POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Biomedica



Tesi di Laurea Magistrale

Generazione di Synthetic CT mediante Deep Learning: Validazione di un modello di Transfer Learning per i Distretti Pelvico e Addominale

Relatore:
Prof. Filippo Molinari

Candidato:
Carlo Crovace

Correlatori:
Prof. Massimo Salvi

Ing. Stefania Zara

Anno Accademico 2025–2026

Indice

Elenco delle figure ..... **iii**
Elenco delle tabelle ..... **iv**
Abstract ..... **v**
1 Introduzione **1**
2 Radioterapia nel trattamento del cancro **2**
2.1 Radiobiologia 2
2.2 Evoluzione tecnologica 3
2.3 Workflow 5
2.3.1 Workflow convenzionale 6
2.3.2 Radioterapia adattativa 7
2.3.3 MRI-Only 9
3 Deep Learning in radioterapia **13**
3.1 Synthetic CT 15
3.1.1 Networks 17
3.1.2 Generatore e discriminatore 20
3.1.3 Architetture GAN e funzioni di Loss 22
3.2 Transfer Learning 24
3.2.1 Transfer Learning per le GAN 27
3.2.2 Strategie di Transfer Learning 29
4 Obiettivo della tesi **31**
5 Materiali e Metodi **32**
5.1 Dataset 32
5.2 Registrazione RM-CT 34
5.3 Preprocessing 35
5.4 Architettura e configurazione GAN 37
5.4.1 Generatore 38
5.4.2 Discriminatore 40
5.5 Metriche 41
5.6 Data cleaning 42
5.7 Divisione in Fold 43
6 Risultati **45**
6.1 Studio d'ablazione 45

6.2 Conversione a 2.5D 48
6.3 Data augmentation 50
7 Discussione dei risultati **52**
8 Conclusione **55**
9 Sviluppi futuri **56**
Appendice A **58**
Appendice B **68**
Appendice C **74**
Bibliografia **80**

Elenco delle figure

Fig. 1 MR-Linac: una macchina che combina strumentazione per radioterapia con uno scanner per MRI. 5
Fig. 2 Sottoclassi dell'Intelligenza Artificiale. 13
Fig. 3 Gerarchia delle architetture deep learning. 17
Fig. 4 Illustrazione del modello generator-only. 18
Fig. 5 Architettura GAN. 19
Fig. 6 (a): ResNet; (b): U-Net. 21
Fig. 7 PatchGAN. 22
Fig. 8 (a): sigmoid cross entropy loss function. (b): least squares loss function. 24
Fig. 9 Risultati dello studio citato. 26
Fig. 10 Rappresentazione della strategia di apprendimento per trasferimento tramite (a) congelamento dei layer di estrazione delle feature e (b) fine tuning. 29
Fig. 11 A sinistra la segmentazione automatica della ROI esterna su RM, a destra su TC. 34
Fig. 12 Visualizzazione delle segmentazioni degli OAR su piano sagittale (paziente 1). 35
Fig. 13 Istogramma (PDF, probability density function) utilizzato per l'histogram matching. .. 36
Fig. 14 Schema dell'architettura del generatore. 40
Fig. 15 Schema dell'architettura del discriminatore. 41
Fig. 16 Esempio di slice scartata per via delle deformazioni alle estremità del volume. 42
Fig. 17 Risultati del k-validation. 48
Fig. 18 Patch estraibili durante il data augmentation. 50
Fig. 19 Risultati della generazione di sCT per il paziente 2. 54

Elenco delle tabelle

Tabella 1 Metriche per la valutazione quantitativa delle sCT. 12
Tabella 2 Confronto tra le caratteristiche del dataset sorgente e del dataset target. 32
Tabella 3 Informazioni ricavate dai TAG DICOM delle acquisizioni presenti nel dataset target. 33
Tabella 4 Parametri di addestramento. 38
Tabella 5 Risultati di prova 1, prova 2 e prova 3. 43
Tabella 6 Divisione in 5 fold. 44
Tabella 7 Valutazione crollo prestazionale del modello sorgente passando da input RM 3T a RM 1.5T. 45
Tabella 8 Risultati delle prove 1 e 2 sul primo fold. 45
Tabella 9 Risultati delle prove 3 e 4 sul primo fold. 45
Tabella 10 Risultati della prova 5 sul primo fold. 46
Tabella 11 Risultati delle prove 6, 7 e 8 sul primo fold. 46
Tabella 12 Risultati delle prove 9, 10, 11 e 12 sul primo fold. 47
Tabella 13 Risultati della prova 13 sul primo fold. 47
Tabella 14 Risultati delle prove 14 e 15 sul primo fold. 47
Tabella 15 Risultati della prova 16 sul primo fold. 47
Tabella 16 Risultati delle prove per la selezione della miglior strategia per la conversione della rete a 2.5D. 49
Tabella 17 Risultati del k-validation dopo conversione a rete 2.5D. 49
Tabella 18 Confronto in termini relativi fra rete 2D e 2.5D per ogni fold. 49
Tabella 19 Confronto in termini relativi fra rete 2D e 2.5D combinata al data augmentation per ogni fold. 51
Tabella 20 Risultati del k-validation dopo conversione a rete 2.5D combinata al data augmentation. 51

Abstract

Lo stato dell'arte della radioterapia (RT) oncologica è sempre più orientato verso workflow basati sulla sola acquisizione della Risonanza Magnetica (RM), al fine di eliminare l'esposizione a radiazioni ionizzanti in fase di imaging e sfruttare il maggior contrasto dei tessuti molli rispetto alla tradizionale Tomografia Computerizzata (TC). Tuttavia, per essere utilizzabile nel calcolo dosimetrico, questo approccio richiede la generazione di synthetic CT (sCT) che compensino la mancanza di informazioni dirette sulla densità elettronica (DE), intrinsecamente codificate nelle Unità Hounsfield (HU) delle TC. Tra le diverse soluzioni proposte in letteratura, le tecniche di Deep Learning, e in particolare le Reti Generative Avversarie (GAN), si sono affermate come gli strumenti più efficaci per questo compito. L'obiettivo di questa tesi è l'adattamento di un modello Pix2Pix per il distretto addominale e la pelvi maschile, originariamente addestrato su RM a 3T, ad un nuovo dominio clinico basato su RM a 1.5T. Per affrontare il problema disponendo di una coorte limitata di pazienti, è stato applicato il Transfer Learning (TL), una tecnica che permette di trasferire la conoscenza già acquisita dal modello, riducendo i tempi computazionali e migliorando la convergenza della rete.

Dal punto di vista metodologico, il lavoro ha previsto la registrazione deformabile tra TC e RM di ogni paziente, supportata dalla segmentazione degli organi, un *preprocessing* mirato a standardizzare gli input e l'utilizzo della *k-fold cross-validation* per garantire robustezza statistica in fase di validazione. Sono state esplorate varie configurazioni di TL, mettendo a confronto il *Fine-Tuning* (FT) con il congelamento dei pesi, e il trasferimento dell'intera GAN con il trasferimento asimmetrico. In seguito, per migliorare la continuità spaziale dei risultati, il TL è stato sfruttato per effettuare la conversione verso un'architettura 2.5D, affiancando l'applicazione della *data augmentation*. Infine, in fase di inferenza, si è adottata una strategia di *bagging* sui modelli addestrati durante la *cross-validation*, aggregando i rispettivi output tramite la mediana.

I risultati mostrano che il modello finale ha colmato il gap prestazionale rispetto al modello sorgente e, nonostante l'inferiore qualità nativa delle immagini RM a 1.5T, ha addirittura superato le performance del modello sorgente. L'algoritmo ha registrato le seguenti metriche: MAE pari a 37,8 HU, RMSE pari a 97,4 HU, SSIM pari a 0,900 e PSNR pari a 30,1 dB, riportando infine un tempo di generazione inferiore ai 3 minuti.

In conclusione, la strategia implementata produce sCT di ottima qualità in tempi clinicamente sostenibili, aprendo alla possibilità di ulteriori adattamenti o estensioni della rete anche in presenza di coorti ridotte di pazienti, con tempi di sviluppo ottimizzati.

1 Introduzione

Attualmente il cancro rappresenta una delle principali cause di mortalità a livello globale; l'International Agency for Research on Cancer (IARC) ha stimato che annualmente si verificano circa 12,7 milioni di nuovi casi e 7,6 milioni di decessi legati a questa patologia. Per fronteggiare questa problematica, negli ultimi anni sono stati raggiunti importanti progressi nelle terapie oncologiche. Tra le diverse opzioni terapeutiche disponibili, la RT rimane tra le più utilizzate [1].

L'obiettivo della RT è fornire un trattamento localizzato, cercando di massimizzare il controllo della malattia tumorale e, al contempo, riducendo al minimo i danni ai tessuti sani circostanti. Per raggiungere questo risultato, le tecniche di somministrazione delle radiazioni hanno subito una profonda evoluzione tecnologica, passando dalla RT conformazionale tridimensionale (3D-CRT), all'impiego di approcci più sofisticati come la RT ad intensità modulata (IMRT) fino ad arrivare all'integrazione della RT guidata dalle immagini (IGRT) [2].

Tradizionalmente, la TC rappresenta lo standard per la simulazione e la pianificazione dei trattamenti radioterapici. Tuttavia, le nuove tecnologie emergenti, in particolare la Magnetic Resonance guided RT (MRgRT), introducono l'uso della risonanza magnetica (RM) nel workflow clinico. L'impiego della RM offre un notevole vantaggio in questo ambito, in quanto garantisce maggior contrasto dei tessuti molli che si traduce nella possibilità di minimizzare ulteriormente la dose assorbita dagli Organi a Rischio (OAR) [3], eliminando inoltre la dose proveniente dalle TC di simulazione.

Nonostante i benefici discussi, il trattamento basato esclusivamente su RM presenta una criticità rilevante che riguarda la mancanza dell'informazione sulla DE dei tessuti, necessaria per poter eseguire i calcoli della dose. Normalmente, i valori DE vengono ricavati direttamente dalle HU della TC. Vista la mancanza di relazione diretta tra il segnale di RM e la DE, il metodo convenzionale non può essere applicato sull'imaging a RM (MRI). Per affrontare questo ostacolo si ricorre alla generazione di sCT, modelli digitali predittivi capaci di mappare accuratamente le intensità di segnale MR di una data anatomia nei corrispondenti valori HU. Nel corso del tempo sono state implementate diverse strategie computazionali per ottenere questa traslazione di dominio [4].

A delineare lo stato dell'arte di tali strategie, un recente sondaggio ha evidenziato come il 52% degli sviluppatori di software prediliga l'impiego di acquisizioni RM pesate in T2 come input primario per i propri algoritmi. Per quanto riguarda il campo magnetico, la metà dei ricercatori si è affidata a tomografi RM a 3 T, il 30% si è orientato su sistemi da 1,5 T e il restante su macchine a 0,35 T. Tali variazioni riflettono la grande adattabilità degli algoritmi per la generazione di sCT a diversi contesti di imaging. Un altro dato emerso dal medesimo studio è che solo un quarto degli sviluppatori dichiara di avvalersi di immagini accoppiate per addestrare i propri modelli. Risulta anche evidente una chiara preferenza per le GAN, con l'80% degli sviluppatori che utilizza questa architettura [5].

2 Radioterapia nel trattamento del cancro

L’insorgenza delle neoplasie maligne trae origine dall’alterazione delle cellule sane, innescata dall’esposizione a diversi fattori di rischio. Questi comprendono sia agenti ambientali esterni, come le radiazioni ionizzanti e i metalli pesanti tossici, sia agenti tossici endogeni, tra cui i radicali liberi. Il progressivo accumulo di danni a carico dell’acido desossiribonucleico nucleare (DNA) determina una condizione di instabilità genomica, caratteristica distintiva di varie patologie tumorali. D’altra parte, questa vulnerabilità è alla base di alcuni trattamenti, quali la chemioterapia e la RT, le quali sono comunque associate all’insorgenza di effetti collaterali [6].

Fisiologicamente, le cellule dei mammiferi attivano complessi meccanismi di riparazione del DNA per preservare la stabilità e l’integrità genetica quando esposte ad agenti endogeni o esogeni. Un’anomalia o una disregolazione di questi meccanismi viene associata all’insorgenza e alla progressione del cancro. Poiché le principali terapie antitumorali, come le radiazioni ionizzanti e gli agenti chemioterapici, mirano a indurre l’apoptosi provocando, in modo diretto o indiretto, un danno a livello del DNA, eventuali alterazioni nella risposta a tale danno possono determinare scenari clinici opposti passando dall’ipersensibilità ai trattamenti allo sviluppo di resistenza da parte delle cellule tumorali [7].

Tra le varie modalità di trattamento che comprendono RT, asportazione chirurgica, chemioterapia, immunoterapia e terapia ormonale, la RT rimane una componente predominante essendo impiegata in circa il 50% dei pazienti oncologici. L’obiettivo della RT è quello di arrestare il potenziale proliferativo e moltiplicativo delle cellule maligne. L'impiego di radiazioni ionizzanti permette infatti di generare ioni e depositare energia all'interno dei tessuti attraversati; questa energia depositata induce la morte delle cellule cancerose per via diretta, oppure indirettamente attraverso l'innesco di cambiamenti genetici [1].

