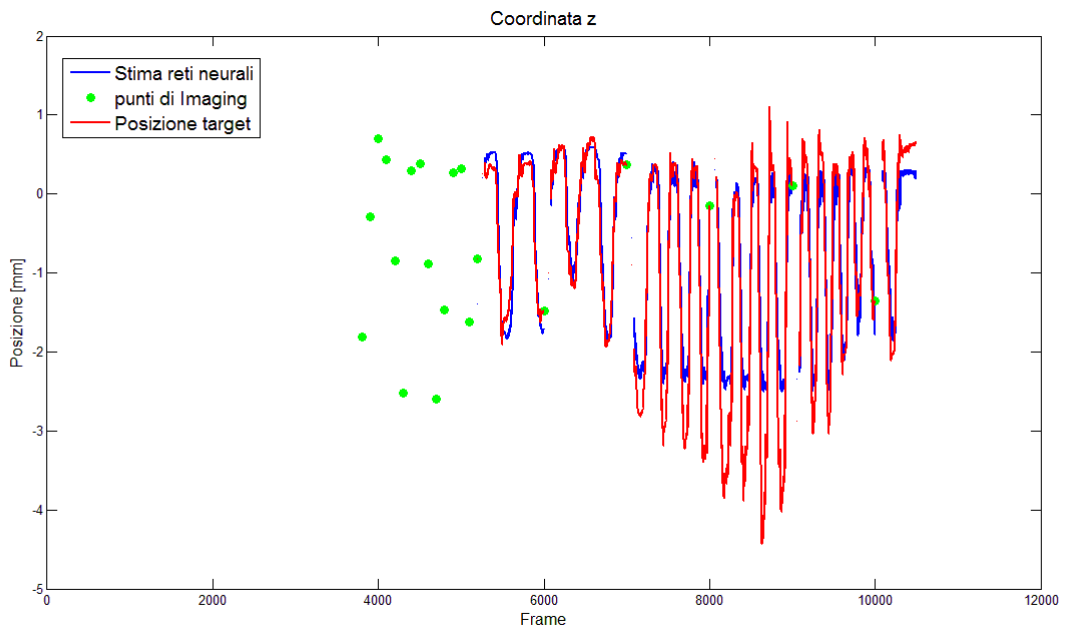


Figura 6.13: Coordinata z (LL). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di *imaging* acquisiti.

In ultimo, la Figura 6.14 mostra l'andamento dell'errore di puntamento 3D, sempre relativamente all'acquisizione numero 4.

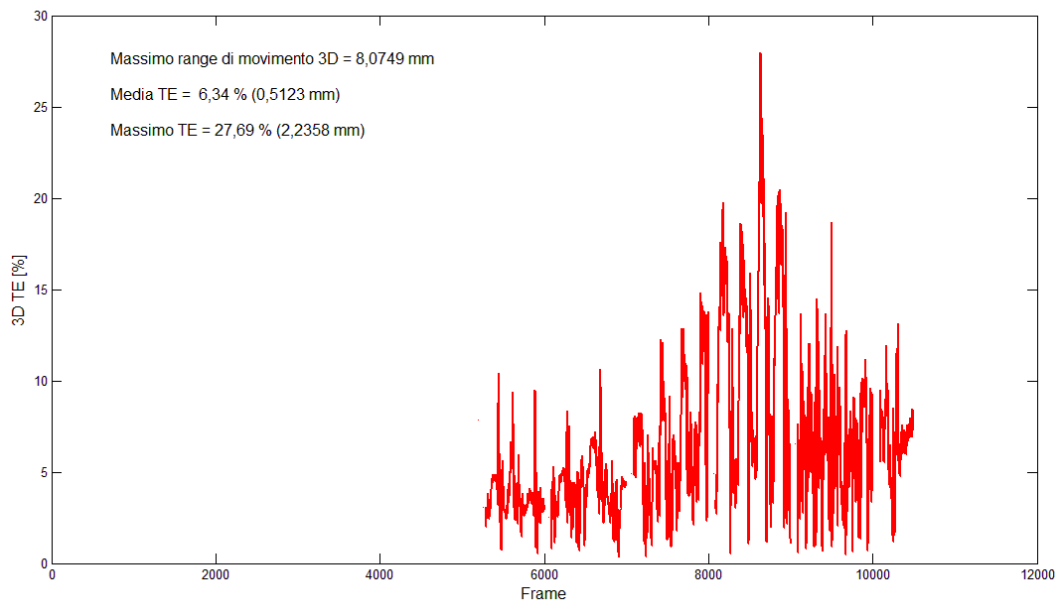


Figura 6.14: errore di puntamento 3D in corrispondenza di ogni *frame*, espresso come percentuale del massimo *range* di movimento 3D del bersaglio.

Acquisizioni online

Le acquisizioni *online* sono state effettuate utilizzando il fantoccio per la simulazione del respiro, ma a differenza delle acquisizioni *offline*, queste sono state eseguite installando l’algoritmo sviluppato direttamente sulla *workstation* del sistema SMART-D [Tabella 6.5]. Dunque è stato possibile stimare il segnale interno in corrispondenza di ogni *frame* acquisito, senza ricorrere all’analisi di un file .tdf relativo ad una acquisizione precedente. La Figura 6.15 mostra come sono stati ottenuti ed analizzati i risultati relativi alle acquisizioni *online*.

Caratteristiche del sistema SMART-D utilizzato

Processore

Modello Intel® Xeon® E5420

Numero di Core 4

Frequenza di Clock 2,5 GHz

Memoria RAM 2 GB

TVC

Numero TVC utilizzate 3

Risoluzione 512x512 pixel

Frequenza di acquisizione 60 Hz

Accuratezza <0.2mm su un volume di 3x2x2m

Tabella 6.5: Principali caratteristiche hardware della *workstation* e delle telecamere del sistema SMART-D [64].

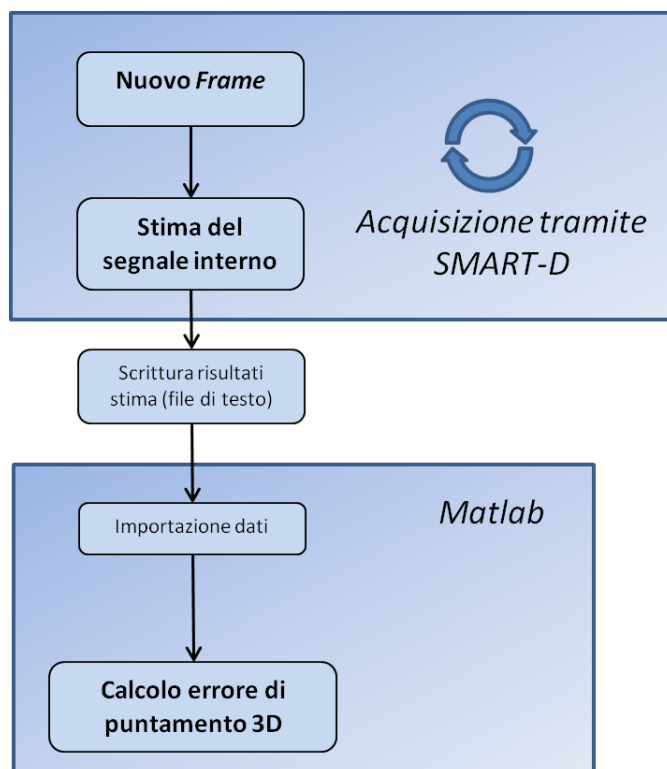


Figura 6.15: Schema a blocchi delle attività svolte per l'esecuzione e la valutazione delle analisi *online*.

Anche nel caso di questo studio su fantoccio, la stima del segnale interno (quarto marcatore) in funzione del segnale esterno (tre marcatori) è stata effettuata con reti neurali con architettura identica a quella descritta sia per le acquisizioni *offline* che per il confronto con il *modeler* di *CyberKnife*.

Come già anticipato all'inizio del presente capitolo, lo scopo principale delle analisi effettuate *online* sul fantoccio per la simulazione del movimento respiratorio consiste nel valutare la capacità del modello sviluppato di eseguire la stima del segnale interno ricevendo i dati direttamente da un sistema di localizzazione ottica. Risulta chiaro che, per poter ottenere una stima in tempo reale, è necessario che le operazioni necessarie per calcolare tale stima richiedano un tempo inferiore al periodo compreso tra l'acquisizione di due *frame* successivi che, si ricorda, è pari a 0,016 secondi operando ad una frequenza 60 Hz.

Differentemente dai casi analizzati *offline*, operando *online* il sistema SMART fornisce una lettura della posizione dei marcatori in ordine casuale, in quanto il riordinamento viene di norma eseguito in una fase successiva all'acquisizione (chiamata *Tracking*). Pertanto è stato necessario integrare nell'algoritmo sviluppato una funzione che ordina i marcatori acquisiti: si è scelto per semplicità, data la natura del problema, di ordinare i marcatori in base alla loro posizione in direzione latero-laterale. In tal modo l'algoritmo riceve la posizione di tali

marcatori sempre nello stesso ordine ed è quindi possibile discriminare i marcatori esterni da quello che simula il bersaglio.

Le acquisizioni *online* confermano i risultati già esposti relativamente alle acquisizioni *offline*. Dunque, anche in questo caso, il tempo di calcolo della stima risulta sempre inferiore a 0,016 secondi, mentre in concomitanza del riaddestramento delle reti neurali si è osservato che, in media, vengono persi circa 34 *frame*, corrispondenti a circa 0,57 secondi. Tale tempo risulta inferiore a quello riportato relativamente alle analisi *offline* (0,80 secondi), dal momento che queste ultime sono state eseguite con un *personal computer* [Tabella 6.6] avente una potenza computazionale inferiore a quella della *workstation* del sistema SMART-D.

Caratteristiche del <i>personal computer</i> utilizzato	
Processore	
Modello	Intel® Core™2 Duo T7300
Numero di Core	2
Frequenza di Clock	2,0 GHz
Memoria RAM	4 GB

Tabella 6.6: Principali caratteristiche hardware della *personal computer* utilizzato.

Analogamente a quanto esposto per l'analisi delle acquisizioni *offline*, la valutazione della performance del modello sviluppato, in termini di accuratezza della stima, è espressa in termini di errore di puntamento 3D in corrispondenza di ogni *frame* acquisito. La Tabella 6.7 mostra i risultati ottenuti.

Acquisizione	Numero di Frame	Massimo range di movimento 3D [mm]	Media TE[%]	Deviazione standard TE[%]	Massimo TE[%]
1	7685	29,8292	4,07	2,40	11,60
2	9820	30,3641	6,00	8,56	42,37
3	9713	32,0333	4,40	1,95	15,82
Valori medi			4,82	4,31	23,26

Tabella 6.7: Dati relativi alle analisi effettuate su tre acquisizioni *online*.

Si osservi che gli errori percentuali appena riportati risultano inferiori a quelli presentati in Tabella 6.4. Questo conferma che, in presenza di *range* di movimento più ampi, il prototipo è in grado di apprendere la correlazione interno/esterno con maggiore efficacia rispetto ai casi in cui tale *range* è particolarmente limitato.

Di seguito vengono mostrati come esempio le tracce relative alla prima acquisizione [Figure 6.16, 6.17 e 6.18], nelle quali si può osservare sia l'effettiva posizione del bersaglio (curva in rosso) sia la stima effettuata *online* dall'algoritmo di correlazione (curva in blu). I punti acquisiti per l'addestramento iniziale e per la valutazione della stima sono evidenziati in verde. Gli intervalli in cui non si ha disposizione né la posizione del bersaglio né la sua stima, corrispondono ai *frame* persi a causa del riaddestramento delle reti neurali.

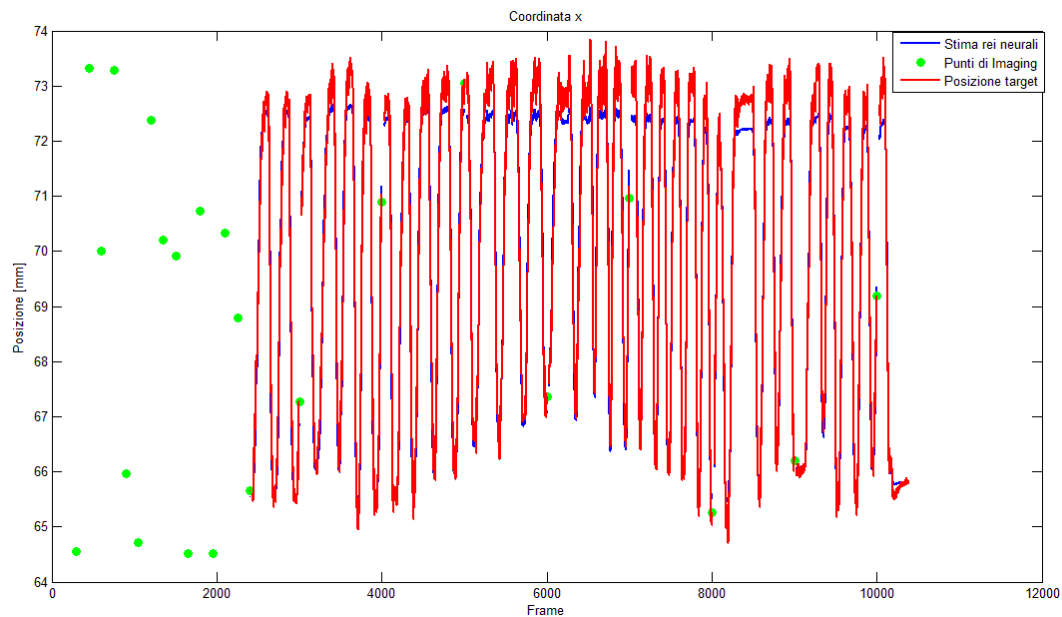


Figura 6.16: Coordinata x (AP). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di *imaging* acquisiti.

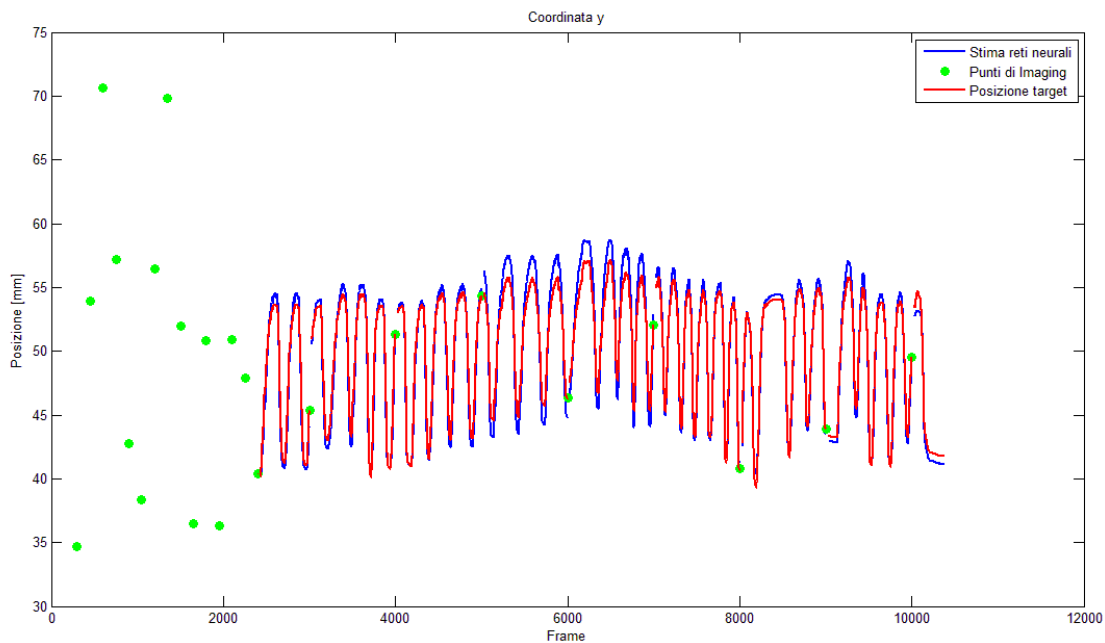


Figura 6.17: Coordinata y (CC). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di *imaging* acquisiti.

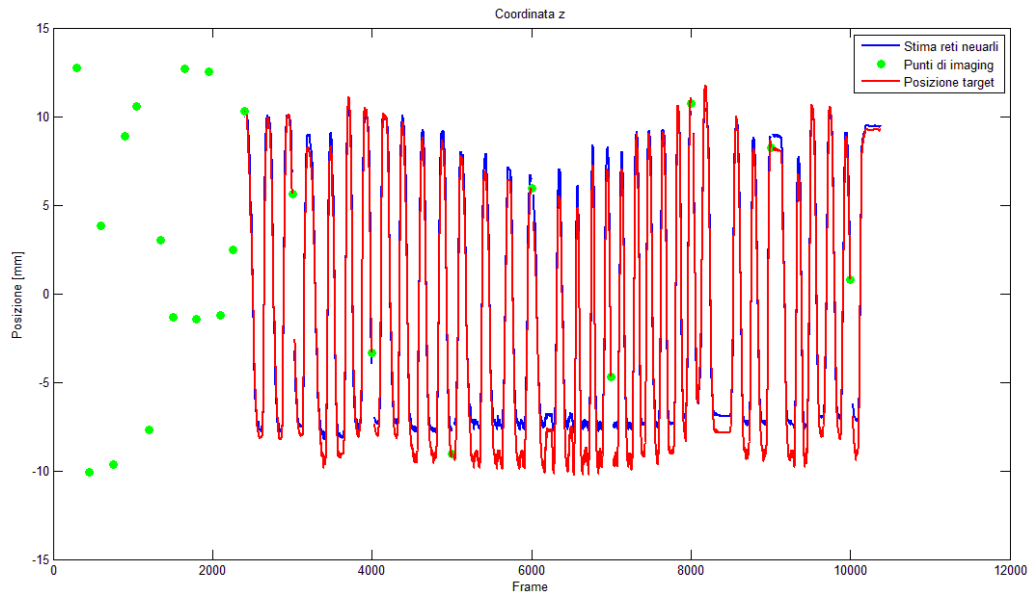


Figura 6.18: Coordinata z (LL). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di *imaging* acquisiti.

In tali figure è possibile osservare come nell'acquisizione siano state introdotte sia variazioni di ampiezza che di frequenza del movimento respiratorio, oltre ad alcune irregolarità come ad esempio la sospensione temporanea del movimento sia in fase di massima inspirazione che di massima espirazione.

Sempre relativamente alla prima acquisizione, la Figura 6.19 mostra l'andamento dell'errore di puntamento 3D.

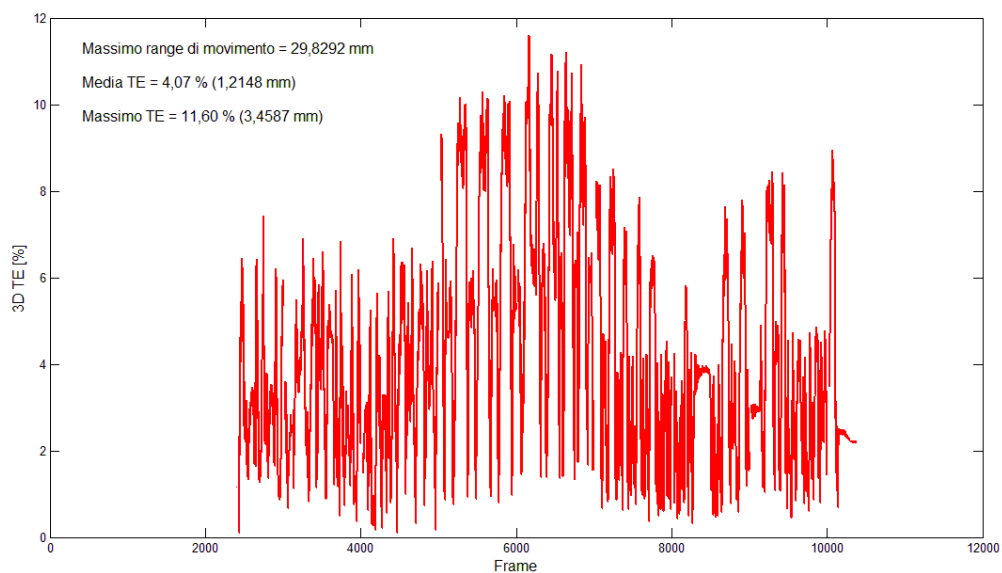


Figura 6.19: errore di puntamento 3D in corrispondenza di ogni *frame*, espresso in percentuale del massimo *range* di movimento 3D del bersaglio.

I risultati presentati in questo capitolo, sia relativi al confronto con il *modeler* di *CyberKnife* sia alle acquisizioni su fantoccio verranno discussi nel prossimo capitolo.

7. Discussione dei risultati e sviluppi futuri

Nell'ambito della radioterapia e, soprattutto, dell'adroterapia, garantire l'accuratezza con la quale si somministra la dose risulta di fondamentale importanza, sia per ottenere un efficace controllo della malattia che per evitare gravi effetti indesiderati. Garantire tale accuratezza è particolarmente difficoltoso nel caso in cui il bersaglio sia soggetto a movimenti fisiologici, come ad esempio la respirazione, il cui effetto non è ignorabile in un intervallo di tempo generalmente ricoperto da una frazione di trattamento (da pochi minuti a decine di minuti). Per limitare inaccurately nella localizzazione del target sono state sviluppate diverse tecniche, tra le quali il *gating*, che prevede il monitoraggio della fase respiratoria del paziente e l'irradiazione soltanto in un particolare intervallo del ciclo respiratorio (tipicamente la fase di massima espirazione), e il *tracking* che, invece, consiste nell'inseguire il movimento del bersaglio facendo sì che il fascio radiante sia sempre collimato su quest'ultimo.