2.1 Radiobiologia

L'effetto più critico indotto dalle radiazioni a livello cellulare è la rottura a doppio filamento della molecola di DNA. Tale lesione determina la perdita irreversibile della capacità riproduttiva della cellula, conducendola alla conseguente morte. Il danno indotto dalle radiazioni può verificarsi attraverso una ionizzazione diretta o, come si osserva più frequentemente nella pratica clinica, attraverso un'azione ionizzante indiretta. Quest'ultima è mediata dai radicali liberi, i quali si generano a seguito del processo di radiolisi dell'acqua presente all'interno della cellula. Le radiazioni ionizzanti sono inoltre in grado di alterare svariati processi biologici, tra cui i processi del ciclo cellulare, la senescenza cellulare e le vie apoptotiche [2].

Il successo di un trattamento radioterapico, in termini di eradicazione tumorale, è strettamente legato a due fattori contrapposti: da un lato la radiosensibilità intrinseca del tumore, dall'altro la tolleranza al danno dei tessuti sani adiacenti. La dose letale tumorale è definita come la dose di radiazioni che produce una regressione completa e permanente del tumore in vivo nella zona irradiata. Il principio del frazionamento della dose, che prevede la somministrazione della dose totale in quote minori (frazioni) all'interno di un regime multi-frazionato convenzionale, trova il suo fondamento razionale nelle cosiddette "4R" della radiobiologia: Riparazione del danno subletale, Ripopolamento, Ridistribuzione e Riossigenazione [2].

La riparazione del danno è considerata la giustificazione biologica principale per l'adozione del frazionamento. Razionare la dose nel tempo permette infatti ai tessuti sani di recuperare tra una seduta e l'altra, limitando l'entità del danno subito. La ridistribuzione delle cellule proliferanti all'interno del ciclo cellulare è un altro elemento chiave e si basa sul fatto che le cellule che si trovano in una fase radioresistente possono progredire verso una fase maggiormente radiosensibile; questo fenomeno incrementa l'efficacia del trattamento frazionato rispetto a un'ipotetica irradiazione in singola seduta. Se l'intervallo è superiore a sei ore, le cellule si ripopolano, con conseguente aumento della frazione di cellule sopravvissute, fenomeno noto come ripopolamento. Infine, le cellule localizzate al centro del tumore sono tipicamente ipossiche e resistenti alle radiazioni. Queste cellule, durante un ciclo di trattamento frazionato, vanno incontro a riossigenazione, diventando progressivamente più radiosensibili alle dosi successive di radiazione [2].

È inoltre fondamentale considerare che la risposta dei tumori alle radiazioni nel tempo presenta un'enorme variabilità da un tipo di tumore all'altro, ma anche all'interno dello stesso tipo tumorale. I tessuti normali possono essere classificati in base alla loro cinetica di risposta. I tessuti a "risposta precoce" presentano un'organizzazione gerarchica, un'elevata attività proliferativa e un rinnovamento cellulare costante. Al contrario, i tessuti a "risposta tardiva" possiedono una struttura più flessibile, una bassa velocità di proliferazione e non sono caratterizzati da un rinnovamento cellulare rapido e continuo. Ne consegue che le peculiarità cinetiche dei tessuti sani e delle neoplasie, le loro relative risposte radiobiologiche e quindi le variazioni di sensibilità sono elementi determinanti nello sviluppo e nell'ottimizzazione dei protocolli di frazionamento. In definitiva, il miglior esito terapeutico si raggiunge somministrando la massima dose tollerabile dai tessuti a risposta tardiva nel minor tempo possibile per mantenere gli effetti acuti entro limiti clinici accettabili [8].

2.2 Evoluzione tecnologica

In base alla localizzazione della sorgente radiogena, la RT può essere classificata in due categorie principali: la RT a fasci esterni (EBRT, *External Beam Radiation Therapy*) e la brachiterapia (una forma di RT che consiste nel posizionamento di sorgenti radioattive direttamente all'interno o in prossimità del tumore). L'EBRT rappresenta l'approccio radioterapico più comune e risulta particolarmente indicata nel trattamento di pazienti affetti da patologie a rischio intermedio o elevato [9], [10].

Dalla metà degli anni Novanta, la ricerca nel campo della RT si è fortemente concentrata sullo sviluppo, uso e valutazione della tecnologia di RT conformazionale, con particolare interesse alle tecniche di RT conformazionale controllata da computer (CCRT, *Computer-Controlled Conformal Radiation Therapy*). L'obiettivo della terapia conformazionale è quello di concentrare la regione ad alta dose, nelle tre dimensioni, in modo che si adatti strettamente alla geometria del volume bersaglio, minimizzando l'irradiazione ai tessuti sani circostanti. Per erogare trattamenti altamente conformi, la CCRT si avvale di avanzati sistemi computerizzati di pianificazione 3D, di collimatori multi-lamellari (MLC, *Multileaf Collimators*) e di acceleratori lineari controllati da computer [11].

L'evoluzione di questo approccio si è avuta con la RT a intensità modulata (IMRT), una tecnica in cui l'intensità del fascio radiante viene modulata dinamicamente in funzione della disposizione spaziale del bersaglio e dei tessuti adiacenti. Nello specifico, l'intensità della radiazione viene

ridotta in corrispondenza delle aree che attraversano gli OAR, mentre viene incrementata lungo le direzioni che attraversano principalmente il volume bersaglio. Il vantaggio principale dell'IMRT consiste nella capacità di adattare la distribuzione della dose anche a bersagli concavi. Con questa tecnica, l'unità minima di irradiazione che può essere controllata individualmente prende il nome di *beamlet*; l'insieme delle intensità assegnate ai vari beamlet all'interno del campo di irradiazione definisce la mappa di intensità. Nei trattamenti a campi multipli il numero di beamlet a cui assegnare un'intensità diventa elevato, al punto da necessitare un'ottimizzazione computerizzata. Il processo prevede che il pianificatore imposti a priori gli obiettivi di dose per il tumore e i limiti per gli OAR; successivamente, un algoritmo determina le intensità ottimali dei fasci per soddisfare tali requisiti, un processo noto come "pianificazione inversa" (*inverse planning*) [12].

Più recentemente, al fine di superare i limiti intrinseci dell'IMRT a campo fisso, l'attenzione clinica si è spostata verso le terapie rotazionali o ad arco. Il principio alla base di queste metodiche prevede l'erogazione della radiazione attraverso una rotazione continua della sorgente attorno al paziente, coprendo un angolo di 360°. Le terapie ad arco garantiscono distribuzioni di dose altamente conformi e costituiscono di fatto un'estensione dell'IMRT tradizionale. Rispetto all'IMRT con gantry fisso, un vantaggio rilevante è la superiore efficienza di erogazione del trattamento: grazie alla riduzione dei tempi di erogazione e del numero di unità monitor (MU) impiegate comporta una minore esposizione alle radiazioni al resto del corpo. La riduzione dei tempi di erogazione del macchinario offre l'ulteriore opportunità di integrare più agevolmente la RT guidata dalle immagini (IGRT) all'interno dello slot temporale della singola seduta. L'IGRT si basa sull'acquisizione di immagini morfologiche prima e/o durante il trattamento, consentendo una verifica più accurata del posizionamento del paziente e l'implementazione di strategie adattative. Sebbene l'IGRT convenzionale possa aumentare il tempo di permanenza sul lettino e la dose radiogena totale a cui è esposto il paziente, la terapia ad arco, grazie alla sua rapidità di erogazione, riesce ampiamente a compensare questi svantaggi. Le due principali declinazioni delle terapie ad arco sono la tomoterapia e la RT ad arco a modulazione volumetrica (VMAT). Le macchine per la tomoterapia combinano le funzionalità di uno scanner TC e di un acceleratore lineare, erogando un fascio a ventaglio (*fan-beam*) da una sorgente in continua rotazione mentre il paziente trasla longitudinalmente seguendo una modalità assiale o elicoidale. La VMAT, introdotta nel 2007, è stata descritta come una sofisticata tecnica di RT che permette di variare simultaneamente tre parametri durante l'erogazione del trattamento: la velocità di rotazione del gantry, la forma dell'apertura di trattamento tramite il movimento delle lamelle del MLC e la dose rate (intensità della dose).

L'innovazione tecnologica più recente si è avuta con lo sviluppo del sistema MR-linac, un dispositivo che integra un acceleratore lineare con uno scanner per RM (Fig. 1). L'obiettivo di questa architettura ibrida è permettere l'acquisizione di immagini RM, caratterizzate da alto contrasto dei tessuti molli, direttamente dal lettino del trattamento, rendendo possibile la visualizzazione di tutte le variazioni anatomiche durante il corso della RT. La RM integrata consente di monitorare accuratamente sia il movimento inter-frazione (tra una seduta e l'altra) sia quello intra-frazione (durante l'erogazione). Le informazioni prodotte dalla RM sono utilizzate come input per l'implementazione della RT adattativa (ART). Nello specifico, la RM giornaliera funge da guida per la verifica della posizione dei tessuti molli o per la ripianificazione giornaliera, mentre l'acquisizione di immagini a fascio attivo (*beam-on*) supporta invece il calcolo dell'accumulo di dose risolto nel tempo o l'adattamento del piano in tempo reale (*real-time replanning*). Infine, il monitoraggio mediante RM anatomica e funzionale ripetuta si rivela un prezioso strumento per la valutazione della risposta tumorale alla terapia [13].

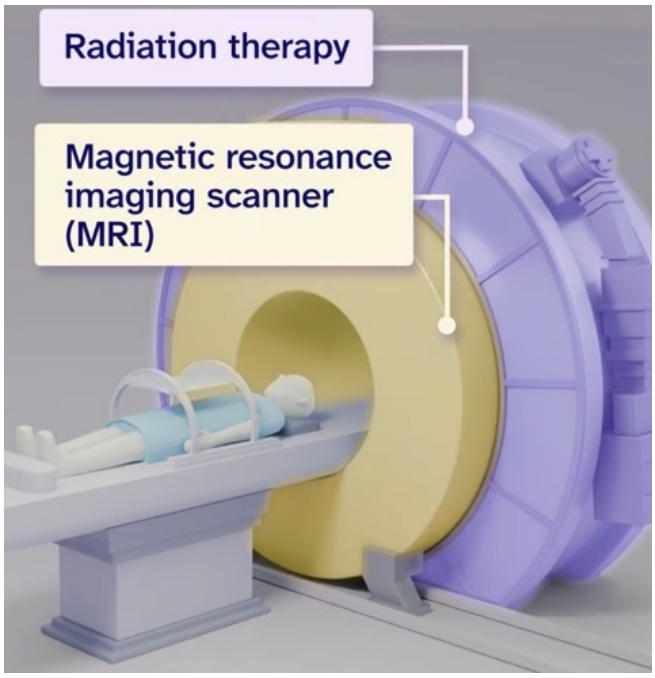


Fig. 1 MR-Linac: una macchina che combina strumentazione per radioterapia con uno scanner per MRI [14].

2.3 Workflow

L'incremento della richiesta di trattamenti oncologici, unito alla crescente personalizzazione delle terapie oncologiche, rende complessa l'erogazione tempestiva delle cure. Tempi di attesa eccessivamente prolungati comportano esiti clinici negativi, come un maggior rischio di recidive locali, un'accelerata progressione neoplastica e un prolungato stress psicologico per i pazienti. Inoltre, non è da trascurare il monitoraggio del carico di lavoro del personale sanitario, considerando che questo si riflette sulla soddisfazione e sul benessere professionale dei medici. Dal punto di vista operativo, il trattamento radioterapico inizia con l'indirizzamento del paziente al dipartimento di RT e prosegue con la successiva visita specialistica con un radio-oncologo, il quale definisce le procedure diagnostiche e preparatorie necessarie, note come "workflow pretrattamento". Questa fase include l'acquisizione delle immagini (TC, RM, PET-TC), la delineazione del volume bersaglio e degli OAR e la pianificazione del trattamento. Al termine del consulto iniziale viene programmata la prima seduta di irradiazione [15].

Per standardizzare la delineazione dei volumi di interesse, si fa riferimento a linee guida come quella del Report 50 dell'International Commission on Radiation Units and Measurements (ICRU) [16], che stabilisce una definizione chiara e coincisa di quest'ultimi:

- Gross Tumor Volume (GTV): rappresenta l'estensione e la localizzazione macroscopica, palpabile o visibile/dimostrabile, della neoplasia maligna.
- Clinical Target Volume (CTV): è un volume di tessuto che contiene un GTV e/o una malattia maligna microscopica subclinica, che deve essere eliminata.
- Planning Target Volume (PTV): è un concetto geometrico, definito per selezionare le configurazioni appropriate di fasci, tenendo conto dell'effetto netto di tutte le possibili variazioni e imprecisioni geometriche al fine di garantire che la dose prescritta venga effettivamente assorbita nel CTV. Le sue dimensioni e la sua forma dipendono principalmente dal CTV, ma anche dalla tecnica di trattamento utilizzata.