Quest'ultima tecnica, ad oggi, è utilizzata in ambito clinico soltanto nel campo della radioterapia con fotoni. Il sistema di trattamento robotico *CyberKnife*, infatti, se dotato di un apposito modulo chiamato *Synchrony™ Respiratory Tracking System (RTS)*, è in grado di stimare la posizione del bersaglio in funzione di un segnale surrogato esterno, dato dal movimento della superficie corporea del paziente, e di inseguire il movimento del *target* in base alle stime calcolate [46].

Nell'ambito dell'adroterapia, invece, non vi sono, ad oggi, sistemi di *tracking* in uso clinico, ma soltanto a livello sperimentale [50,51] (paragrafo 3.4.2). Nel caso di trattamenti con particelle, infatti, tale sistema richiede una maggiore complessità, in quanto la compensazione del movimento non può avvenire soltanto sul piano perpendicolare al fascio ma deve avvenire anche in direzione longitudinale, tramite la modulazione dell'energia del fascio (paragrafo 2.5.2). È chiaro che, una volta implementati con successo, i sistemi di *tracking* permetterebbero, anche in adroterapia, un efficace trattamento delle lesioni mobili.

Tali sistemi, oltre a richiedere adeguati dispositivi per la somministrazione della dose (paragrafo 3.4.2), necessitano sia di uno strumento per la stima della posizione interna del bersaglio sia di uno in grado di predire tale posizione in istanti temporali successivi. Quest'ultimo si rende necessario dal momento che la modulazione del fascio (in direzione ed energia) necessita di tempi non trascurabili, nei quali il bersaglio si sposta dalla posizione stimata. Poiché, però, la stima della posizione del bersaglio costituisce la principale fonte di incertezza nella localizzazione del *target*, il progetto proposto è stato concentrato su tale aspetto, senza considerare le problematiche legate alla predizione.

Di conseguenza, il presente lavoro, come presentato nel capitolo 5, è stato focalizzato sull'obiettivo di sviluppare un metodo di correlazione interno/esterno, ovvero uno strumento in grado di apprendere la correlazione tra il movimento esterno di surrogati della superficie corporea e quello interno della lesione da irradiare, indipendentemente dalla collocazione anatomica di quest'ultima e dalla complessità del movimento a cui è soggetta. Infine, ci si è proposti di utilizzare la correlazione appresa per stimare *online* il segnale interno (posizione del bersaglio) in funzione del segnale esterno (posizione di marcatori esterni).

Si è proceduto effettuando, per prima cosa, uno studio preliminare (paragrafo 5.1) volto ad identificare quale strumento computazionale utilizzare per l'implementazione del modello di correlazione. Tale studio ha valutato la performance di un metodo basato su una rete neurale a doppio strato e di uno basato su un modello a variabili di stato [49,60]. I risultati del confronto tra le performance dei due metodi, condotto su reali casi clinici, hanno portato ad identificare nelle reti neurali lo strumento computazionale che, relativamente ai casi considerati, si è dimostrato più adeguato nell'apprendimento della correlazione interno/esterno.

Per questo motivo si è deciso di realizzare il modello proposto utilizzando algoritmi che implementano reti neurali. Nel corso dello sviluppo di quest'ultimo, il lavoro è stato orientato verso l'implementazione delle seguenti funzionalità:

- capacità di apprendere la correlazione interno/esterno,
- capacità di stimare in tempo reale il segnale interno in funzione del segnale esterno,
- possibilità di aggiornare la correlazione appresa,
- possibilità di integrare il prototipo in sistemi commerciali, come il sistema di localizzazione ottica SMART-D in uso presso il Centro Nazionale di Adroterapia Oncologica (CNAO).

La progettazione di tale prototipo è avvenuta in linguaggio C/C++, implementando un algoritmo che costruisce ed addestra tre distinte reti neurali, ognuna avente in ingresso il segnale proveniente da un unico marcatore esterno e in uscita la posizione stimata del bersaglio. L'uscita complessiva dell'algoritmo consiste nella media delle stime effettuate dalle singole reti. Tale algoritmo, inoltre, è in grado di riaddestrare una o più reti neurali se l'errore di stima, misurato in determinati punti di controllo in cui è disponibile l'informazione relativa alla posizione reale della lesione, risulta superiore ad una soglia imposta dall'utente. In aggiunta, il prototipo è stato anche modificato in modo da poter ricevere in ingresso un flusso di dati *real time*, generato ad esempio da un sistema di localizzazione ottica.

Terminato il suo sviluppo, il modello implementato è stato quindi validato. In una prima fase (paragrafo 6.1) si è valutata la sua capacità di stimare accuratamente la posizione del bersaglio, confrontando tale accuratezza con quella mostrata dal sistema *CyberKnife* in 20 distinte frazioni di trattamento.

Una seconda fase di validazione (paragrafo 6.2), invece, è stata effettuata per valutare la capacità di operare in tempo reale e di poter integrare il prototipo sviluppato all'interno di sistemi commerciali.

Per condurre quest'ultima validazione è stato impiegato un fantoccio per la simulazione del movimento respiratorio (paragrafo 5.3), appositamente ottimizzato. Tale fantoccio, infatti, prevedeva un azionamento tramite motore elettrico, il quale non consentiva la variazione dell'ampiezza del movimento durante l'esecuzione di quest'ultimo. Per questo motivo si è progettato un azionamento manuale, basato su una puleggia a sezione variabile appositamente dimensionata e realizzata. In aggiunta, si è anche ottimizzata la traiettoria di movimento del marcatore che, nel suddetto fantoccio, simula la posizione del bersaglio, in modo tale che questa risulti compatibile con quella di una lesione soggetta ad *organ motion*.

I risultati relativi alle due fasi di validazione appena citate sono stati presentati nel capitolo precedente e saranno discussi nei prossimi paragrafi.

7.1. Confronto con il sistema *CyberKnife*

Per prima cosa si discuterà dell'architettura e delle modalità di addestramento delle reti neurali utilizzate. Si ricorda che l'uscita complessiva del modello sviluppato, ovvero la stima del segnale interno, è stata ottenuta calcolando la media delle uscite di tre reti neurali distinte ma aventi la medesima architettura. Ogni rete presenta infatti tre ingressi (le coordinate di un marcatore esterno), ha un singolo strato nascosto formato da dieci neuroni con funzione di attivazione a tangente iperbolica e uno strato d'uscita formato da 3 neuroni (le coordinate della posizione stimata del *target*), sempre con funzione d'attivazione a tangente iperbolica.

Come è evidente, il numero di neuroni dello strato d'uscita e il numero di ingressi è dettato dal problema che si vuole affrontare e dalla scelta di realizzare tre reti distinte. Per quanto riguarda, invece, le decisioni circa la numerosità dello strato nascosto e le funzioni di attivazione, queste sono state prese considerando principalmente la natura dei segnali relativi ai trattamenti con il sistema *CyberKnife* ed in particolare quelli riguardanti la posizione dei marcatori esterni. Tali segnali, infatti, sono stati acquisiti durante reali trattamenti radioterapici e quindi presentano caratteristiche e problematiche legate alla situazione patologica del soggetto.

A titolo di esempio si osservi la Figura 7.1 in cui è riportato un estratto dell'andamento della coordinata x (direzione cranio-caudale, CC) di un marcatore esterno durante uno dei trattamenti analizzati.

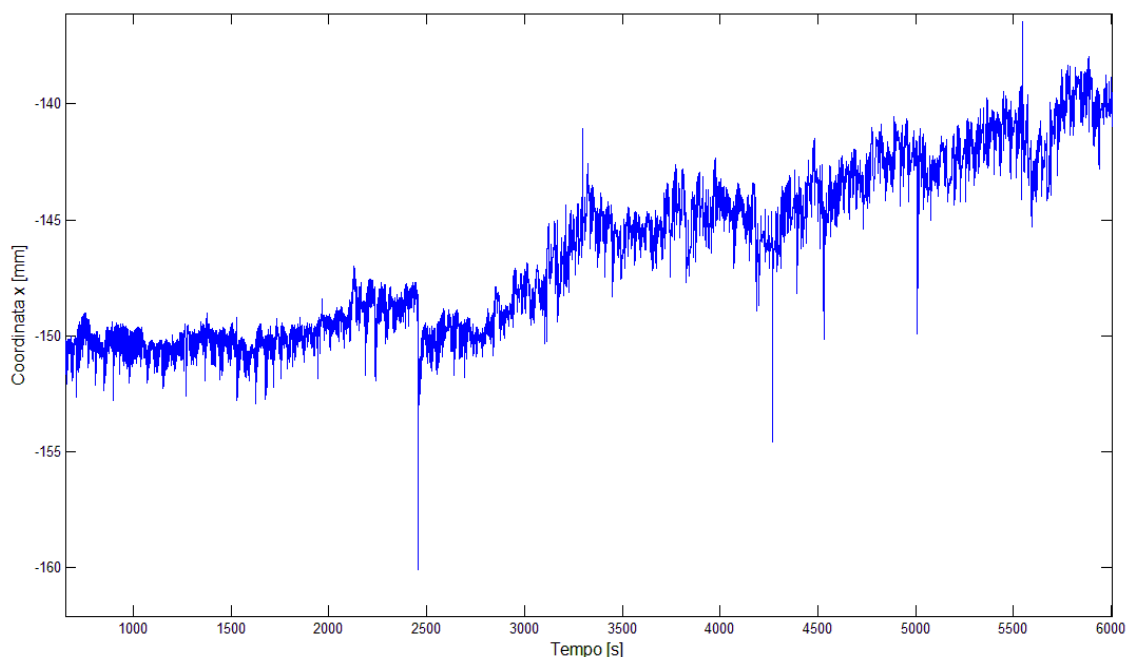


Figura 7.1: Coordinata x (CC) di un marcatore esterno in un trattamento analizzato eseguito da CyberKnife.

Di conseguenza si è deciso di utilizzare reti neurali sufficientemente complesse per poter apprendere la correlazione interno/esterno anche in presenza di tali segnali.

Allo scopo di ottenere un soddisfacente compromesso tra complessità e costo computazionale, è stato quindi adottato un unico strato nascosto formato da 10 neuroni.

Per quanto riguarda le funzioni di attivazione, la scelta di adottare la tangente iperbolica in entrambi gli strati è stata effettuata con l'intenzione di sfruttare l'effetto di saturazione per favorire la stabilità dell'uscita. Come è noto, infatti, la funzione tangente iperbolica tende asintoticamente a valori costanti al crescere del modulo del suo argomento. Di conseguenza, eventuali picchi improvvisi, probabilmente attribuibili a rumore o gravi irregolarità presenti sul segnale in ingresso, non vengono riportati sull'uscita.

In ultimo, si sottolinea che l'introduzione di funzioni di trasferimento non lineari, come la tangente iperbolica, è necessario in quanto la correlazione interno/esterno che deve essere appresa è una funzione certamente non lineare.

Dal confronto statistico tra le distribuzioni degli errori di puntamento 3D commessi dal *modeler* e di quelli commessi dall'algoritmo che implementa le reti neurali appena discusse (paragrafo 6.1), si osserva una significativa differenza tra le due ed in particolare si osserva che, sia nel caso del *subset control* che nel caso del *subset worst*, il modello di correlazione

sviluppato esegue, relativamente ai punti in cui si ha l'acquisizione di immagini di controllo, stime della posizione interna del *target* con maggiore accuratezza rispetto al *modeler*.

Si precisa, però, che l'errore di puntamento 3D può essere calcolato solo in corrispondenza dell'acquisizioni di immagini radiografiche, ovvero quando si ha a disposizione, nello stesso istante, sia la stima della posizione del bersaglio sia la sua effettiva collocazione. È quindi chiaro che, sia per quanto riguarda il modello sviluppato sia per quanto riguarda il *modeler*, non è possibile valutare l'accuratezza della stima al di fuori dei punti di *imaging*.

I risultati relativi al confronto appena citato mostrano che le reti neurali sono uno strumento potenzialmente valido per implementare un sistema di *tracking* in quanto si è dimostrato che sono in grado di apprendere la correlazione interno/esterno e di stimare la posizione del bersaglio con accuratezza confrontabile, relativamente ai casi analizzati, a quella di un sistema commerciale.

Nonostante l'algoritmo sviluppato si sia mostrato in grado di apprendere la correlazione interno/esterno e di stimare il segnale interno con una accuratezza che, relativamente ai casi analizzati e limitatamente ai punti di *imaging* di controllo, si è mostrata maggiore rispetto a quella del sistema *CyberKnife*, è chiaro che esso necessita di un processo di ottimizzazione che non è stato svolto in modo sistematico nel corso del progetto proposto, il cui obiettivo, si ricorda, consiste nell'ottenere un prototipo. Tale ottimizzazione dovrà riguardare sia l'architettura delle reti neurali sia le modalità con cui queste vengono addestrate. In particolare si evidenzia la necessità di introdurre metodi di addestramento più sofisticati di una semplice *back propagation*. Ad esempio, potrebbe essere implementato un algoritmo di addestramento basato sul metodo di Levenberg–Marquardt. In questo modo, essendo tale algoritmo più efficiente nella minimizzazione della funzione costo, sarebbe possibile limitare il numero di epoche utilizzate per l'addestramento, riducendo di conseguenza i tempi richiesti da questa procedura. Si precisa, però, che tale metodo comporta un notevole incremento del costo computazionale, in quanto richiede il calcolo e l'inversione della matrice Hessiana, dunque può essere impiegato solo nel caso in cui tale incremento di costo computazionale non causasse un aumento dei tempi di addestramento tale da non riscontrare più alcun vantaggio rispetto alla semplice *back propagation*.

In aggiunta, si ritiene necessario ottimizzare il numero di esempi con il quale costruire il *training set* per l'addestramento iniziale e le soglie in base alle quali si decide riguardo alla necessità di eseguire un riaddestramento. Nell'ambito di questo studio, il numero di esempi contenuti nel *training set* è stato scelto coincidente con il numero di acquisizioni radiografiche che il sistema *CyberKnife* esegue prima dell'irradiazione al fine di costruire il modello di

correlazione. Di conseguenza tale numero differisce a seconda del trattamento considerato. A questo proposito sarebbe utile valutare se è possibile utilizzare sempre un dato numero di esempi, definito tramite uno studio di ottimizzazione, o se tale numero deve essere personalizzato per ogni paziente.

Riguardo alle soglie utilizzate per valutare la necessità del riaddestramento, non disponendo di informazioni dettagliate riguardo alle modalità con le quali il sistema *CyberKnife* aggiorna il modello di correlazione, sono stati impostati valori di 1 mm per la soglia relativa all'addestramento senza reset dei pesi e 2 mm per quella relativa all'addestramento con reset dei pesi (paragrafo 5.2.2), ritenendo quindi accettabili soltanto errori di stima sub-millimetrici, in quanto confrontabili con l'accuratezza di un sistema di localizzazione ottica.

In ultimo, si sottolinea la necessità di valutare accuratamente l'ipotesi di stazionarietà assunta per il segnale respiratorio, in quanto da questa dipende l'affidabilità dell'interpolazione effettuata allo scopo di ottenere un sufficiente numero di coppie interno/esterno per l'apprendimento della correzione da parte delle reti neurali.

7.2. Attività sperimentale su fantoccio

L'attività sperimentale svolta con l'ausilio del fantoccio per la simulazione del movimento respiratorio è stata utilizzata per valutare la capacità dello strumento sviluppato di operare in tempo reale, ovvero di essere in grado di calcolare la stima del segnale interno in funzione di ogni campione del segnale esterno reso disponibile da un sistema di localizzazione ottica.

I test effettuati mostrano che è possibile operare in *real time* nel caso in cui l'acquisizione della posizione dei marcatori avvenga ad una frequenza minore o uguale a 60 Hz. Di conseguenza, fatta eccezione per gli istanti in cui l'algoritmo riaddestra una o più reti neurali, in corrispondenza di ogni *frame* acquisito può essere calcolata la stima del segnale interno. Si è anche calcolato che la procedura di riaddestramento comporta la sospensione della stima del segnale interno per un tempo inferiore ad un secondo, tempo per il quale l'irradiazione dovrebbe essere necessariamente sospesa. Le tempistiche appena esposte consentirebbero, in ogni caso, di ottenere un elevato *duty cycle* (rapporto tra tempo di irradiazione e tempo totale di trattamento), riducendo di conseguenza la durata di ogni singola frazione senza rinunciare ad un *targeting* accurato. Questo costituirebbe il principale vantaggio del *tracking* rispetto alle tecniche di *gating*, dove il *duty cycle* risulta notevolmente minore, dal momento che la dose è somministrata soltanto durante una ridotta frazione del ciclo respiratorio.

Le prove eseguite *online*, ovvero installando l’algoritmo di correlazione direttamente sulla *workstation* del sistema SMART-D, hanno mostrato che il prototipo sviluppato può essere integrato all’interno di sistemi più complessi. Inoltre, dal momento che il suddetto sistema di localizzazione ottica è installato nelle sale di trattamento del CNAO, di fatto si è condotto con successo un primo studio di fattibilità riguardo alla possibilità di integrazione nei sistemi di tale struttura.

Oltre a permettere di valutare il prototipo riguardo alla capacità di operare in tempo reale, le prove su fantoccio hanno permesso di valutare l’accuratezza della stima, non soltanto in corrispondenza di alcuni punti, bensì in corrispondenza di ogni *frame* acquisito.

Si ricorda, infatti, che tali prove sono state effettuate acquisendo la posizione di quattro marcatori, dei quali tre sono stati utilizzati per la stima della posizione del quarto, che simula il bersaglio [Figura 7.2].

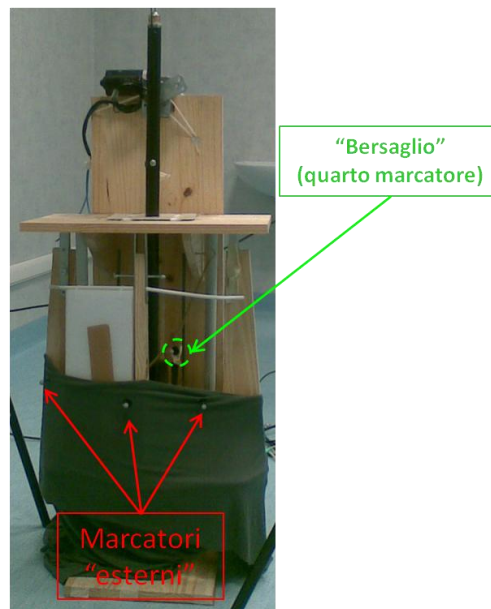


Figura 7.2: Fantoccio per simulazione del movimento respiratorio. Sono evidenziati i marcatori che simulano il segnale esterno (in rosso) e quello che simula il bersaglio (in verde).

In questo modo è stato possibile confrontare l’andamento, istante per istante, della posizione stimata con quello dalla posizione effettiva del bersaglio, raccogliendo informazioni riguardo all’errore di stima altrimenti irreperibili. Non è infatti assolutamente possibile raccogliere informazioni così dettagliate effettuando acquisizioni su pazienti o su soggetti sani, dal momento che la localizzazione in tempo reale del bersaglio richiederebbe procedure altamente invasive quali l’acquisizione di immagini radiografiche temporalmente ravvicinate o fluoroscopiche.

Premettendo che le condizioni sperimentali in cui si è operato, pur essendo realistiche in termini di ampiezza e frequenza dei movimenti eseguiti, sono decisamente più controllate di una situazione di reale impiego clinico di un sistema di *tracking*, dai grafici riportati nel capitolo precedente [Figure 6.11, 6.12 e 6.13] si è notato che le reti neurali, aventi una architettura identica a quella descritta nel paragrafo precedente, permettono di ottenere un'uscita stabile e poco rumorosa. Inoltre, si osserva che le variazioni di frequenza non causano alcuna riduzione dell'accuratezza, in quanto l'andamento della stima si adatta alla frequenza del respiro senza alcun ritardo temporale.

Si è notato che il principale fattore che limita l'accuratezza della stima consiste nella evidente dipendenza di quest'ultima dalla posizione dei punti con i quali è stata addestrata la rete. In particolare, si osserva che l'accuratezza risulta soddisfacente se il *target* si trova all'interno del *range* di variabilità dei punti con i quali è avvenuto l'addestramento. Se, invece, il bersaglio si trova fuori da tale *range*, allora la stima perde accuratezza.

Relativamente agli errori di stima percentuali presentati nel paragrafo 6.2 (Tabella 6.4), si sottolinea che questi sono riferiti a movimenti 3D del bersaglio la cui ampiezza è limitata. Nonostante ciò, il loro valore in termini assoluti, risulta, in media, sub-millimetrico quindi clinicamente accettabile.