- OAR: sono tessuti normali la cui radiosensibilità può influenzare significativamente la pianificazione del trattamento e/o la dose prescritta.

Inoltre, in funzione della tecnica terapeutica utilizzata, si possono identificare altri due volumi:

- Treated Volume: è il volume racchiuso da una superficie di isodose, selezionata e specificata dal radioterapista oncologo come appropriata per raggiungere lo scopo del trattamento (ad esempio, eradicazione del tumore, palliazione).
- Irradiated Volume: rappresenta il volume di tessuto che riceve una dose considerata significativa in relazione alla normale tolleranza del tessuto.

Un piano radioterapeutico ottimale deve mirare a un'irradiazione omogenea del PTV, salvaguardando al contempo gli OAR e il restante volume a rischio rimanente (VRR) dai danni indotti dalle radiazioni. Questo richiede la necessità di prescrivere una dose da somministrare al PTV, cercando simultaneamente di contenere al livello più basso possibile l'esposizione dei tessuti sani. Per affrontare questo bilanciamento dosimetrico si impiegano sistemi computerizzati di pianificazione del trattamento (TPS, *Treatment Planning Systems*), i quali simulano una serie di piani provvisori. Per ogni piano elaborato, il TPS genera delle curve di isodose sull'intero dominio di trattamento e gli istogrammi Dose-Volume (DVH), i quali quantificano la dose ricevuta dal PTV e da ciascun OAR contornato. Tipicamente, la selezione del piano definitivo si basa sull'analisi visiva dei due strumenti descritti. Tuttavia, è opinione diffusa che questo processo decisionale risulti spesso altamente soggettivo. Per ovviare a questo limite ed effettuare valutazioni più rigorose, è stato sviluppato l'Indice di Distribuzione della Dose (DDI), un parametro che aggrega diverse metriche parziali offrendo uno strumento oggettivo per il confronto dosimetrico dei piani preliminari [17].

Un ulteriore elemento che influenza la distribuzione della dose clinica è la presenza di disomogeneità tissutali all'interno del corpo del paziente. Già a partire dal 1974, i calcoli dosimetrici computerizzati per i fasci di fotoni tentavano di introdurre dei fattori correttivi per compensare le differenze di densità tessutali. Tuttavia, solo con la disponibilità della TC è stato possibile applicare questi metodi di correzione in modo accurato. Il ruolo centrale della TC è imputabile a due caratteristiche: in primo luogo, le scansioni TC forniscono un'esatta mappatura spaziale delle disomogeneità anatomiche; in secondo luogo, i valori numerici intrinseci della TC (HU) forniscono informazioni quantitative dirette sulle proprietà radiologiche, evidenziando la densità elettronica dei vari tessuti [18].

2.3.1 Workflow convenzionale

Nella pratica clinica, il processo ha inizio con l'acquisizione di una TC di simulazione (o di pianificazione), i cui dati vengono esportati verso il TPS nel formato DICOM (*Digital Imaging and Communications in Medicine*). In questa fase preparatoria, i medici radio-oncologi procedono alla delineazione dei volumi target e degli OAR direttamente sulle immagini TC, seguendo le linee guida standard. Successivamente, i fisici medici elaborano i piani di trattamento avvalendosi di tecniche di erogazione ad elevata conformazione spaziale, quali l'IMRT e la VMAT.

Una volta ultimato e approvato il piano, ha inizio la fase di trattamento vera e propria. Nella RT moderna, il corretto posizionamento del paziente (comunemente definito setup) riveste un prerequisito indispensabile per assicurare un'erogazione accurata della dose al volume bersaglio,

limitando l'esposizione dei tessuti sani circostanti. Per massimizzare la precisone del posizionamento, i protocolli clinici adottano le tecniche IGRT.

Tra le varie modalità di imaging possibili per l'IGRT, la TC a fascio conico (CBCT, *Cone-Beam Computed Tomography*) ha acquisito un ruolo di primo piano, grazie alla sua capacità di generare acquisizioni volumetriche 3D dell'anatomia del paziente mentre è nella posizione di trattamento. Queste acquisizioni, generate giornalmente, vengono sfruttate per individuare e correggere gli errori di setup attraverso il confronto con le immagini di riferimento della TC di pianificazione. L'implementazione di questo workflow richiede l'impiego di un acceleratore lineare equipaggiato con uno scanner CBCT integrato all'interno del gantry [19]. L'impiego di una scansione CBCT, in alternativa a una TC convenzionale, offre vantaggi significativi riconducibili principalmente a tempistiche di imaging più rapide e una marcata riduzione della dose radiogena complessiva somministrata al paziente [20].

L'acquisizione delle CBCT giornaliere viene effettuata prima di ogni singola frazione. Il set di immagini CBCT ottenuto viene co-registrato con la TC di pianificazione di riferimento: questo processo, chiamato *image-registration*, consiste nell'allineamento spaziale delle due acquisizioni al fine di rilevare e correggere le eventuali deviazioni di posizionamento del paziente prima di procedere con l'irradiazione [19].

2.3.2 Radioterapia adattativa

All'interno dei trattamenti radioterapici convenzionali, le tossicità e le complicanze a carico dei tessuti sani più frequentemente documentate interessano l'apparato urologico (ad esempio cistite, ematuria, stenosi uretrale e ureterale, disfunzione erettile) e il tratto gastrointestinale (come diarrea, proctite, stenosi rettale e anale). Allo stesso tempo, risulta importante preservare una copertura dosimetrica adeguata al target tumorale. L'efficacia di tale copertura è intrinsecamente vincolata dalla precisione con cui viene posizionato il paziente, dalla delineazione dei volumi bersaglio, dai movimenti fisiologici degli organi, dalla conformazione geometrica del fascio e dalla precisione del calcolo tridimensionale della dose [21].

Poiché le immagini TC offrono un limitato contrasto per i tessuti molli, si preferisce integrare l'acquisizione di RM, le quali vengono successivamente fuse con la TC di simulazione per ottimizzare la delineazione delle strutture target. L'implementazione della RM nel processo di pianificazione del trattamento, rispetto alla delineazione del bersaglio mediante la sola TC, ha dimostrato di poter ridurre il volume contornato per il CTV della prostata di una percentuale compresa tra il 32% e il 40%. Tale ottimizzazione porta benefici sul risparmio dosimetrico a carico di strutture critiche quali la vescica, il retto e il bulbo penieno. Inoltre, la RM assicura una discriminazione spaziale più accurata per l'apice prostatico, permettendo di ridurre l'irradiazione del diaframma genitourinario, struttura che ospita lo sfintere uretrale esterno. Poiché l'apice prostatico rappresenta una sede comunemente interessata dal carcinoma prostatico, la sua corretta identificazione è fondamentale per evitare un sotto-dosaggio che inalzerebbe il rischio di recidiva locale. Inoltre, l'imaging RM facilita l'identificazione dell'uretra membranosa, permettendo di inserirla tra le strutture da proteggere dalle radiazioni al fine di evitare l'insorgenza di stenosi indotte [22].

In questo scenario, l'introduzione dei sistemi integrati MR-linac sfrutta il miglior contrasto della RM per offrire la possibilità di tracciare le variazioni anatomiche del paziente durante il trattamento, consentendo l'adattamento dinamico del piano. Tuttavia, l'implementazione clinica

richiede una valutazione del trade-off tra le tempistiche necessarie alla ripianificazione e l'effettivo beneficio dosimetrico ottenuto. L'impiego dei dati RM nella ART richiede l'interpretazione dei volumi, operazione che solitamente viene eseguita propagando i contorni dalle immagini pretrattamento di riferimento alle scansioni dell'ultima sessione temporale mediante algoritmi di registrazione. Questo processo richiede protocolli specifici per il controllo qualità in base al sito anatomico considerato per garantire risultati realistici. Dal momento che ogni frazione di trattamento può richiedere un aggiornamento del piano, il team clinico deve selezionare quotidianamente la strategia adattativa più idonea [13].

La procedura consiste nell'acquisire una TC e una RM pretrattamento, avendo cura di garantire un posizionamento riproducibile. Su queste immagini di base, il medico radio-oncologo provvede alla delineazione del target e degli OAR, fondamentali per la generazione del piano di trattamento tramite TPS. Mentre, per ogni singola seduta terapeutica, viene acquisita una RM direttamente sul lettino del paziente. La TC, i contorni dei volumi e il piano terapeutico originario, congiuntamente alla nuova RM, costituiscono i nuovi dati di input del TPS per l'adattamento del piano per quella specifica frazione [13].

A livello temporale è possibile condurre l'ART con tre tempistiche differenti: offline, online e real-time. L'adattamento offline avviene nell'intervallo di tempo tra una frazione e l'altra; questo approccio offre agli operatori clinici il tempo necessario per eseguire una ri-pianificazione dettagliata utilizzando workflow e controlli di qualità (QA, *Quality Assurance*) standard. Tuttavia, l'approccio offline si dimostra meno efficace in distretti dove le variazioni anatomiche avvengono con una frequenza elevata, rischiando di applicare correzioni basate su un'anatomica già mutata nel tempo. Al contrario, l'adattamento online si svolge immediatamente prima dell'erogazione della frazione, con il paziente mantenuto sul lettino nella posizione di trattamento. Questa strategia richiede un flusso di lavoro rapido e altamente integrato, ma offre il vantaggio di rispondere tempestivamente all'anatomia del giorno, mitigando gli errori dovuti alle ampie variazioni inter-frazione. Infine, l'adattamento real-time spinge l'integrazione ancora oltre, utilizzando un imaging continuo dell'anatomia (o di surrogati, tra cui l'impianto e il tracciamento di marcatori radiopachi o il monitoraggio dell'espansione dell'addome durante la respirazione tramite telecamere ad infrarossi e marcatori appositi) per modificare il piano terapeutico, idealmente istante per istante, durante l'erogazione del fascio radiante. Pur rappresentando la forma più accurata di ART in assoluto, riuscendo a compensare in diretta i movimenti intra-frazione, il real-time richiede livelli di automazione estremamente elevati e protocolli di QA ben più complessi rispetto alle metodiche offline e online [23], [24].

L'adattamento del piano terapeutico può seguire due approcci: il metodo "Adapt-to-Position" (ATP) e il metodo "Adapt-to-Shape" (ATS). Nella strategia ATP, la TC pretrattamento viene allineata spazialmente alla RM giornaliera mediante un algoritmo di registrazione rigida; sulla base di tale allineamento, la posizione dell'isocentro nei dati di riferimento viene aggiornata. Il ricalcolo o la re-ottimizzazione della distribuzione della dose viene quindi effettuato basandosi sulla geometria della TC di pianificazione e sui relativi contorni. Di conseguenza, questo approccio non prevede (né rende possibile) una modifica delle delineazioni dei volumi per il nuovo piano adattato [13].

Al contrario, la strategia ATS, consente l'ottimizzazione del piano in base all'anatomia che il paziente presenta ad ogni seduta. Anche nel workflow ATS, l'operazione iniziale consiste nella registrazione della TC pretrattamento e della RM online. Successivamente i contorni stabiliti durante la pianificazione vengono propagati in modo automatico sulla RM giornaliera mediante algoritmi di registrazione deformabile, tali contorni vengono verificati ed eventualmente corretti

dal radio-oncologo. Sebbene l'approccio ATP garantisca tempi di esecuzione decisamente più rapidi rispetto al metodo ATS, esso mostra limiti evidenti nella copertura del target, anche in presenza di minime variazioni anatomiche dei tessuti [13].

Il processo di registrazione delle immagini costituisce l'elemento cardine della ART e può essere di tipo rigido o deformabile. La registrazione rigida multimodale è un'operazione geometrica mirata ad allineare due differenti set di immagini per integrarne le informazioni complementari. Essa si basa sull'applicazione di trasformazioni spaziali (quali traslazione, rotazione, scalatura, campionamento, ecc.) applicate a un'immagine target per allinearla a un'immagine di riferimento [25].

L'approccio più avanzato della registrazione deformabile dell'immagine (DIR, *Deformable Image Registration*) prevede la modellazione delle variazioni locali dei tessuti. Esistono diversi tipi di algoritmi per questo fine, segue la presentazione delle due possibilità fornite dal software RayStation [26], utilizzato in questa tesi:

- Hybrid DIR: combina informazioni dell'immagine, come l'intensità, con informazioni anatomiche fornite da set di immagini contornate. Ciò consente di mantenere l'integrità delle strutture anatomiche, anche in aree in cui le informazioni dell'immagine sono meno affidabili. L'algoritmo ANACONDA di RayStation, ad esempio, si basa su una formulazione matematica in cui il problema di registrazione è formulato come un problema di ottimizzazione non lineare. La funzione obiettivo è una somma ponderata di tre termini non lineari:
  1. Termine di similarità dell'immagine, misurato tramite il coefficiente di correlazione.
  2. Termine di regolarizzazione della griglia, progettato per mantenere la griglia dell'immagine deformata liscia e invertibile.
  3. Termine di penalità, aggiunto al problema di ottimizzazione quando vengono utilizzate strutture di controllo, al fine di deformare la struttura selezionata dell'immagine di riferimento nella struttura corrispondente dell'immagine target.
- Biomechanical DIR: la deformazione è guidata dalle condizioni al contorno calcolate a partire dalle regioni di interesse (ROI, *Region of Interest*), definite sia nell'immagine di riferimento che in quella di destinazione. A ciascun elemento della griglia vengono assegnate proprietà dei materiali lineari ed elastiche, in base alle informazioni strutturali. Per risolvere l'equazione differenziale parziale risultante viene utilizzato un approccio di modellazione agli elementi finiti.