Da un punto di vista clinico, tale problematica potrebbe essere affrontata progettando una tecnica di irradiazione che combini il *tracking* con il *gating*, allo scopo di limitare gli svantaggi associati a ciascuna dei suddetti metodi. Sarebbe possibile irradiare il bersaglio con un fascio fisso durante la fase di massima espirazione, ovvero quando si suppone che i movimenti dovuti all'*organ motion* siano trascurabili, eseguendo il *tracking* durante il transitorio tra la massima espirazione e la massima inspirazione (e viceversa). In quest'ultima fase, nella quale la stima della posizione del bersaglio risulta scarsamente accurata, l'irradiazione potrebbe essere sospesa. Con una simile tecnica sarebbe possibile limitare sostanzialmente sia il principale svantaggio del *gating* (il ridotto *duty cycle*), in quanto il fascio rimarrebbe attivo per la maggior parte della durata del ciclo respiratorio, sia la principale limitazione del modello implementato, ovvero l'inaccuratezza in prossimità dei massimi e dei minimi del ciclo respiratorio.

La limitazione appena esposta può essere spiegata osservando che i segnali ottenuti tramite il fantoccio sono stati acquisiti in condizioni molto più controllate di quelle relative ai trattamenti tramite *CyberKnife*, per i quali l'architettura delle reti neurali precedentemente descritta era stata progettata. Le acquisizioni su fantoccio, infatti, sono avvenute mantenendo quest'ultimo fisso in una determinata posizione ed eseguendo movimenti respiratori ripetibili. Dunque, utilizzando i segnali così ottenuti, si può considerare l'adozione di reti neurali meno complesse

o di una combinazione di reti con differenti architetture. Si potrebbe, ad esempio, ottenere la stima utilizzando un'alternanza di due reti con funzione di attivazione differenti, una lineare e una non lineare come, ad esempio, la tangente iperbolica. La prima potrebbe essere utilizzata negli istanti in cui il ciclo respiratorio è prossimo alla massima espirazione e alla massima inspirazione, in quanto è in grado di fornire uscite anche esterne al *range* di variabilità degli esempi del *training set*. La seconda, invece, può essere usata nelle restanti parti del ciclo. Un approccio di questo tipo, sebbene costituisca uno sviluppo futuro reso possibile dagli strumenti che sono stati implementati, necessiterebbe di una specifica fase validazione da svolgersi nell'ambito di un apposito studio.

A titolo di esempio, la Figura 7.3 propone un confronto tra la stima della coordinata z (direzione antero-posteriore, AP) del *target* ottenuta con reti a singolo strato e funzione d'attivazione lineare e quella ottenuta con reti a doppio strato e funzione d'attivazione a tangente iperbolica.

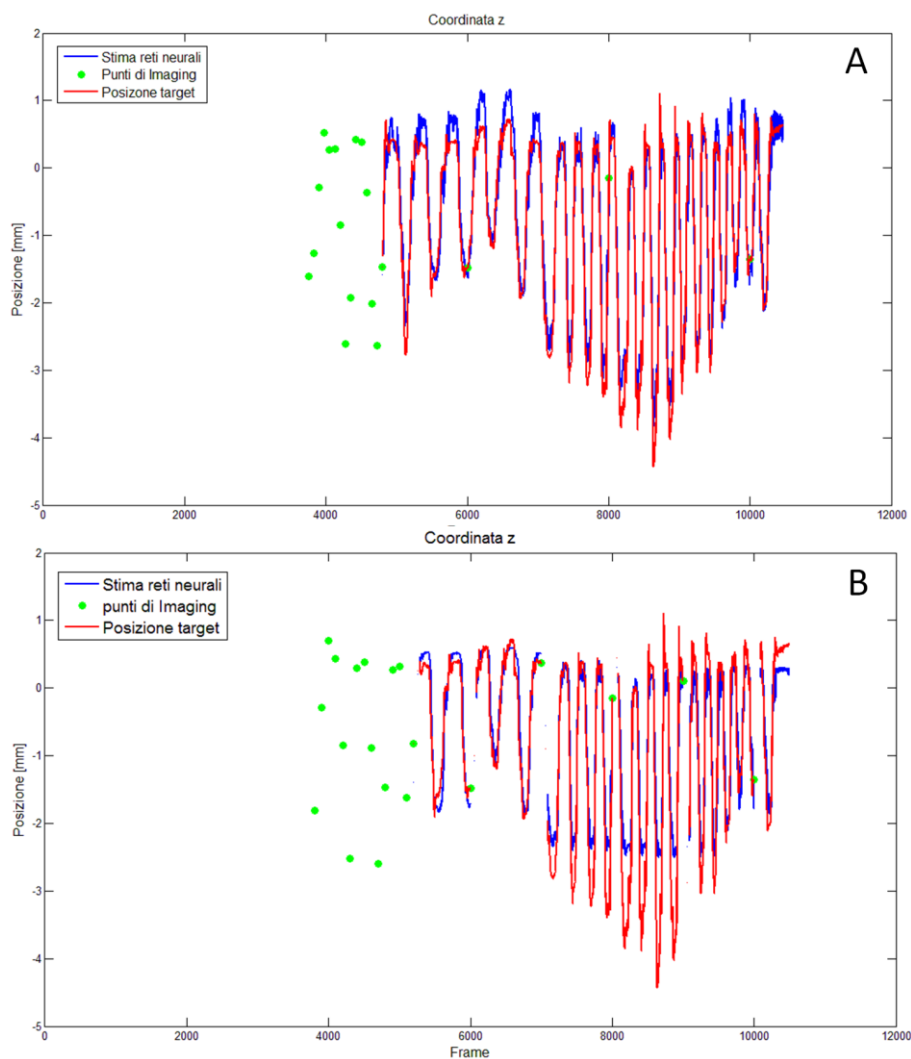


Figura 7.3: Pannello A: stima della coordinata z (AP) del bersaglio con rete a singolo strato (f.d.t lineare).
 Pannello B: stima della coordinata z (AP) del bersaglio con rete a due strati (f.d.t tanh).

Come si osserva in figura, la rete a singolo strato con funzione di trasferimento lineare è in grado di stimare accuratamente la posizione del bersaglio anche in prossimità dei massimi e dei minimi del segnale respiratorio. Si nota, però, che tale stima presenta una instabilità visibilmente maggiore di quella osservata nel caso della rete neurale a doppio strato e con funzione di trasferimento a tangente iperbolica, soprattutto nel momento in cui si sospende il movimento respiratorio. Quanto detto si osserva nel dettaglio relativo all'ultimo ciclo respiratorio, proposto in Figura 7.4.

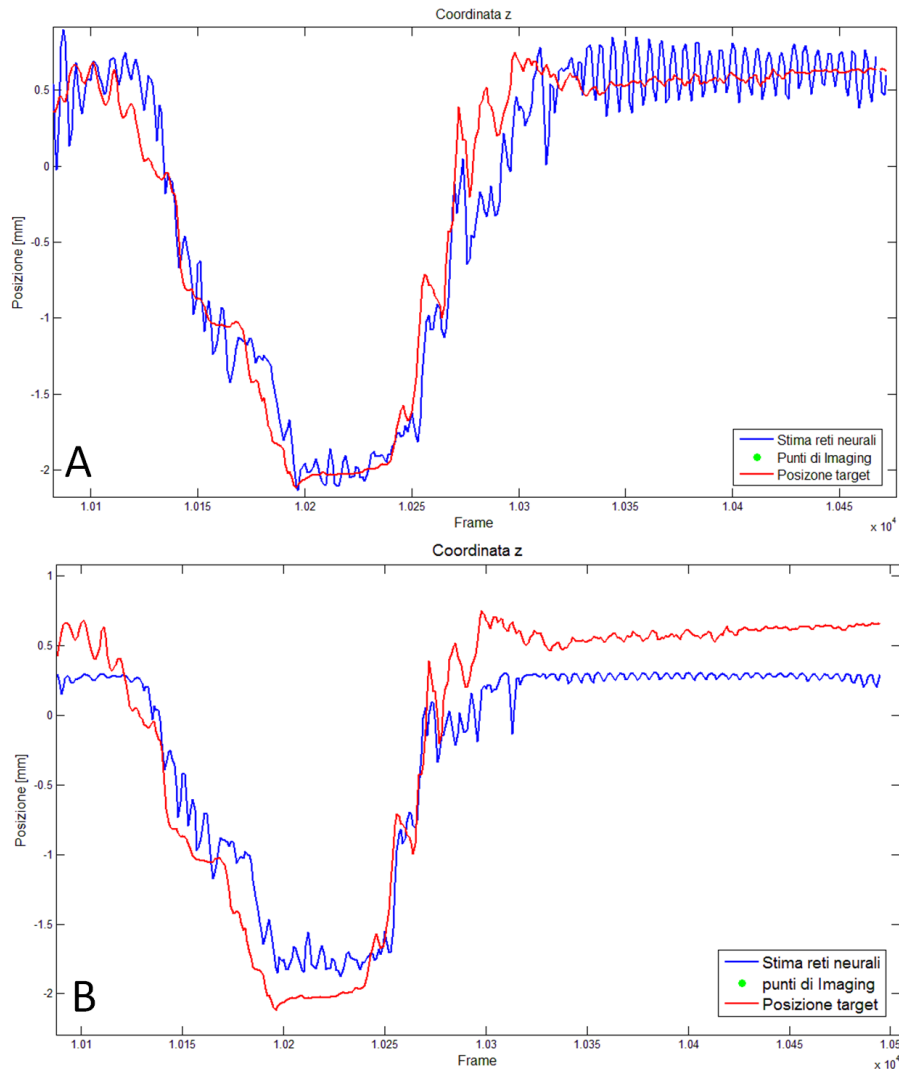


Figura 7.4: Pannello A: dettaglio della stima della coordinata z (AP) del bersaglio con rete a singolo strato (f.d.t lineare). Pannello B: dettaglio della stima della coordinata z (AP) del bersaglio con rete a due strati (f.d.t tanh).

Le prove eseguite su fantoccio hanno permesso, oltre che di testare la possibilità di operare in tempo reale, anche di valutare l'accuratezza delle stime effettuate per ogni singolo *frame* acquisito. Per questi motivi si ritiene che il fantoccio per la simulazione del respiro sia stato un ausilio estremamente valido nell'ambito dello studio condotto e, come tale, possa essere ulteriormente ottimizzato. Ad esempio, le parti in legno potrebbero essere sostituite con

materiali plastici (quali il PMMA), con il duplice vantaggio della riduzione degli attriti tra parti mobili e della possibilità di utilizzare tale fantoccio anche per studi basati su acquisizioni radiografiche o tomografiche, impossibili al momento data la presenza di numerosi elementi metallici che comporterebbero artefatti nelle immagini acquisite. È chiaro che, per poter effettuare tali studi è necessaria anche la progettazione e l'implementazione di un sistema di azionamento automatico in grado di consentire sia variazioni di frequenza che di ampiezza del movimento. A tal fine potrebbero essere impiegati due distinti motori elettrici passo-passo, governati da un'apposita elettronica di controllo: uno per fornire il movimento rotatorio e un secondo per impostare la lunghezza della manovella, in modo tale da poter modificare l'ampiezza del movimento durante la sua esecuzione.

Inoltre, nel caso i cui si volessero ottenere movimenti del bersaglio di maggiore ampiezza, si renderebbe necessario, oltre all'azionamento tramite motori elettrici appena discusso, anche una riprogettazione delle struttura di supporto del fantoccio, in modo tale da permettere al pistone un escursione maggiore. In alternativa, sarebbe anche possibile modificare le parti mobili della cassa toracica e i supporti elastici che sorreggono il marcatore che simula i *target*, affinché, a parità di escursione verticale del pistone, si abbiano movimenti più ampi delle suddette parti.

In ultimo, si ritiene che tale fantoccio possa essere utilizzato anche nell'ambito di studi volti ad ottimizzare il numero e la disposizione dei marcatori esterni, in quanto permetterebbe di testare, in modo semplice e rapido, varie configurazioni di marcatori e di osservare, in corrispondenza di ognuna, l'andamento della stima del segnale interno.

7.3. Osservazioni conclusive

In conclusione dell'intero elaborato e alla luce dei risultati esposti e commentati precedentemente, si sottolinea che lo strumento sviluppato, pur avendo dimostrato di poter operare in tempo reale e, relativamente ai casi considerati, di eseguire stime (in corrispondenza dei punti di *imaging* di controllo) con accuratezza superiore al sistema *CyberKnife*, necessita di un esteso processo di ottimizzazione e di ulteriore validazione, allo scopo di affrontare e risolvere le limitazioni esposte nei paragrafi precedenti.

Nonostante ciò, si può allo stesso modo affermare che il prototipo realizzato implementa tutte le specifiche richieste (paragrafo 3.5) e potrebbe quindi costituire un valido ausilio per la realizzazione di un sistema di *tracking online* per l'adroterapia.

8. Bibliografia

1. Medicine OnLine [<http://www.meds.com/index.html>]. Stratford, CT. Medicine OnLine An Affiliate of UltiTech, Inc; [consultato il 09/08/10].
2. Nation Cancer Institute [<http://www.cancer.gov>]. Bethesda, MD. National Institutes of Health (NIH). [consultato il 09/8/10].
3. Ekberg L, Holmberg O, Wittgren L, Bjelkengren G, Landberg T. What margins should be added to the clinical target volume in radiotherapy treatment planning for lung cancer? *Radiotherapy and Oncology* 48(1), 71-77 (1998).
4. Agosteo S, Adroterapia con protoni e ioni leggeri, dispense del corso di “Applicazioni Medicali delle radiazioni”, Politecnico di Milano, AA2008-2009.
5. Wilson RR. Radiological Use of Fast Protons. *Radiology* 47, 487-491 (1946).
6. Bragg W, Kleemann R. On the α -Particles of Radium and their Loss of Range in Passing Through Various Atoms and Molecules. *Phil. Mag.* 3, 18-340 (1903).
7. Kraft G. The Radiobiological and Physical Basis for Radiotherapy with Protons and Heavier Ions. *Strahlenther. Onkol.* 166, 10-13 (1990).
8. Sisterson J. Particles, a Newsletter for those Interested in Proton Light Ion and Heavy Charged Particle Radiotherapy, Particle Therapy Cooperative Group (PTCOG) Number 29, January 2002.
9. Lagniel JM. Status of the hadrontherapy projects in Europe. Proceedings of Particle Accelerator Conference 2007. 2007 June 25-29; Albuquerque, NM.
10. Kaft G. Tumor Therapy with Heavy Charged Particles, Progress in Particle and Nuclear. *Physics* 45, S473-S544 (2000).
11. Kraft G. Radiobiological effects of highly charged ions: Their relevance for tumor therapy and radioprotection in space, The Physics of Highly and Multiply Charged Ions Ed. F. J. Currell Kluwer Academic Publisher (2002).
12. Barkas HW. Nuclear Research Emulsions, Vol. I, Academic Press New York and London (1963).
13. Harvard Medical School [<http://www.med.harvard.edu/jpnm/physics/nmltd/radprin>] Boston, MA. [consultato il 12/08/2010].
14. Ahlen SP. Theoretical and Experimental Aspects of the Energy Loss of Relativistic Heavily Ionizing Particles. *Rev. Mod. Phys* 52(12), 1-173 (1980).
15. Highland VL. Some Practical Remarks on Multiple Scattering. *Nucl. Instr. and Meth.* 129-497 (1975).

16. Weber U. Volumenkonforme Bestrahlung mit Kohlenstoffionen, PhD-Thesis, Universitat Gh Kassel, (1996).
17. Enghardt W, Fromm WD, Geissel H, Keller H, Kraft G, Magel A, Manfral P, Mtinzenberg G, Nickel F, Pawelke J, Schardt D, Scheidenberger C, Sobiella M. The Spatial Distribution of Positron-Emitting Nuclei Generated by Relativistic Light Ion Beams in Organic Matter. *Phys. Med. Biol.* 37(11), 2127-2131 (1992).
18. Agosteo S. Acceleratori lineari di elettroni per radioterapia, dispense del corso di “Applicazioni Medicali delle radiazioni”, Politecnico di Milano, AA2008-2009.
19. Agosteo S. Il Ciclotrone, dispense del corso di “Applicazioni Medicali delle radiazioni”, Politecnico di Milano, AA2008-2009.
20. Agosteo S. Il Sincrotrone, dispense del corso di “Applicazioni Medicali delle radiazioni”, Politecnico di Milano, AA2008-2009.
21. Symonds RP. Recent advances Radiotherapy; *BMJ* 323(7321), 1107–1110 (2001).
22. Dearnaley DP, Khoo VS, Norman AR. Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised trial. *Lancet.* 353(9149), 267-72 (1999).
23. Taylor A, Powell MEB. Intensity-modulated radiotherapy—what is it?. *Cancer Imaging* 4(2): 68–73 (2004).
24. Chu WT, Ludewigt BA, Renner TR. Instrumentation for Treatment of Cancer Using Proton and Light-Ion Beams, Review of Scientific Instruments, Accelerator & Fusion Research Division, LBL-33403 UC-406 Preprint (1993).
25. Haberer TH, Becher W, Schardt D, Kraft G. Magnetic scanning system for heavy ion therapy. *Nucl. Instr. and Meth. in Phys. Res. A* 330, 296-305, (1993).
26. Giordanengo S. “Sistemi di distribuzione per fasci di adroni terapeutici”. Centro Nazionale di Adroterapia Oncologica, Pavia, Italy (2010).
27. Donetti M. “Dose delivery at CNAO ”. Centro Nazionale di Adroterapia Oncologica, Pavia, Italy (2010).
28. Paganetti H, Niemierko A, Ancukiewicz M, Gerweck LE, Goitein M, Loeffler JS, Suit HD. Relative Biological Effectiveness (RBE) values for proton beam therapy, *Int. J. Radiation Oncology Biol. Phys.* 53(2),407–421 (2002).
29. Weyrather WK, Kraft G. RBE of carbon ions: Experimental data and the strategy of RBE calculation for treatment planning. *Radiotherapy and Oncology* 73(2), S161-S169 (2004).

30. Weyrather WK, Ritter S, Scholz M, Kraft G. RBE for Track-Segment Irradiation in Cell Lines of Differing Repair Capacity, *Int. J. Radiat. Biol.* 75, 1357-1364 (1999).
31. ICRU-Report 40. The Quality Factor in Radiation Protection. Int. Commission on Radiation, Units and Measurements, Washington (1986).
32. Kraft G, Scholz M, Bechthold U. Tumor Therapy and Track Structure, *Radiat. Environm. Biophys.* 38, 229-237 (1999).
33. Todd PW. Heavy-Ion Irradiation of Human and Chinese-Hamster Cells. *Radiat. Res.* 61, 288 (1965).
34. Blakely EA, Tobias CA, Ngo FQH, Curtis SB. Physical and Cellular Radiobiological Properties of Heavy Ions in Relation to Cancer Therapy Application, Biological and Medical Research with Accelerated Heavy Ions at the Bevalac, eds.: M.D. Pimcello, CA. Tobias, LBL 11220, 73-88, (1980).
35. Wambersie A. The Future of High-LET Radiation in Cancer Therapy, in: Chauvel P. and Wambersie A., Eulima Workshop on the Potential Value of Light Ion Beam Therapy, Publication No. EUR 12165, Commission of the European Community, Brussels, (1989).
36. Seco J, Sharp GC, Wu Z, Gierga D, Buettner F, Paganetti H. Dosimetric impact of motion in free-breathing and gated lung radiotherapy: A 4D Monte Carlo study of intrafraction and interfraction effects. *Med. Phys.* 35(1), 356–366 (2008).
37. Langen KM, Jones DTL. Organ motion and its management. *Int J Radiat Oncol Biol Phys.* 50, 265–278 (2001).
38. Mori S, Lu HM, Wolfgang JA, Choi NC, Chen GT. Effects of interfractional anatomical changes on water-equivalent pathlength in charged-particle radiotherapy of lung cancer. *J Radiat Res* 50(6), 513-9 (2009).
39. Dong L. “ In room CT imaging: Conventional CT”. AAPM Summer School, University of Windsor, Ontario, Canada, June 28-22 2006.
40. Varian Medical Systems [<http://www.varian.com>]. Palo Alto, CA. [consultato il 23/10/2010].
41. Dobbins JT. Tomosynthesis imaging: At a translational crossroads. *Med Phys.* 36(6), 1956–1967 (2009).
42. The PSI Proton Therapy Facility: Patient Handling. [http://radmed.web.psi.ch/asm/gantry/handl/n_handl.html]. Paul Scherrer Institut, Villigen, Switzerland. [consultato il 24/10/10].