Infine, per procedere alla validazione qualitativa e quantitativa del DIR creato esistono diversi strumenti, tra cui l'ispezione visiva della griglia deformata, lo studio dei campi vettoriali di spostamento e le mappe di fusione delle immagini [27].

2.3.3 MRI-Only

Nel distretto pelvico, l'accuratezza della registrazione delle immagini può essere compromessa dalle continue variazioni anatomiche che interessano organi cavi come la vescica, il retto e le anse intestinali, oltre che dalle fluttuazioni della tensione muscolare o dei tessuti molli lungo la traiettoria delle radiazioni. Le maggiori incertezze geometriche si riscontrano a livello del GTV e della vescica; a ciò si sommano le ulteriori difficoltà legate alla variabilità volumetrica del retto, che risulta considerevolmente ridotta passando dalla TC alla RM [28]. Diversi studi presenti in letteratura hanno stimato che l'errore associato a queste procedure di registrazione si attesta

attorno ai 2 mm; un'incertezza rilevante di cui bisogna tenerne conto ampliando il margine del PTV, producendo di conseguenza un errore sistematico ad ogni seduta del trattamento [29].

Il frequente ricorso a scansioni TC (o CBCT) comporta la somministrazione di una dose cumulativa non trascurabile, correlata a un potenziale incremento di sviluppare cancro o carcinogenesi secondaria. Per ridurre questo rischio, quando possibile, si dovrebbe sostituire la TC con altre opzioni adeguate [30].

A differenza della metodologia ART, in cui rimane ancora la necessità di utilizzare una TC per ottenere la registrazione, l'adozione di un workflow basato esclusivamente su RM (*workflow MRI-Only*) risolve i problemi legati alle incertezze geometriche intrinseche alla co-registrazione multimodale e, contemporaneamente elimina totalmente l'esposizione del paziente alle radiazioni della TC. Oltre ai benefici appena discussi, questo approccio si dimostra altamente efficiente ed economico grazie all'eliminazione della sessione precedentemente necessaria per acquisire la TC. Per affrontare il limite della RM applicata nella RT, ossia l'assenza di informazioni dirette sulla DE dei tessuti, i ricercatori hanno proposto diversi algoritmi per la generazione di sCT. Tali immagini permettono di simulare i valori HU, necessari per il calcolo della dose [31].

Storicamente, il primo metodo utilizzato per affrontare questa problematica è basato sulla generazione di una "bulk-density" sCT. Questa strategia prevede la segmentazione delle macrostrutture anatomiche sulla RM e la successiva assegnazione di valori fissi di DE a ciascuna classe di tessuto. Nella sua forma più semplice, il metodo assegna una singola DE all'intero paziente; in soluzioni più elaborate, vengono assegnate DE distinte al tessuto osseo, al tessuto adiposo e alle sacche d'aria al fine di incrementare l'accuratezza della sCT e generare una distribuzione di dose più conforme. Sebbene la tecnica bulk-density abbia fornito risultati promettenti, in special modo nella pianificazione dei trattamenti nella regione pelvica, essa presenta dei limiti da considerare. Il tessuto osseo, infatti, presenta una DE fortemente eterogenea al suo interno che non può essere ridotta ad un singolo valore medio. Tale approssimazione può causare una sensibile distorsione della dose, in particolar modo se il volume tumorale è localizzato all'interno o in prossimità dell'osso, come nel caso delle metastasi ossee spinali [4].

Sebbene i metodi bulk-density siano semplici, richiedono molto tempo e dipendono dall'operatore. Per superare queste limitazioni, sono stati introdotti i metodi basati su atlante (*atlas-based*), i quali sfruttano complesse registrazioni non rigide tra uno o più coppie di immagini RM-TC di riferimento (atlanti) co-registrate e la RM target del paziente. Completata la fase di registrazione, i volumi TC degli atlanti utilizzati vengono fusi per generare la sCT. Questa tecnica offre il vantaggio di essere totalmente automatica ma perde robustezza in presenza di anomalie o di ampie variazioni anatomiche oltre a necessitare di un notevole costo computazionale per effettuare le registrazioni.

Nell'ultimo periodo la generazione di sCT ha trovato una svolta significativa grazie all'utilizzo del Deep Learning (DL). Questi modelli basati su architetture neurali consentono una rapida e accurata mappatura da valori di RM a valori HU senza la necessità di utilizzare registrazioni durante l'inferenza. Per contro, i modelli DL richiedono grandi database ed elevati costi computazionali per l'addestramento, oltre ad essere suscettibili alle variazioni nelle apparecchiature o nei protocolli di acquisizione.

Infine, vi sono i metodi basati su patch (*patch-based*), originariamente concepiti per la segmentazione e recentemente adattati alla sintesi di immagini. Essi si suddividono principalmente in due categorie: gli approcci a partizionamento semplice seguito da modellazione

e gli approcci basati sulla media (*non-local means*). Questi ultimi impiegano finestre scorrevoli per estrarre patch parzialmente sovrapponibili alla RM target da dataset TC-RM registrati affinemente (coorte di addestramento); tale estrazione viene opportunamente vincolata e guidata dall'intorno spaziale del voxel RM in analisi [32].

Al fine di valutare e confrontare quantitativamente la qualità delle sCT prodotte da questi diversi approcci, sono disponibili una serie di metriche d'immagine (basate su geometria o intensità) e dosimetriche. Queste sono riportate nella Tabella 1.

	Tipo di metriche	Metrica	Definizione	Valore ideale
Valutazione dell'immagine	Metriche basate sull'intensità	ME: mean error	$ME = \frac{1}{N}\sum_{i=1}^{N}(sCT_i - CT_i)$	0 HU
		MAE: mean absolute error	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert sCT_i - CT_i \rvert$	0 HU
		PSNR: peak signal to noise ratio	$PSNR = 10\log_{10}\left(\frac{Q^2}{MSE}\right)$	∞
		SSIM: structural similarity metric	$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$	1
		MSE: mean square error	$MSE = \frac{1}{N}\sum_{i=1}^{N}(sCT_i - CT_i)^2$	0
		RMSE: root mean square error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(sCT_i - CT_i)^2}$	0 HU
		NCC: normalized cross-correlation	$NCC = \frac{1}{N}\sum_{x,y,z}\frac{(I_{CT}(x,y,z) - \mu_{CT})(I_{sCT}(x,y,z) - \mu_{sCT})}{\sigma_{CT}\sigma_{sCT}}$	
	Metriche di fedeltà geometrica	DSC: dice score coefficient	$DSC = \frac{2\lvert V_{CT} \cap V_{sCT}\rvert}{\lvert V_{CT}\rvert + \lvert V_{sCT}\rvert}$	1
		HD: Hausdorff distance	$H(sCT, CT) = \max\big(h(sCT, CT), h(CT, sCT)\big)$	0 mm
		MASD: mean absolute surface distance	$MASD(A,R) = \frac{d_{ave}(S_A, S_R) + d_{ave}(S_R, S_A)}{2}$	0 mm
Valutazione della dose	Metriche per la differenza nella dose	Voxel-to-voxel dose differences	Differenza tra la distribuzione della dose calcolata sulla TC di riferimento e sulla sCT	0 Gy or 0%
		DVH difference	Differenze di dose su punti specifici del DVH (D_{max}, V_{70Gy}, etc.) per una data struttura	0 Gy or 0%
		Mean gamma	Valore della gamma media	0

	Metriche per l'analisi Gamma	Gamma pass-rate	Percentuale di pixel/voxel con un valore gamma inferiore a 1	100%

Tabella 1 Metriche per la valutazione quantitativa delle sCT. Abbreviazioni: N: numero di voxel; MSE: Mean square error; Q: intervallo di valori dei voxel di sCT e CT di riferimento; x: CT di riferimento; y: sCT; μ_x: valore medio di x; μ_y: valore medio di y; σ_x^2: varianza di x; σ_y^2; varianza di y; C_1 e C_2 sono espressi come $(k_1Q)^2$ e $(k_2Q)^2$; I_{CT}: valore HU della CT di riferimento, I_{sCT}: valore HU della sCT, μ_{CT}: valore medio dell'intensità della CT di riferimento, μ_{sCT}: valore medio dell'intensità della sCT, σ_{CT} e σ_{sCT}: deviazione standard della CT di riferimento e della sCT; V: volume della CT e della sCT; d_{ave}: distanza Euclidea assoluta; S_A: superficie del volume di segmentazione automatizzata; S_R: superficie della delimitazione dell'organo di riferimento.

In conclusione, l'impiego clinico dei modelli di DL garantisce una generazione rapida e accurata di sCT a partire da immagini RM. Gli studi documentano valori di MAE prossimi a 76 HU per i distretti testa-collo e fegato, mentre per la regione pelvica il MAE è intorno ai 40 HU. Per la valutazione dosimetrica si evidenziano incertezze inferiori al 2% per la RT cerebrale e inferiori al 1% per i distretti testa-collo, fegato, addome e aree pelviche. È stato anche dimostrato che l'impiego di input multimodali consente il raggiungimento di una qualità superiore delle sCT rispetto all'utilizzo di un singolo input. Attualmente, le sfide aperte più significative in questo ambito riguardano la standardizzazione dei protocolli di valutazione e l'analisi della robustezza dei modelli in scenari multicentrici [33].

3 Deep Learning in radioterapia

L'implementazione dell'intelligenza artificiale (IA) sta attualmente interessando numerosi settori, incluso quello medico. Nello specifico ambito della radio-oncologia, i modelli di apprendimento automatico consentono di automatizzare significativamente il workflow clinico. Tuttavia, una limitata conoscenza e interpretabilità di questi modelli può rallentarne l'integrazione clinica. Per agevolare l'introduzione dei modelli AI nei percorsi radioterapici, risulta indispensabile aderire a specifiche linee guida di carattere generale riguardanti la loro implementazione e il rispettivo QA. Tra gli impieghi più frequenti in RT si presentano la segmentazione automatica di immagini/volumi, la pianificazione automatizzata del piano di cura e la generazione di sCT.

I workflow di RT sono suddivisi in diverse fasi, le quali richiedono spesso molto tempo e incidono sulla qualità del trattamento e, di conseguenza, sugli esiti clinici per il paziente. In questo contesto, l'IA emerge come uno strumento ideale per migliorare la qualità, la standardizzazione e le tempistiche di queste fasi, ottenendo benefici come l'ottimizzazione dell'erogazione della dose in termini di accuratezza e sicurezza. Nello specifico, raggiungere un flusso di lavoro semplificato e rapido è fondamentale per la messa in atto di modalità altamente conformi come l'ART. Concettualmente, l'IA raggruppa una serie di algoritmi progettati per eseguire task assimilabili alle capacità intellettive umane; al suo interno si distinguono come principali sottodomini il Machine Learning (ML) e il DL [34].

In tempi recenti, il DL si è affermato come l'approccio computazionale dominante all'interno del ML, raggiungendo traguardi straordinari in molteplici compiti cognitivi complessi, eguagliando/oltrepassando le performance umane. Rispetto ai metodi ML convenzionali, Il DL vanta la capacità di apprendere pattern complessi da enormi dataset e di automatizzare l'estrazione del set di feature rilevanti per svolgere determinati task. Essendo il DL un sottoinsieme del ML (Fig. 2), trae ispirazione dalle reti di elaborazione neuronale e dalle modalità di gestione delle informazioni tipiche del cervello umano. Il suo funzionamento quindi, non si fonda sulla progettazione di rigide regole, bensì sfrutta una vasta quantità di dati per cogliere le relazioni utili per mappare l'input fornito nelle rispettive etichette.

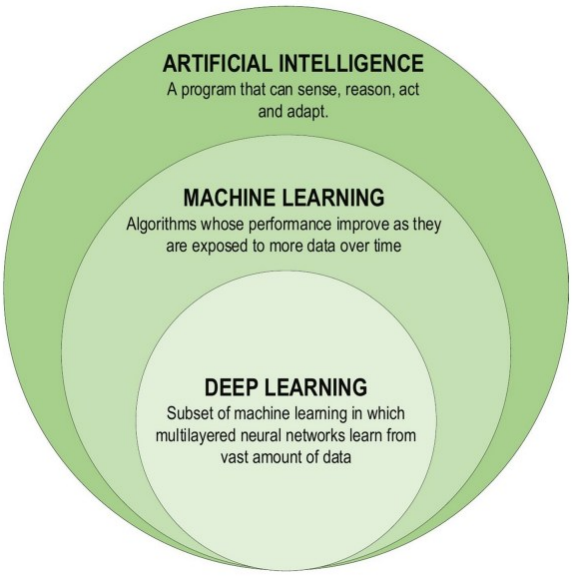


Fig. 2 Sottoclassi dell'Intelligenza Artificiale [35].