43. Siemens Healthcare. [<http://www.medical.siemens.com>], Siemens AG Healthcare Sector, Erlangen, Germany. [consultato il 24/10/10]
44. Baroni G, "Patient Set-up Aspects of CNAO Experimental Phase", CNAO-NIRS Second Joint Symposium on Hadrontherapy, Pavia, Italy. March 20-21 2010.
45. Riboldi M, Pella A, Tagaste B, Baroni G, Ciocca M, Rossi S, Orecchia R. Installation and preliminary testing of a newly designed patient setup and monitoring system for particle beam therapy. Proceedings of PTCOG 48, Heidelberg, Germany, p91 (2009).
46. Schweikard A, Glosser G, Bodduluri M, Murphy MJ, Adler JR. Robotic Motion Compensation for Respiratory Movement during Radiosurgery. *Computer Aided Surgery* 5, 263–277 (2000).
47. CyberKnife® Brochure [<http://www accuray.com/products/CyberKnife-System.aspx>] Accuray Inc., Sunnyval, CA. [consultato il 26/08/2010]
48. Schweikard A, Shiomi H, Adler JR. Respiration tracking in radiosurgery. *Med Phys.* 31(10), 2738–2741 (2004).
49. Seppenwoolde Y, Berbeco RI, Nishioka S, Shirato H, Heijmen B. Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: a simulation study. *Med Phys.* 34(7), 2774-84 (2007).
50. Bert C, Saito N, Schmidt A, Chaudhri N, Schardt D, Rietzel E. Target motion tracking with a scanned particle beam. *Med Phys* 34, 4768-4771 (2007).
51. Bert C, Gemmel A, Saito N, Chaudhri N, Schardt D, Durante M, Kraft G, Rietzel E. Dosimetric precision of an ion beam tracking system. *Radiation Oncology* 5, 61 (2010).
52. Hertz J, Krogh A, Palmer RG. Introduction to the theory of neural computation, Capitolo 1, Introduzione. Redwood City, CA: Addison-Wesley (1991).
53. Dayhoff JE, Deleo JM. Artificial neural network, Conference on Prognostic Factors and Staging in Cancer Management: Contributions of Artificial Neural Networks and Other Statistical Methods. *Cancer Supplement* 91(8), 1615-1635 (2001).
54. Rosenblatt F, Principles in neurodynamics. Spartan Books, Washington, DC (1962).
55. Hertz J, Krogh A, Palmer RG. Introduction to the theory of neural computation, Capitolo 5, Il perceptrone semplice. Redwood City, CA: Addison-Wesley (1991).
56. Werbos PJ. Beyond regression: new tools for prediction and analysis in the behavioral sciences. Ph.D. thesis, Harvard Univ, Cambridge, MA (1974).
57. Widrow B, Hoff M. Adaptive switching circuits. August IREWESCON Convention Record, Part 4, 96–104 (1960).

58. Hertz J, Krogh A, Palmer RG. Palmer, Introduction to the theory of neural computation, Capitolo 6, Reti Multi strato. Redwood City, CA: Addison-Wesley (1991).
59. Ciceri M, Applicazione delle reti neurali per la predizione della tossicità acuta nel trattamento radiante del carcinoma prostatico, Tesi di Laurea, Politecnico di Milano, Milano, Italy, (2009).
60. Ruan, D, Fessler JA, Balter JM, Berbeco RI, Nishioka S, Shirato H. Inference of hysteretic respiratory tumor motion from external surrogates: a state augmentation approach. *Phys. Med. Biol.*, 53, 2923-2936 (2008).
61. Luenberger DG. Optimization by Vector Space Methods, New York: Wiley (1969).
62. Riboldi M, Real-Time Tumor Targeting in External Beam Radiotherapy. Brain Cancer, Tumor Targeting and Cervical Cancer. Nova Science Publishers. In Press (2010).
63. Ferrante S. Feed forward Networks: Codice Matlab, esercitazioni del corso di Neuroingegneria, AA2009-2010.
64. BTS Bioengineering [<http://www.btsbioengineering.com>]. BTS S.p.A, Garbagnate Milanese, Milano, Italia. [consultato il 18/10/2010]
65. Pella A, Riboldi M, Pedotti A, Baroni G. Design of a 4D breathing phantom for motion tracking studies in extra-cranial particle therapy, Particle Therapy Cooperative Group (PTCOG) 49. May 17-22 2010, Chiba, Japan.
66. Torshabi AE, Pella A, Riboldi M, Baroni G. Targeting Accuracy in Real-Time Tumor Tracking via External Surrogates: A Comparative Study. *Technology in Cancer Research and Treatment*, 9(6), December 2010.

Indice delle figure

Figura 1.1: Curva di Bragg per gli ioni carbonio accelerati a 290 MeV/u.	4
Figura 1.2: Schema a blocchi dell'iter seguito per lo svolgimento del progetto.	7
Figura 2.1: Spettro elettromagnetico.	18
Figura 2.2: Schema a blocchi dell'iter di un generico trattamento di radioterapia.	19
Figura 2.3: Definizioni ICRU 50 di GTV, CTV, PTV.	20
Figura 2.4: andamento del LET in funzione dell'energia per diverse particelle, dal protone allo ione uranio.	22
Figura 2.5: Carica efficace della particella in funzione della sua energia.	23
Figura 2.6: Andamento della deposizione di energia in funzione della profondità di penetrazione.	24
Figura 2.7: Dose in funzione della profondità in acqua per un fascio di ioni carbonio con energia iniziale di 290 MeV/u	24
Figura 2.8: Pannello A: allargamento trasversale del fascio in funzione della profondità di penetrazione.	26
Figura 2.9: curva di Bragg per un fascio di ioni carbonio con energia di 270 MeV/u in acqua: è evidenziato il contributo di frammenti leggeri (linea tratteggiata).	27
Figura 2.10: Acceleratore lineare con gantry isocentrico.	29
Figura 2.11: Schema delle parti fondamentali di un ciclotrone.	30
Figura 2.12: Accelerazione di una particella in un ciclotrone. In blu: traiettoria della particella. In rosso: campo magnetico.	31
Figura 2.13: Rappresentazione schematica del sincrotrone del CNAO.	33
Figura 2.14: Pannello A: Esempio di un micro-collimatore multi lamellare da applicare alla testa dell'acceleratore lineare	35
Figura 2.15: Picco di Bragg allargato per protoni, ioni neon e ioni carbonio.	37
Figura 2.16: Modulatore di energia a spirale.	38
Figura 2.17: Compensatore per un trattamento al fegato.	38
Figura 2.18: Pannello A: Sistema dispersore per ottenere una distribuzione trasversale di energia piatta.	39
Figura 2.19: raster scanning. Pannello A: andamento a zig-zag del fascio radiante. Pannello B: segnali di controllo dei magneti.	40
Figura 2.20: Strategia di deposito della dose con sistema attivo [25].	41
Figura 2.21: Curva di sopravvivenza per una coltura di cellule CHO irradiate con raggi X[30].	43
Figura 2.22: Calcolo dell'RBE ₁₀ dalle curve di sopravvivenza per un generico fascio di particelle.	44
Figura 2.23: Andamento dell'RBE ₁₀ in funzione del LET, per un fascio di ioni carbonio che irradia una coltura di cellule CHO.	45
Figura 2.24: Andamento dell'RBE in funzione del LET, per varie particelle.	45
Figura 2.25: Confronto tra l'RBE (in funzione del LET) di una coltura cellulare con forte capacità di riparazione e quello di una coltura con scarsa capacità di riparazione.	46
Figura 2.26: RBE in funzione della dose per frazione. Grafico A: studi in-vitro. Grafico B: studi in-vivo. I cerchi rappresentano fasci con energia < 100 MeV, i triangoli fasci con energia > 100 MeV.	48
Figura 2.27: RBE in funzione della distanza dal picco di Bragg, per gli ioni carbonio, neon, silicio ed argon.	49
Figura 2.28: Dose fisica e dose biologica (per tre differenti colture cellulari) in funzione della profondità di penetrazione in acqua.	49
Figura 2.29: OER in funzione del LET per una coltura di cellule V79 irradiate da ioni carbonio	50
Figura 2.30: RBE in funzione del LET per una coltura di cellule V79, in presenza ed in assenza di ossigeno.	50
Figura 2.31: RBE in funzione del LET (Pannello A) e del range residuo in acqua (Pannello B), per colture di cellule CHO-k1 e xrs-5.	51
Figura 2.32: RBE _α in funzione del rapporto β/α, per tre differenti valori di LET.	52
Figura 2.33: Dose relativa in funzione della profondità di penetrazione in acqua per Raggi X a 120 KeV, fotoni γ emessi dal ⁶⁰ Co, fotoni di Bremsstrahlung a 18 MeV, ioni carbonio a 250 MeV/u e a 300 MeV/u.	53
Figura 3.1: Schermata del software di gestione del trattamento, in cui si osservano i reperi radiopachi impiantati nel bersaglio (in questo caso, la prostata).	59
Figura 3.2: Sistema di IGRT basato in room CT.	59
Figura 3.3 Sistema Varian® On-Board Imager®, si nota la coppia sorgente-detettore installata sull'acceleratore lineare.	60
Figura 3.4: Rappresentazione schematica del funzionamento di un sistema di localizzazione ottica con marcatori riflettenti.	63