Tra le varie caratteristiche distintive del DL troviamo:

- Approccio di apprendimento universale: la versatilità del DL concede la possibilità di utilizzo in quasi tutti i domini applicativi.
- Robustezza: nelle tecniche di DL non sono richieste feature ingegnerizzate ed estratte manualmente. Al contrario, le feature vengono apprese ed ottimizzate in modo automatico in relazione al compito in esame. In questo modo, si ottiene robustezza rispetto alle normali variazioni dei dati di input.
- Generalizzazione: una singola architettura di DL può essere riadattata per analizzare tipologie di dati differenti, trovando applicazione in domini differenti. Questo approccio, spesso definito TL, risulta utile in problemi in cui i dati sono insufficienti.
- Scalabilità: capacità di un modello di migliorare le prestazioni all'aumentare delle risorse impiegate.

Le tecniche di DL possono essere classificate in tre categorie principali in base alla disponibilità di dati etichettati: non supervisionate, parzialmente supervisionate e supervisionate. Il vantaggio principale della tecnica supervisionata è la capacità di produrre risultati a partire da conoscenza pregressa. Tuttavia, un rischio intrinseco di questo approccio è la possibilità di generare confini decisionali sovradeterminati (*overfitting*), un fenomeno che può verificarsi da molteplici fattori, tra cui l'utilizzo di un modello eccessivamente complesso rispetto alla quantità e alla varietà dei dati forniti nel set di addestramento. Ciononostante, si tratta di una strategia tendenzialmente più semplice rispetto ad altre e che assicura performance elevate [35].

La procedura di integrazione clinica si divide in due parti: l'addestramento del modello di IA e la successiva verifica della accuratezza e riproducibilità. Si parla quindi di una prima fase di addestramento/validazione e di una seconda fase di test. Il training necessita solitamente di voluminosi database, preferibilmente costituiti da dati annotati e di massima qualità. L'impiego di dati acquisiti in sede garantisce la piena aderenza ai protocolli di imaging e alle direttive cliniche del proprio dipartimento. È fondamentale che tali dati vengano attentamente revisionati per descrivere quanto questi siano rappresentativi dell'intera popolazione a cui si intende estendere l'utilizzo del modello. All'atto pratico, per via del continuo aggiornamento dei protocolli clinici, si prediligono dataset non eccessivamente grandi ma meticolosamente curati, piuttosto che collezioni enormi ma di scarsa qualità.

La validazione del modello viene condotta avvalendosi di un set di dati ridotto, chiamato *validation-set*, il quale viene sottoposto ad analisi sia di stampo qualitativo che quantitativo. Questo set deve fungere da campione rappresentativo dei dati su cui il modello verrà applicato. È opportuno considerare una serie di metriche di valutazione coerenti con lo scopo applicativo della rete. Il primo passo prevede l'utilizzo del modello pre-addestrato sul validation-set e la valutazione degli output forniti attraverso metriche quantitative, le quali forniscono informazioni sulla qualità delle predizioni del modello rispetto al *ground truth*. Soddisfatti questi requisiti, gli output vengono sottoposti alla valutazione qualitativa di professionisti del settore, come medici, fisici o tecnici di RT. Questa fase di revisione umana permette di individuare le aree di disaccordo tra la predizione del modello e il giudizio di un esperto, offrendo spunti per l'ulteriore affinamento delle caratteristiche del modello. I modelli sviluppati da fornitori commerciali spesso omettono questo passaggio, ma necessitano della prova della marcatura CE/FDA e di report di convalida dettagliati che ne attestino le prestazioni, sotto responsabilità del fornitore. Per rendere il modello più trasparente e comprensibile è raccomandabile fornire informazioni relative all'uso previsto, al dominio di applicabilità, ai limiti del modello, descrivendo dettagliatamente i set di training e validazione, gli standard utilizzati e le metriche di valutazione adottate. Inoltre, i fornitori sono tenuti a fornire un protocollo di valutazione affinché i clienti, su base annuale, possano monitorare

le performance del modello. L'obiettivo della fase di test è quello di ottenere una valutazione indipendente delle prestazioni finali della rete, attestando la robustezza e le capacità di generalizzazione raggiunte, in modo da poter definire la tipologia di pazienti sulla quale il modello risulta applicabile [34].

Dal punto delle risorse, lo sviluppo e la gestione operativa dei modelli di DL si necessita di infrastrutture hardware capaci di sostenere calcoli paralleli ad altissima velocità. Le unità di elaborazione grafica (GPU, *Graphics Processing Unit*) sono quindi preferite alle unità di elaborazione centrale (CPU, *Central Processing Unit*) in questo contesto, principalmente perché dispongono di un numero maggiore di unità aritmetiche e logiche (ALU, *Arithmetic Logic Unit*), consentendo l'esecuzione simultanea di più operazioni. Inoltre, le GPU offrono la possibilità di gestire grandi insiemi di dati, grazie alla loro progettazione che incorpora bus più ampi e frequenze di clock della memoria più elevate rispetto alle CPU, il che rende la scelta preferita soprattutto per l'addestramento di reti neurali tridimensionali [36].

3.1 Synthetic CT

La sCT, ricavata direttamente dalle acquisizioni di RM, rappresenta una valida e innovativa alternativa alla TC nell'ambito della RT. L'adozione di questo approccio elimina le imprecisioni geometriche dovute alla co-registrazione di immagini provenienti da modalità di acquisizione differenti, riducendo al contempo i costi operativi e l'esposizione del paziente a radiazioni ionizzanti. Attualmente, esistono già soluzioni per la generazione di sCT dotate di marchio CE/FDA impiegabili clinicamente per alcuni distretti corporei quali pelvi, encefalo e complesso testa-collo, mentre per le altre regioni anatomiche sono in corso di sviluppo e validazione delle architetture DL più avanzate [36].

Dal punto di vista architetturale, i modelli di DL impiegati per la generazione di sCT a partire da RM possono essere suddivisi in due categorie: le reti basate esclusivamente su generatori (*only-generator*) e le GAN con le rispettive varianti come conditional-GAN (cGAN), Least Square GAN (LSGAN) e Cycle-GAN. Un parametro progettuale rilevante è la dimensionalità dei dati in ingresso: le reti che elaborano volumi 3D risolvono il problema della discontinuità tra le slice, tipica delle reti 2D. Tuttavia, il processamento tridimensionale è notoriamente ostico da ottimizzare a causa del massiccio incremento di parametri addestrabili, oltre a richiedere una quantità esponenzialmente maggiore di memoria GPU e di dati. Un valido compromesso si realizza con l'approccio 2.5D, il quale integra informazioni spaziali tridimensionali vicine alla slice in esame, mantenendo la complessità computazionale su livelli gestibili e riuscendo comunque a ridurre gli artefatti di continuità causati dalle reti bidimensionali. In ultima analisi, i requisiti quantitativi dei dati e le metodologie di apprendimento da adottare dipendono strettamente dalla specifica architettura neurale selezionata [33] [36].

A prescindere dall'architettura adottata, un passaggio fondamentale è il pre-processing delle RM, utile per incrementare la robustezza del modello attraverso la riduzione della variabilità della qualità d'immagine e la mitigazione dell'influenza negativa di eventuali artefatti. Solitamente queste operazioni preparatorie comprendono la correzione delle distorsioni di campo, il ricampionamento spaziale, le correzioni di fedeltà geometrica, l'equalizzazione dell'istogramma o la normalizzazione delle intensità RM e, se necessario, la registrazione d'immagini. Spesso si adotta anche il clipping dell'intensità per rimuovere i valori anomali, riconducibili principalmente al rumore di fondo. Inoltre, qualora la tipologia di modello impiegato richieda coppie di immagini

registrate (RM e TC), è fondamentale considerare il tempo intercorso tra le due acquisizioni. Questo intervallo è specifico per la sede anatomica considerata: per il distretto addominale, fortemente soggetto a movimenti fisiologici, sono raccomandabili intervalli brevi, al di sotto della mezz'ora; per aree soggette a variazioni morfologiche più lente, come il cranio o il distretto testa-collo, si possono tollerare intervalli più lunghi, dell'ordine di uno o due giorni [36].

Oltre all'adeguata preparazione delle immagini, l'efficacia e l'affidabilità di un modello di DL per la sintesi di sCT dipendono intrinsecamente dalla qualità e dalle caratteristiche del dataset impiegato durante la sua fase di training. Il compromesso da affrontare durante la progettazione del modello risiede nello scegliere tra un modello specifico per sede anatomica e per un preciso macchinario, quindi altamente accurato per questo caso specifico e compromettere l'accuratezza a favore della capacità di generalizzazione considerando un approccio multicentrico e considerando più sedi anatomiche. Di conseguenza, i dati di addestramento devono rispecchiare fedelmente la reale coorte clinica su cui l'algoritmo andrà ad operare (ad esempio, pazienti di sesso maschile trattati a livello pelvico e privi di protesi d'anca). Questo implica la necessità di descrivere la popolazione target attraverso la definizione di parametri demografici e antropometrici (ad esempio età, altezza, sesso, indice di massa corporea, ecc.); il modello non dovrebbe essere applicato al di fuori di tali intervalli. Al contempo, è opportuno che il *training-set* inglobi la massima eterogeneità anatomica possibile per il distretto di interesse. È raccomandabile includere scenari clinici diversificati, come precedenti interventi chirurgici, presenza di stent e variazioni morfologiche indotte dalla neoplasia stessa (in termini di volume, posizionamento e grado di distorsione indotto sui tessuti sani adiacenti). L'introduzione di un'ampia variabilità inter-paziente estende notevolmente il dominio di applicabilità del modello rispetto all'utilizzo di criteri di esclusione troppo restrittivi, ma comporta l'utilizzo di dataset sensibilmente più ampi, il che si traduce nell'impiego di reti neurali strutturalmente più complesse e di maggior tempo computazionale. Al contrario, i criteri di esclusione vengono tipicamente applicati per scartare RM caratterizzate da evidenti artefatti d'immagine o da alterazioni anatomiche anomale [36].

Una volta definita la composizione della coorte, la determinazione della quantità ideale di dati necessari per un addestramento ottimale può essere guidata dal monitoraggio delle curve di apprendimento sia in fase di training che di validazione. Per compensare la scarsità di campioni e ottimizzare l'addestramento, si può ricorrere al TL da modelli pre-addestrati, sfruttando le feature già apprese in altri domini. Inoltre, per espandere esplicitamente la variabilità e la quantità di dati a disposizione, vengono implementate tecniche di *data augmentation*. Quest'ultima si divide in due categorie:

- La manipolazione tramite algoritmi dei dati originali tramite trasformazioni spaziali come traslazione, rotazione, ribaltamento, ridimensionamento, ritaglio e deformazione (o qualsiasi combinazione di queste).
- La generazione di dati artificiali tramite modelli generativi [36].

Passando dal piano algoritmico a quello pratico, lo sviluppo e l'implementazione clinica della generazione di sCT mediante DL si scontrano attualmente con diverse sfide tecniche:

- Differenze in termini di scanner RM e protocolli di acquisizione tra i dati di training e i dati di input utilizzati nella pratica clinica. I rischi che ne conseguono dovrebbero essere mitigati dal produttore, che è tenuto a dichiarare le limitazioni di applicabilità dell'algoritmo.

- Incertezze di registrazione tra RM e TC nel set di training unite alla mancanza di un *ground truth* per verificarne l'affidabilità.
- Potenziali allucinazioni di rete, che producono un output sCT non verificabile in assenza di una TC di pianificazione, ad esempio in presenza di impianti metallici.
- L'assenza di dataset standardizzati per la validazione rende difficile determinare quantitativamente la superiorità di un modello specifico.

A fronte di tali criticità, è opportuno condurre rigorose analisi del rischio per rilevare potenziali modalità di malfunzionamento. Tutte le questioni del caso devono essere esaminate ricorrendo ad approcci come Health FMEA (*Failure Mode and Effects Analysis*) o procedure analoghe. Inoltre, per favorire l'adozione del workflow MR-only nei reparti di RT, le linee guida raccomandano ai fornitori di garantire massima trasparenza, mettendo a disposizione degli utilizzatori:

- Una descrizione del database di addestramento, utile per definire i valori di tolleranza per i controlli di qualità clinici, come ad esempio: intervallo di peso, altezza ed età dei pazienti, situazioni cliniche (ad esempio, presenza di pazienti operati), sequenze RM da utilizzare, intensità del campo magnetico, protocolli da seguire, intervalli di parametri, ecc.
- Metodologie algoritmiche mirate alla riduzione degli artefatti indotti da mezzi di sintesi, impianti e viti chirurgiche.
- Flessibilità nella selezione del FOV (field of view), assistita dall'esclusione automatica delle aree con segnale RM basso.
- Protocolli RM dedicati per la RT stereotassica corporea (SBRT) che garantiscono una maggiore accuratezza geometrica.
- Tag DICOM che consentono di tracciare la versione dell'algoritmo utilizzato e le immagini RM di origine impiegate per generare la sCT.

Infine, si incoraggiano le aziende produttrici a sviluppare e rilasciare strumenti per la QA che possano facilitare i fisici medici nello sviluppo di protocolli di controllo robusti [36].

3.1.1 Networks

L'insieme delle architetture di DL a disposizione per la generazione di sCT è vasto ed eterogeneo; come mostrato in Fig. 3, queste reti sono classificate secondo una struttura gerarchica.

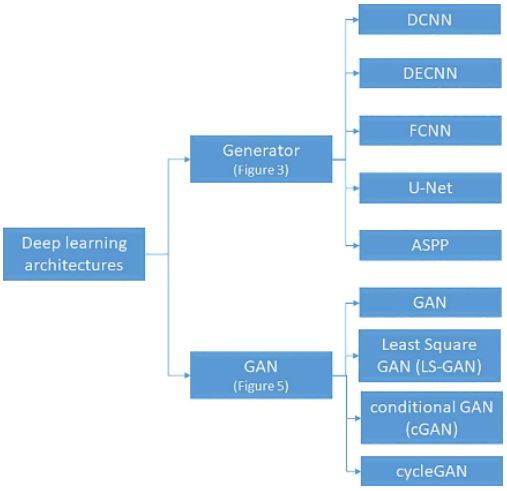


Fig. 3 Gerarchia delle architetture deep learning [33].

I modelli di tipo generator-only operano come funzioni avanzate di mappatura end-to-end, con lo scopo di convertire l'immagine RM in ingresso nella sua corrispondente immagine TC. Nella fase di addestramento, il generatore tenta di minimizzare una specifica funzione obiettivo, nota come loss function (generalmente una *voxel-wise loss function*). Questa funzione è una misura di similarità basata sull'intensità tra sCT predetta dalla rete e la TC reale utilizzata come riferimento; perciò, la sua minimizzazione comporta il raggiungimento di predizioni più simili alla realtà. In Fig. 4 si presenta l'architettura globale del modello basato esclusivamente sul generatore.

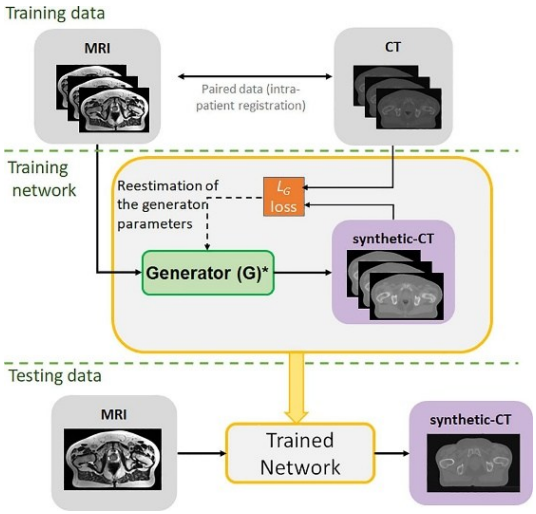


Fig. 4 Illustrazione del modello generator-only (il modello del generatore varia a seconda delle reti) [33].

Per la generazione di sCT a partire da dati RM, le architetture dei generatori si basano prevalentemente sulle Reti Convoluzionali Encoder-Decoder (CED). Tra le varianti più note in letteratura si piazzano le Deep CED network (DCNN), Deep Embedding CNN (DECNN), Fully Convolutional Network (FCN), U-net, efficient CNN (eCNN), ResNet, SE-ResNet e DenseNet. Le reti CED sono ampiamente utilizzate grazie alle loro eccellenti prestazioni. L'architettura alla base di questi modelli prevede un primo stadio di codifica in cui l'encoder applica un sottocampionamento (*downsampling*) alle feature maps di basso livello per estrarre feature ad alto livello. Successivamente, nella fase di decodifica, il decoder sovracampiona (*upsampling*) le feature di alto livello appena ottenute, ricorrendo all'applicazione di strati convoluzionali trasposti per costruire la sCT finale. Il limite dei modelli generator-only basati su CNN è che portano a risultati sfocati, un difetto imputabile al disallineamento generale tra RM e TC.

Per aggirare questo ostacolo, l'attenzione si è spostata verso le GAN, architetture basate sulla strategia di apprendimento avversaria che permette di produrre sCT qualitativamente superiori rispetto alle soluzioni generator-only. Il concetto alla base (schematizzato in Fig. 5) prevede l'addestramento simultaneo e competitivo di due distinte reti neurali: il generatore e il discriminatore. Il generatore tenta di creare immagini talmente verosimili da ingannare il discriminatore, mentre questo viene addestrato ad affinare la sua capacità di distinzione fra dati reali e artificiali [33].

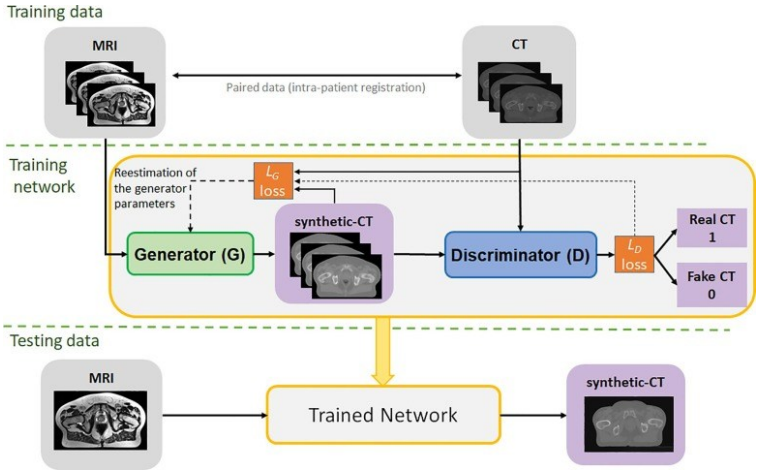


Fig. 5 Architettura GAN [33].

Evidenze scientifiche [37], [38], [39] confermano che l'impiego delle GAN contrasta efficacemente l'effetto sfocatura, preservando i dettagli anatomici e le caratteristiche dei bordi. Questo si traduce in un marcato incremento dell'accuratezza della sCT, in particolar modo a livello delle strutture ossee. Inoltre, il discriminatore analizza le caratteristiche locali attraverso la suddivisione delle immagini (reali o sintetiche) in patch, contribuendo a mitigare gli errori di registrazione causati da un allineamento imperfetto fra RM e TC. Dall'altra parte, la difficolta delle GAN tradizionali risiede nella loro instabilità durante l'addestramento e quindi una difficile convergenza del modello. Questo implica che la regolazione degli iperparametri diventa un'operazione altamente sensibile per il corretto funzionamento del modello, scongiurando eventi come la scomparsa o l'esplosione dei gradienti. Al fine di stabilizzare il processo di training, i ricercatori hanno sviluppato svariate estensioni e sottoclassi [33]:

- LSGAN: le GAN classiche impiegano la *sigmoid/binary cross-entropy* come loss function, esponendosi al rischio di saturazione dell'apprendimento per via del problema dei gradienti evanescenti. Le LSGAN, invece, utilizzano la *least square loss function*, una funzione obiettivo basata sui minimi quadrati, che ha il vantaggio di penalizzare fortemente i campioni artificiali che si collocano a notevole distanza dal confine decisionale, stabilizzando drasticamente il processo di apprendimento.
- cGAN: poiché le GAN non permettono alcun controllo esplicito sulla generazione dei dati, le varianti cGAN superano questo limite condizionando l'output di entrambe le reti (generatore e discriminatore) attraverso l'immissione di informazioni supplementari, come le etichette di classe o l'immagine RM stessa. Una delle architetture di maggior successo in questo scenario è il framework Pix2Pix, appositamente progettato per la traduzione image-to-image ad alta risoluzione. Questo modello è tipicamente composto da un generatore U-Net accoppiato a un discriminatore PatchGAN. Uno dei principali vantaggi delle cGAN è che garantiscono ragionevoli traduzioni *image-to-image* anche partendo da set di addestramento modesti, dall'altro lato esigono che le immagini RM e TC utilizzate durante il training siano registrate rigorosamente.
- Cycle-GAN: concepite per effettuare traduzione tra due modalità di imaging, il principio delle Cycle-GAN si fonda sull'estrazione delle peculiarità distintive di entrambe le tipologie di immagine per individuare la relazione tra di esse. L'infrastruttura prevede l'impiego di GAN interconnesse: la prima converte la RM in sCT, mentre la seconda genera un'immagine RM sintetica partendo dalla sCT appena creata, e tentando di riottenere la stessa immagine fornita in input. L'addestramento simultaneo è governato

da un *cyclic loss function* che mira a minimizzare lo scarto tra la TC reale e quella sintetizzata dal ciclo generativo. L'enorme vantaggio del framework Cycle-GAN risiede nel non necessitare di dataset RM/TC geometricamente accoppiati.

Va riconosciuto che i modelli descritti non sono direttamente comparabili, vi sono diverse fonti di disparità che coinvolgono i protocolli di imaging, i parametri operativi degli scanner, la dimensione della coorte di addestramento e validazione, oltre che nei criteri di valutazione. Per una valutazione appropriata sarebbe necessario utilizzare gli stessi dati per confrontare le prestazioni delle varie reti [33].

3.1.2 Generatore e discriminatore

Nel contesto delle architetture generative, le reti di tipo encoder-decoder rivestono un ruolo di primo piano per la sintesi d'immagine. Il funzionamento di questi modelli prevede che l'input attraversi una serie di layer deputati al sottocampionamento progressivo (*downsampling*) fino a raggiungere uno strato di minima dimensionalità, definito *bottleneck*. Da questo punto, il processo subisce un'inversione, procedendo con il sovracampionamento (*upsampling*) per ricostruire l'immagine finale. Tuttavia, questo comporta che l'architettura imponga la compressione del flusso informativo, forzandolo ad attraversare la strozzatura centrale della rete.

In numerose applicazioni di traduzione image-to-image, l'input e l'output condividono intrinsecamente una vasta gamma di informazioni di basso livello, come ad esempio la posizione e la geometria dei contorni principali. Al fine di preservare e trasferire direttamente tali caratteristiche senza disperderle nella fase di compressione, risulta vantaggioso introdurre dei collegamenti diretti noti come skip connection, dando così origine all'architettura "U-Net". Nella pratica, una skip connection viene instaurata tra l'i-esimo strato della fase di codifica e il corrispondente strato (n-1) nella fase di decodifica, con *n* che rappresenta il numero totale di stati. Questo collegamento agisce concatenando direttamente i canali delle feature maps tra encoder e decoder, bypassando così il *bottleneck* [40].

Come illustrato in Fig. 6, l'architettura U-Net si configura come una struttura CED. Il percorso di codifica si avvale di una serie di filtri convoluzionali bidimensionali per l'estrazione di feature, tipicamente seguiti da un'operazione di normalizzazione (come la batch o instance normalization), dall'applicazione di una funzione di attivazione non lineare (ad esempio ReLU, Leaky-ReLU o PReLU) e da un livello di max-pooling per la riduzione dimensionale. Invece, il ramo di decodifica aggrega le feature e le informazioni spaziali impiegando una sequenza di layer convoluzionali trasposti simmetrici (*up-convolution*) e operatori di upsampling. Queste operazioni culminano nell'integrazione delle feature maps provenienti dall'encoder tramite concatenazione e in un passaggio finale di convoluzione con funzione di attivazione ReLU.

Un'ulteriore architettura generativa ampiamente adottata all'interno delle reti avversarie è la ResNet (*Residual Neural Network*). Il principale punto di forza risiede nella facilità di ottimizzazione, anche in presenza di una marcata profondità strutturale, conferendo la possibilità di poter raggiungere un maggior livello di precisione. Come visibile in Fig. 6, lo schema alla base di un generatore ResNet comprende tipicamente tre blocchi convoluzionali iniziali (costituiti da un'operazione di convoluzione, uno strato di normalizzazione batch e attivazione ReLU), seguiti da nove blocchi residui. Ogni blocco residuo reitera l'uso di convoluzioni, normalizzazione batch e ReLU, fino ad arrivare al fully connected layer nel caso della classificazione [33] o nel caso della generazione di immagini al percorso di upsampling e Softmax layer.