Figura 3.5: Sala Gantry 2 presso il PSI	64
Figura 3.6: Sala di trattamento presso l'HIT. Sono visibili sia il sistema di posizionamento robotizzato sia l'arco a C.	65
Figura 3.7: sistema di IGHT presso al GHMC. Si notano lo scanner CT e il lettino robotico.	66
Figura 3.8: Telecamera e sensore utilizzati per il gating respiratorio presso HIMAC.	66
Figura 3.9: Dispositivo per il gating respiratorio utilizzato presso GHMC.	67
Figura 3.10: Rappresentazione schematica dei tre sottosistemi integrati nelle sale di trattamento laterali del CNAO.	68
Figura 3.11: Pannello A: PPS vincolato al lettino. Pannello B: PPS vincolato alla sedia.	68
Figura 3.12: Distribuzione degli errori di posizionamento traslazionali e rotazionali del PVS installato presso il CNAO [45].	69
Figura 3.13: PVS. Pannello A: i tubi radiogeni e i detettori sono ritratti nel cilindro di supporto. Pannello B: i tubi radiogeni e i detettori sono estratti per l'utilizzo.	70
Figura 3.14: OTS installato presso CNAO. Sono visibili le tre telecamere del sistema SMART-D.	71
Figura 3.15: Sistema CyberKnife, con evidenziati i componenti principali.	73
Figura 3.16: Rappresentazione schematica delle principali componenti del sistema Synchrony.	73
Figura 3.17: Set-up sperimentale del sistema di tracking realizzato presso il GSI.	75
Figura 3.18: Schema a blocchi di un generico software di correlazione interno/esterno per un sistema tracking.	77
Figura 4.1: Pannello A: soluzione di un problema in maniera deduttiva. Pannello B: soluzione di un problema a scatola nera tramite ANN.	80
Figura 4.2: Rappresentazione schematica di un neurone. D: dendriti, S: Soma, A: Assone.	80
Figura 4.3: Modello di neurone di McCulloch e Pitts (MCP).	81
Figura 4.4: Schema di un perceptrone semplice avente 3 ingressi e 2 uscite.	82
Figura 4.5: Pannello A: MLP con 3 neuroni nello strato nascosto e 2 neuroni nello strato d'uscita. Pannello B: MLP con due strati nascosti (da 2 e da 3 neuroni) e un neurone nello strato d'uscita.	83
Figura 4.6: Esempio di rete retroazionata.	84
Figura 4.7: Generica rete multi strato feed forward con singolo strato nascosto.	86
Figura 4.8: Andamento dell'RMS per il training, il validation set e il testing set [59].	88
Figura 4.9: Modello a variabili di stato di un generico sistema.	90
Figura 4.10: Pannello A: Il sistema fisico è noto a priori. Pannello B: Il sistema fisico è incognito (problema a scatola nera).	92
Figura 5.1: Schema a blocchi dell'iter seguito per lo svolgimento del progetto. È evidenziato l'ambiente di sviluppo utilizzato in ogni fase.	96
Figura 5.2: Schema a blocchi dell'implementazione della rete neurale in Matlab/Simulink.	99
Figura 5.3: Topologia della rete neurale creata.	101
Figura 5.4: Implementazione della rete neurale in Simulink, con dettagli dell'architettura e dello strato nascosto.	102
Figura 5.5: Schermata di Simulink al termine della simulazione. È visibile, nella finestra dello "scope", la stima del segnale interno (3 coordinate).	103
Figura 5.6: Posizione interna del tumore (p) in funzione del surrogato esterno (r) [60].	105
Figura 5.7: Schema a blocchi dell'implementazione del modello a variabili di stato in Matlab/Simulink.	107
Figura 5.8: Modello a variabili di stato realizzato in Simulink. Sono evidenziati e numerati i principali blocchi funzionali.	109
Figura 5.9: Blocco funzionale n°1.	109
Figura 5.10: Blocco funzionale n°2.	110
Figura 5.11: Interfaccia grafica per l'utente (GUI).	111
Figura 5.12: Finestra per la selezione del trattamento da importare.	112
Figura 5.13: Modello in Simulink completo. Sono evidenziati i blocchi relativi alla rete neurale e quelli relativi al modello di stato.	113
Figura 5.14: La GUI al termine dell'analisi dei risultati. A sinistra si notano i grafici dei segnali differenza d_{nn} (sopra) e d_{sm} (sotto), a destra i valori medi, le deviazioni standard e i valori massimi dei medesimi segnali.	114
Figura 5.15: Schema a blocchi dell'algoritmo della prima versione del software di correlazione.	119
Figura 5.16: Procedura per l'identificazione di una coppia ingresso/uscita.	121
Figura 5.17: Algoritmo di addestramento di una singola rete neurale.	124
Figura 5.18: In blu è riportata la stima della coordinata x del bersaglio eseguita dal software. I punti rossi indicano le effettive coordinate x del bersaglio in corrispondenza delle acquisizioni radiografiche. I campioni di imaging a sinistra della linea verde sono quelli utilizzati per l'addestramento.	126
Figura 5.19: Diagramma di flusso del ciclo di simulazione.	127

Figura 5.20: Diagramma di flusso della valutazione dell'errore di stima.	130
Figura 5.21: Confronto tra l'errore di stima, espresso come errore di puntamento 3D, commesso dalle reti neurali. Il tracciato rosso si riferisce alle stime eseguite con reti che non vengono mai riaddestrate, il tracciato verde, invece, si riferisce a reti che possono essere riaddestrate nel corso della simulazione.	131
Figura 5.22: Schema del sistema impiegato per la stima online del segnale interno.	133
Figura 5.23: Diagramma di flusso dell' algoritmo del software di correlazione.	134
Figura 5.24: In blu è mostrata la stima della coordinata x del segnale interno, mentre i punti rossi mostrano l'effettiva coordinata x del quarto marcatore negli istanti in cui viene registrata.	137
Figura 5.25: Immagine del fantoccio in cui sono evidenziati i componenti descritti in seguito.	138
Figura 5.26: Disegno schematico della puleggia a sezione ellittica variabile.	140
Figura 5.27: Set-up sperimentale: sono evidenziati la puleggia a sezione variabile, il tirante e il pistone.	141
Figura 5.28: Schematizzazione dei movimenti eseguiti dal fantoccio per la simulazione della fase inspiratoria ed espiratoria.	142
Figura 5.29: Isteresi della traiettoria della "lesione interna" ottenuta a seguito di una acquisizione eseguita alimentando il motore a differenti tensioni [65]. Unità di misura in [mm].	142
Figura 6.1: Schema della procedura utilizzata nella prima fase di validazione dell' algoritmo sviluppato .	145
Figura 6.2: Errore di puntamento (TE) rilevato in corrispondenza di ogni immagine radiografica di controllo acquisita nel corso del trattamento.	146
Figura 6.3: Distribuzione di probabilità cumulativa del TE per i casi control. In blu la curva relativa al modeler, in rosso la curva relativa all' algoritmo di correlazione sviluppato (reti neurali).	147
Figura 6.4: Distribuzione di probabilità cumulativa del TE per i casi worst. In blu la curva relativa al modeler, in rosso la curva relativa all' algoritmo di correlazione sviluppato (reti neurali).	148
Figura 6.5: Estratto della simulazione di un trattamento "control". In blu la stima della coordinata x (CC) del segnale interno eseguita dalle reti neurali, in verde quella eseguita dal modeler. I punti rossi evidenziano l'effettiva coordinata x del bersaglio rilevata tramite immagini radiografiche.	149
Figura 6.6: Estratto della simulazione di un trattamento "worst". La traccia blu indica la stima della coordinata x (CC) del segnale interno eseguita dalle reti neurali, la traccia verde indica quella eseguita dal modeler. I punti rossi evidenziano l'effettiva coordinata x del bersaglio rilevata tramite immagini radiografiche.	149
Figura 6.7: Set-up sperimentale utilizzato per le acquisizioni su fantoccio.	150
Figura 6.8: Andamento della coordinata x (AP) del marcatore numero 3.	151
Figura 6.9: Traiettoria tridimensionale del bersaglio. L'asse X coincide con la direzione antero-posteriore (AP), l'asse Y con la direzione cranio-caudale (CC) e l'asse Z con la direzione latero-laterale (LL).	152
Figura 6.10: Schema a blocchi delle attività svolte per l'analisi offline delle acquisizioni.	153
Figura 6.11: Coordinata x (AP). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di imaging acquisiti. L'ellisse nera tratteggiata evidenzia i frame persi a seguito del primo riaddestramento.	155
Figura 6.12: Coordinata y (CC). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di imaging acquisiti.	156
Figura 6.13: Coordinata z (LL). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di imaging acquisiti.	156
Figura 6.14: errore di puntamento 3D in corrispondenza di ogni frame, espresso come percentuale del massimo range di movimento 3D del bersaglio.	157
Figura 6.15: Schema a blocchi delle attività svolte per l'esecuzione e la valutazione delle analisi online.	158
Figura 6.16: Coordinata x (AP). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di imaging acquisiti.	160
Figura 6.17: Coordinata y (CC). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di imaging acquisiti.	160
Figura 6.18: Coordinata z (LL). In blu la stima eseguita dalle reti neurali, il rosso l'effettiva posizione del bersaglio e in verde i punti di imaging acquisiti.	161
Figura 6.19: errore di puntamento 3D in corrispondenza di ogni frame, espresso in percentuale del massimo range di movimento 3D del bersaglio.	161
Figura 7.1: Coordinata x (CC) di un marcatore esterno in un trattamento analizzato eseguito da CyberKnife.	165
Figura 7.2: Fantoccio per simulazione del movimento respiratorio. Sono evidenziati i marcatori che simulano il segnale esterno (in rosso) e quello che simula il bersaglio (in verde).	168
Figura 7.3: Pannello A: stima della coordinata z (AP) del bersaglio con rete a singolo strato (f.d.t lineare).	170
Figura 7.4: Pannello A: dettaglio della stima della coordinata z (AP) del bersaglio con rete a singolo strato (f.d.t lineare). Pannello B: dettaglio della stima della coordinata z (AP) del bersaglio con rete a due strati (f.d.t tanh).	171

Indice delle tabelle

<i>Tabella 1.1: Valori medi del segnale differenza ottenuti per la rete neurale e per il modello di stato.</i>	8
<i>Tabella 1.2: Valori medi del targeting error per modeler e per le reti neutrali.</i>	9
<i>Tabella 4.1: Database relativo ai trattamenti con Synchrony presso il Georgetown University Medical Center.</i>	93
<i>Tabella 5.1: Riassunto dei dati ricavati dalle informazioni sul trattamento.</i>	100
<i>Tabella 5.2: Variabili contenenti pesi e bias della rete addestrata.</i>	102
<i>Tabella 5.3: Dati relativi alla performance della rete neurale.</i>	115
<i>Tabella 5.4: Dati relativi alla performance del modello di stato.</i>	115
<i>Tabella 5.5: Possibili architetture delle reti neurali in funzione delle costanti definite dall'utente.</i>	122
<i>Tabella 5.6: Confronto tra i valori minimi e massimi dei range di movimento del bersaglio osservati in 20 frazioni di trattamento con CyberKnife e in 7 simulazioni su fantoccio.</i>	143
<i>Tabella 6.1: Caratteristiche delle reti neurali implementate nell'algoritmo di correlazione.</i>	145
<i>Tabella 6.2: Valori medi del targeting error per modeler e per le reti neutrali.</i>	147
<i>Tabella 6.3: p-value relativi al confronto tra le distribuzioni dell'errore di puntamento 3D osservate per le stime eseguite dalle reti neurali e dal modeler, sia per il subset control che per il subset worst.</i>	148
<i>Tabella 6.4: Dati relativi alle analisi effettuate su quattro acquisizioni offline.</i>	154
<i>Tabella 6.5: Principali caratteristiche hardware della workstation e delle telecamere del sistema SMART-D [64].</i>	157
<i>Tabella 6.6: Principali caratteristiche hardware della personal computer utilizzato.</i>	159
<i>Tabella 6.7: Dati relativi alle analisi effettuate su tre acquisizioni online.</i>	159

Ringraziamenti

Al termine di questo lavoro, desidero ringraziare il Professor Guido Baroni per avermi dato la possibilità di svolgere questa tesi e per la fiducia ed il sostegno dimostratami, non solo ora, ma già in occasione della Laurea Triennale e del Tirocinio.

Un grazie particolare va ad Andrea Pella che ha seguito il mio lavoro di un anno con enorme pazienza, assistendomi con grande competenza e disponibilità. Ringrazio Andrea anche per la progettazione iniziale e la realizzazione del fantoccio “Aigor”, che abbiamo poi completato per la mia tesi.

Ringrazio sinceramente Marco Riboldi per i preziosi consigli con i quali ha costantemente indirizzato il mio lavoro e contribuito al suo sviluppo.

Desidero ringraziare anche tutto lo staff del CART lab e dell’Unità di Bioingegneria del Centro Nazionale di Adroterapia Oncologica, con i quali ho condiviso numerose giornate di lavoro.

In ultimo, un grazie di cuore va ai miei genitori e ai miei amici che, particolarmente in quest’ultimo anno, mi hanno supportato e sopportato.