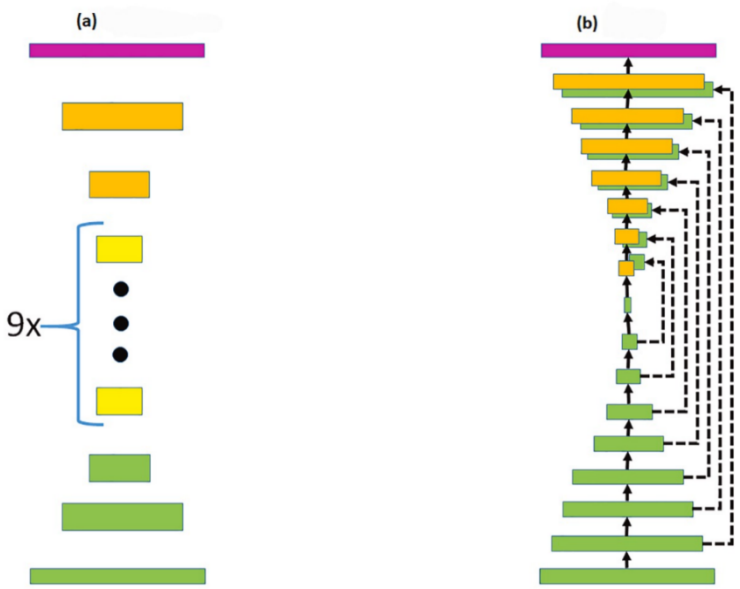


Fig. 6 (a): ResNet; (b): U-Net. In giallo ResNet block (Conv+BN+ReLU+FC+Sigmoid); in arancione: Conv transpose (Upsampling+Conv+ReLU); in verde: Convolution block (Conv+ReLU+Norm+Pooling); in viola: Softmax layer; le frecce indicano le skip connections [33].

Per quanto riguarda le dinamiche di ottimizzazione, è noto che l'applicazione esclusiva di loss L1 o L2 generi predizioni sfocate e bordi non definiti. Sebbene queste loss non favoriscano la nitidezza dei dettagli ad alta frequenza, risultano comunque estremamente affidabili nel catturare accuratamente le caratteristiche a basse frequenze delle immagini in input. Per i casi in cui si verifica questa situazione, non risulta necessario implementare un nuovo framework. Per assicurare la fedeltà delle basse frequenze il termine L1 è più che adeguato allo scopo. Mentre, per le alte frequenze, la strategia ottimale consiste nel delegare alla rete avversaria (il discriminatore) il compito di individuare e penalizzare sfocature e bordi non definiti.

Affinché il discriminatore possa focalizzarsi adeguatamente sui dettagli ad alta frequenza, è sufficiente limitare il suo campo ricettivo all'analisi di porzioni dell'immagine, anche chiamate "patch". Questo principio ha portato allo sviluppo dell'architettura PatchGAN, progettata per penalizzare gli errori strutturali su scala locale. Invece di emettere un singolo giudizio globale, un discriminatore PatchGAN esamina le patch di dimensione NxN e classifica ciascuna di esse come reale o fittizia (Fig. 7). La procedura consiste quindi nel far scorrere il modello sull'intera immagine in modo convoluzionale, mediando infine le valutazioni locali per produrre il giudizio finale della rete. Il valore di *N* può essere sensibilmente inferiore rispetto alla dimensione totale dell'immagine senza compromettere la qualità del risultato. Questo è un vantaggio in termini computazionali, in quanto permette di ridurre drasticamente il numero di parametri, permettendo di processare le immagini in un tempo inferiore. Da un punto di vista concettuale, il discriminatore PatchGAN interpreta l'immagine come un campo casuale di Markov, assumendo che i pixel distanti più del diametro della patch siano statisticamente indipendenti tra loro. Grazie a questa proprietà, esso agisce come una loss function per la texture e lo stile visivo. Per quanto concerne il valore dimensionale da attribuire alle patch, l'adozione del formato 16x16 è sufficiente a garantire un'ottima nitidezza, ma tende a generare artefatti di tassellatura (*tiling*); al contrario, incrementare il campo ricettivo a una dimensione di 70x70 pixel consente di attenuare efficacemente queste distorsioni [40].

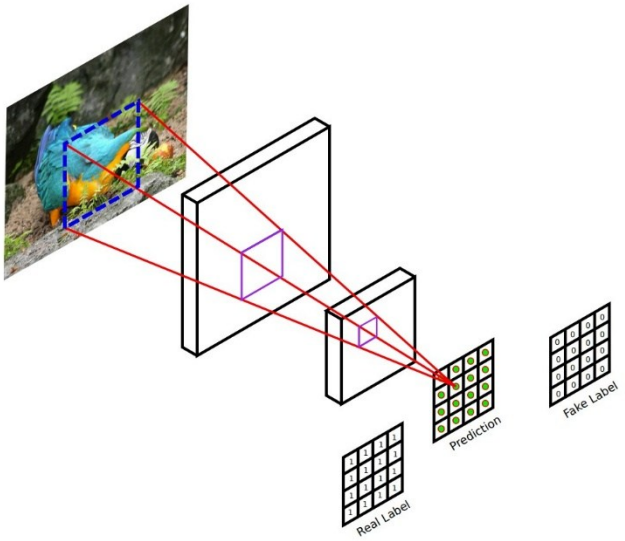


Fig. 7 PatchGAN [41].

3.1.3 Architetture GAN e funzioni di Loss

Le GAN sono modelli generativi progettati per apprendere una funzione di mappatura in grado di trasformare un vettore di rumore casuale z in un'immagine di output y, secondo la relazione G: z → y. Dal punto di vista matematico, la *loss function* generica di una GAN assume la seguente formulazione:

$$\mathcal{L}_{GAN}(G,D) = \mathbb{E}_{y\sim p_{data}(y)}[logD(y)] + \mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))]. \quad (1)$$

Le varianti condizionali (cGAN), invece, estendono questa azione apprendendo una trasformazione che parta non solo dal rumore z, ma anche da un'immagine osservata in input x, definendo così la mappatura G: {x,z} → y. Dal punto di vista matematico, la loss function di una cGAN assume la seguente formulazione:

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[logD(x,y)] + \mathbb{E}_{x,z}[log(1-D(x,G(x,z)))], \quad (2)$$

Dove il generatore (G) viene addestrato per minimizzare questa funzione, mentre il discriminatore (D), agendo da avversario, mira a massimizzarla. La soluzione ottima del problema si configura quindi come un "gioco minmax":

$$G^* = \arg min_G max_D \mathcal{L}_{cGAN}(G,D)\,. \quad (3)$$

Diversi studi hanno dimostrato come l'integrazione dell'obiettivo avversario con una funzione di costo tradizionale porti a risultati decisamente superiori. Pur lasciando inalterato il ruolo del discriminatore, si impone al generatore un duplice compito: ingannare la rete avversaria e, simultaneamente, generare immagini sintetiche che si discostino il meno possibile dal *ground truth*. Poiché metriche come la norma L2 tendono a produrre risultati eccessivamente levigati, si predilige spesso l'impiego della distanza L1, la quale incentiva una minore sfocatura dell'immagine predetta. La loss L1 è definita come:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}\left[||y - G(x,z)||_1\right]. \quad (4)$$

Di conseguenza, la funzione obiettivo complessiva da ottimizzare diventa:

$$G^* = \arg min_G max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G). \quad (5)$$

All'interno di questo processo, è cruciale considerare il ruolo della variabile aleatoria z. Omettendo tale rumore, la rete sarebbe comunque in grado di mappare l'input x nell'output y, ma genererebbe soluzioni puramente deterministiche (assimilabili a distribuzioni a delta di Dirac), perdendo la capacità di modellare la naturale varianza dei dati. Sebbene le architetture cGAN iniziali tentassero di risolvere la problematica concatenando semplicemente un rumore gaussiano z all'input x, il generatore tendeva autonomamente a ignorare tale segnale aleatorio. Una strategia più robusta consiste nell'introdurre stocasticità veicolando il rumore mediante strati di dropout all'interno della rete generativa, mantenendoli attivi sia durante il training sia in fase di validazione/test. Malgrado questa contromisura, la variabilità introdotta nelle immagini sintetiche rimane spesso modesta; lo sviluppo di modelli cGAN capaci di restituire output marcatamente stocastici e di catturare appieno l'entropia delle distribuzioni condizionali rappresenta tuttora un quesito aperto nella ricerca [42] [40].

Analizzando più a fondo la dinamica algoritmica, l'obiettivo principale di G (parametrizzato dai pesi θ_g) è assimilare la distribuzione intrinseca dei dati reali, p_{data}. Il processo prende avvio campionando un vettore z da una distribuzione nota $p_z(z)$ (solitamente uniforme o gaussiana), il quale viene poi mappato nello spazio dell'immagine $G(z; \theta_g)$ attraverso la rete differenziabile. Parallelamente, il discriminatore D (definito dai pesi θ_d) funge da classificatore binario incaricato di distinguere i campioni originali da quelli artificiali. Le architetture GAN convenzionali sfruttano tipicamente la sigmoid/binary *cross-entropy* come funzione di penalità. Tuttavia, in fase di ottimizzazione del generatore, questa formulazione induce frequentemente la scomparsa del gradiente (*vanishing gradient*) per quei campioni generati che, pur trovandosi sul versante corretto del limite decisionale stabilito dal discriminatore, restano ancora significativamente distanti dalla distribuzione dei dati target.

Per ovviare a questo problema sono state proposte le LSGAN. Implementando uno schema di classificazione a-b, in cui le costanti *a* e *b* codificano rispettivamente le label per le immagini sintetiche e per i dati reali, le funzioni obiettivo avversarie basate sui minimi quadrati si traducono nelle seguenti equazioni:

$$min_D \mathcal{L}_{LSGAN}(D) = \frac{1}{2}\mathbb{E}_{x\sim p_{data}(x)}[(D(x) - b)^2] + \frac{1}{2}\mathbb{E}_{z\sim p_z(z)}[(D(G(z)) - a)^2] \quad (6)$$

$$min_G \mathcal{L}_{LSGAN}(G) = \frac{1}{2}\mathbb{E}_{z\sim p_z(z)}[(D(G(z)) - c)^2], \quad (7)$$

In queste espressioni, il parametro *c* rappresenta il valore numerico a cui il generatore tenta di far convergere le valutazioni del discriminatore per i campioni artificiali.

I punti di forza algoritmici delle LSGAN si manifestano attraverso due aspetti chiave. In primis, mentre le reti avversarie tradizionali quasi azzerano la loss per i campioni situati molto distanti dal confine di decisione (purché classificati correttamente dalla rete), la funzione ai minimi quadrati continua a sanzionare tali occorrenze, spingendole attivamente verso il decision boundary (il quale rimane fisso durante l'aggiornamento dei pesi di G). In secondo luogo, proprio questa marcata penalizzazione garantisce gradienti di modulo elevato anche per campioni distanti, mitigando alla radice l'evanescenza degli stessi e stabilizzando drasticamente l'intero iter di apprendimento. Questa peculiarità risulta evidente dal punto di vista analitico: come illustrato in Fig. 8, la curva della funzione *least squares* presenta un gradiente nullo in un singolo punto, in

netta contrapposizione alla curva *cross-entropy*, la quale tende a saturare, annullando i gradienti non appena la classificazione assume un grado di confidenza sufficiente [43].

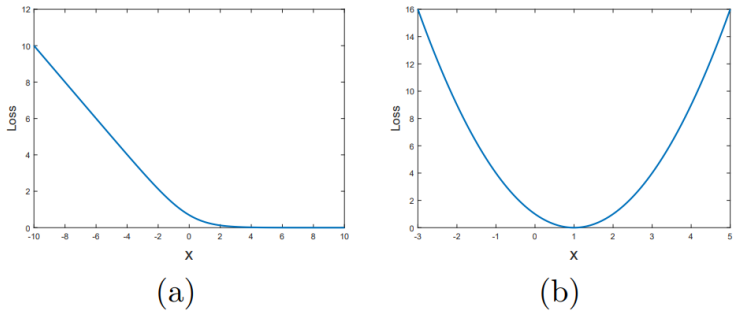


Fig. 8 (a): sigmoid cross entropy loss function. (b): least squares loss function [43].

Infine, per completezza rispetto alle architetture introdotte, è opportuno citare le Cycle-GAN. Mentre le cGAN richiedono dataset di immagini rigorosamente co-registrate, le Cycle-GAN superano questo limite consentendo la traduzione tra domini non accoppiati (es. RM e TC non allineate spazialmente). A tale scopo, l'architettura impiega due generatori distinti operanti in direzioni opposte: il generatore principale G, che traduce le immagini dal dominio di origine X a quello di destinazione Y ($G: X \rightarrow Y$), e un generatore inverso F, che compie l'operazione contraria riportando l'immagine da Y al dominio di origine X ($F: Y \rightarrow X$). Per garantire che un'immagine tradotta nel nuovo dominio mantenga un'elevata coerenza anatomica, viene introdotta la Cycle Consistency Loss, basata sulla norma L1:

$$\mathcal{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}\left[||F\big(G(x)\big) - x||_1\right] + \mathbb{E}_{y\sim p_{data}(y)}\left[||G\big(F(y)\big) - y||_1\right]. \quad (8)$$

Attraverso questa formulazione, la rete è costretta a rispettare la consistenza ciclica: un'immagine tradotta in avanti ($G(x)$) e poi riportata indietro dal generatore inverso ($F\big(G(x)\big)$) deve coincidere il più possibile con l'immagine di partenza x. La funzione obiettivo totale della Cycle-GAN combina quindi la loss avversaria di entrambe le reti con la loss di consistenza ciclica, permettendo di apprendere le mappature e preservando le caratteristiche strutturali intrinseche dei tessuti anche in assenza di un ground truth geometricamente corrispondente [44].

3.2 Transfer Learning

Nelle moderne reti neurali profonde si osserva frequentemente una particolarità: l'addestramento su immagini tende ad estrarre feature assimilabili a macchie di colore o filtri di Gabor nel primo layer convoluzionale. Questi filtri si manifestano con una tale rigorosità che la loro assenza indurrebbe a sospettare errori nell'implementazione algoritmica o un'errata calibrazione degli iperparametri. Dato che l'estrazione di queste feature di basso livello si verifica a prescindere dal dataset utilizzato e dalla specifica loss function, tali feature vengono classificate come "generali". Al contrario, le feature elaborate nello strato terminale di una rete addestrata sono fortemente legate al dominio applicativo e al task di riferimento (ad esempio, in un layer di output Softmax finalizzato alla classificazione supervisionata, ogni unità neurale diventerà sensibile a una determinata classe). Per questo motivo, le feature di alto livello vengono definite "specifiche". Considerando che i livelli iniziali catturano pattern universali e quelli finali pattern altamente

specializzati, è evidente che all'interno della rete debba verificarsi una transizione tra questi due estremi [45].

L'utilità della conoscenza trasferita attraverso i pesi della rete è tanto maggiore quanto più questa è considerata generale. Affinché il TL sia efficace, emergono alcuni quesiti fondamentali da considerare: come quantificare il grado di generalità di un layer neurale e determinare se la transizione avviene su un singolo strato o gradualmente su più livelli, identificandone la posizione nell'architettura di rete. Operativamente, il TL sfrutta un modello addestrato su un dataset e un task di base per importare le feature apprese nell'architettura target, concepita per l'addestramento su un dataset e un task differente. L'efficacia di questa operazione è legata alla presenza di feature generali, ovvero adatte sia al task di base che a quello target. In scenari clinici in cui i campioni del dataset target sono spesso limitati, il TL si rivela una strategia conveniente per addestrare modelli complessi limitando il problema dell'*overfitting* [45].

Un contributo rilevante per l'analisi delle caratteristiche e delle criticità del TL è stato fornito dallo studio sperimentale di Yosinski et al. [45]. La metodologia proposta dagli autori prevede la scomposizione di un vasto dataset in due gruppi, indicati come task A e task B (es. divisione del dataset ImageNet in due set, ognuno con 500 classi equamente rappresentate). Su questi gruppi vengono addestrate due reti neurali convoluzionali distinte, denominate rispettivamente "baseA" e "baseB". Vengono poi strutturate quattro diverse configurazioni per valutare le dinamiche di trasferimento sul livello neurale *n*:

- Rete di controllo BnB: i primi n layer vengono copiati dalla rete baseB e "congelati" (i pesi restano inalterati durante l'allenamento). I livelli successivi vengono inizializzati in modo casuale e addestrati sul dataset B. Questo modello funge da baseline per le reti di trasferimento.
- Rete di trasferimento AnB: i primi n layer vengono copiati dalla rete baseA e congelati. I livelli successivi vengono inizializzati in modo casuale e addestrati sul dataset B. Se l'accuratezza finale del modello AnB eguaglia quella della rete originale baseB, è possibile dedurre che le feature estratte fino al layer n siano generali e compatibili con il task B. Al contrario, un decadimento delle prestazioni indica che le feature apprese sono specifiche per il dominio sorgente A.
- Rete di controllo con ottimizzazione BnB^+: come BnB, ma tutti i layer apprendono.
- Rete di trasferimento con ottimizzazione AnB^+: come AnB, ma tutti i layer apprendono.

I risultati dello studio, schematizzati graficamente in Fig. 9, permettono di decifrare in modo analitico le complesse dinamiche del TL.

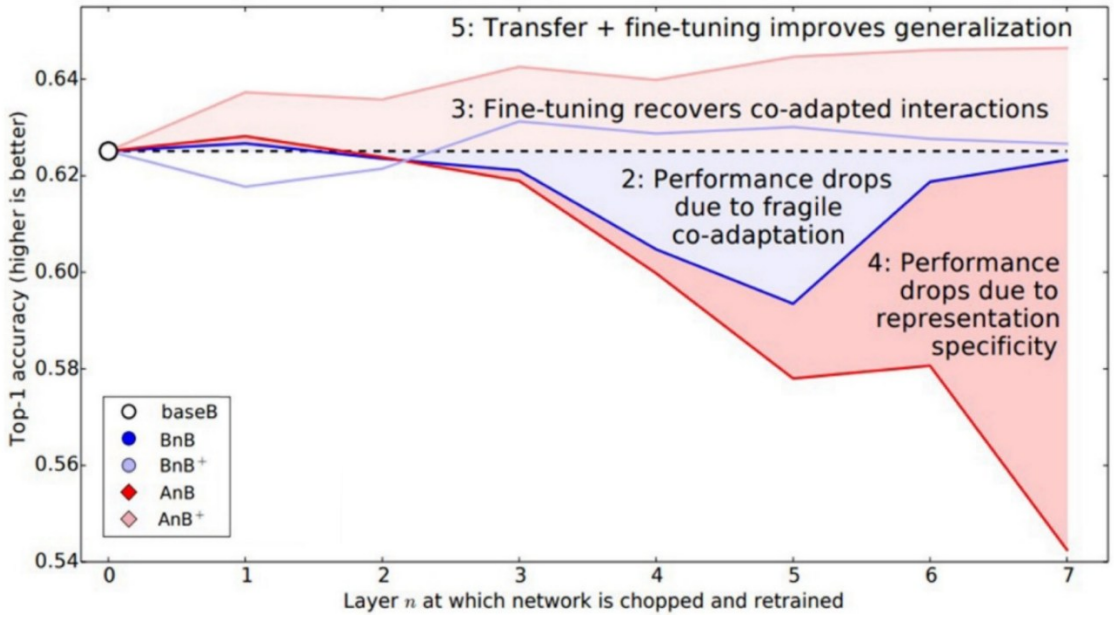


Fig. 9 Risultati dello studio citato. In alto: Ogni valore nella figura rappresenta l'accuratezza media sul set di validazione per una rete addestrata. Il cerchio bianco sopra n = 0 rappresenta l'accuratezza di baseB. I test sono stati eseguiti su quattro diverse suddivisioni casuali A/B [45].

Osservando le prestazioni del modello BnB (linea blu scuro), si nota un marcato calo prestazionale causato dalla frammentazione delle feature "co-adattate" sviluppatesi tra strati adiacenti. In pratica, bloccare i primi layer costringendo i successivi ad addestrarsi da zero, impedisce al modello di ripristinare le complesse e fragili interazioni che si erano create originariamente a ridosso del punto di divisione. Tale deterioramento viene tuttavia neutralizzato se si sbloccano i pesi e si consente l'apprendimento all'intera rete (come mostrato dalla azzurra BnB^{+}).

Analizzando invece la linea rosso scuro (configurazione AnB), è possibile mappare il grado di trasferibilità strato per strato. I livelli più superficiali mantengono un'ottima resa nel passaggio dal dominio A al dominio B, confermando la loro natura di estrattori di caratteristiche universali. Spingendosi verso gli strati più profondi, si registra invece un calo progressivo dell'accuratezza. Integrando questa osservazione con l'effetto di rottura mostrato dalla curva BnB, si deduce che questa decisa perdita di prestazioni sia imputabile alla combinazione dell'interruzione del co-adattamento strutturale tra i layer e della progressiva specializzazione delle feature al solo dominio sorgente.

L'ultimo aspetto rilevante emerge dall'analisi della linea rosso chiaro (variante AnB^{+}), che mostra come il trasferimento delle informazioni seguito dal FT dell'intera rete, consente di ottenere modelli che generalizzano meglio rispetto alle reti addestrate direttamente sul solo dataset target. Questo suggerisce che l'esposizione iniziale ai dati di addestramento sorgente lasci una sorta di "imprinting", che si traduce in un potenziamento stabile delle performance di generalizzazione [45].

Si conclude affermando che le dinamiche del trasferimento, nel caso in cui i pesi vengano "congelati" (senza FT), dipendono dalla profondità dei layer considerati:

- Universalità delle feature: I layer più superficiali estraggono caratteristiche di base che risultano altamente trasferibili, mantenendo un'ottima efficacia quasi inalterata indipendentemente dalla natura e dal dominio del nuovo task target.

- Rottura del co-adattamento: La separazione della rete in corrispondenza dei layer intermedi frammenta le feature fragili e "co-adattate", generando notevoli difficoltà di ottimizzazione. La discesa del gradiente non riesce a ripristinare le complesse interazioni formatesi originariamente tra gli strati adiacenti.
- Specializzazione delle feature: Trasferendo i layer più profondi, il degrado delle performance diventa dominato dalla marcata specificità delle rappresentazioni di alto livello, che risultano eccessivamente orientate al dominio originale per adattarsi al nuovo task.

A queste criticità si aggiunge l'impatto della distanza semantica tra i domini di partenza e di arrivo:

- Distanza tra i task: Il calo delle prestazioni diventa sensibilmente più marcato all'aumentare della diversità tra task sorgente e target.
- Confronto con l'inizializzazione casuale: Nonostante le perdite dovute alla distanza tra i task, l'inizializzazione con feature trasferite si dimostra costantemente superiore all'uso di pesi casuali.

Infine, l'approccio risulta particolarmente vantaggioso quando il trasferimento non è statico, ma prevede l'aggiornamento dei pesi:

- Miglioramento della generalizzazione: L'inizializzazione con le feature trasferite, seguita da cicli di FT sull'intera rete, produce modelli che generalizzano meglio rispetto alle reti addestrate da zero direttamente sul dataset target.
- Effetto di "imprinting": L'esposizione iniziale ai dati di addestramento sorgente lascia una traccia stabile che resiste anche dopo intensi cicli di FT, garantendo un incremento dell'accuratezza e affermandosi come una tecnica utile anche in presenza di ampi dataset target [45].

3.2.1 Transfer Learning per le GAN

L'impiego del TL assume una particolare rilevanza nelle architetture GAN allenate su dataset di dimensioni limitate, in quanto consente di aggirare le criticità associate al *data augmentation*. Quest'ultimo approccio, sebbene molto diffuso, presenta effetti collaterali significativi in ambito generativo. Il discriminatore inserisce erroneamente le immagini manipolate all'interno della distribuzione del dominio applicativo, mentre i modelli GAN sono tipicamente destinati a operare su immagini reali. Invece, il generatore impara a riprodurre artefatti indesiderati dovuti alle tecniche di *augmentation*, come distorsioni cromatiche o tagli innaturali. Di conseguenza, nelle implementazioni più recenti delle GAN ci si limita all'uso esclusivo di *cropping* e *flipping* per il *data augmentation*. Trasferire informazioni addizionali da modelli pre-allenati su database di larga scala risulta la strategia più robusta per compensare la scarsità dei dati a disposizione [46].

In ambito clinico, soprattutto nel caso specifico della generazione di sCT, gli algoritmi di DL sono vulnerabili al fenomeno del *data drift*. Tale deriva emerge quando si verificano discrepanze tra le condizioni su cui i modelli sono allenati e quelle riscontrate durante l'applicazione clinica. Sebbene le lievi fluttuazioni nella qualità delle immagini, imputabili al naturale degrado dell'hardware, possano essere efficacemente gestite tramite regolari procedure di controllo qualità e manutenzione proattiva, i costanti aggiornamenti software introdotti dai produttori (ad esempio per migliorare gli algoritmi di ricostruzione dell'immagine o per ridurre gli artefatti) generano

variazioni nella qualità delle immagini ottenendo dati che divergono da quelli utilizzati per addestrare la rete, creando quindi un *bias* che può rendere il modello inutilizzabile. In contesti simili, in cui il dominio applicativo resta vicino a quello originale, ricorrere al TL permette di ripristinare tempestivamente le performance originali con una coorte ridotta di pazienti. Infatti, nella pratica, il trasferimento della conoscenza evita di dover attendere i tempi necessari per raccogliere un set di dati di dimensioni paragonabili a quello di partenza, che richiede spesso diverse settimane. Solitamente un campione di 10-15 pazienti si rivela sufficiente per allenare e validare il modello, ricalibrandolo sul nuovo task [47].

Per approfondire le dinamiche del trasferimento di conoscenza all'interno delle GAN, lo studio condotto da Wang et al. [48] risulta utile per capire come valutare l'adattamento del dominio, scomponendo il problema in tre categorie:

- Configurazione di trasferimento per generatore e discriminatore: Le due reti antagoniste possono essere inizializzate con pesi casuali o con pesi ereditati da un modello sorgente. Le analisi sperimentali indicano che il trasferimento del discriminatore è più critico del trasferimento del generatore. Quest'ultimo risulta utile solo se il discriminatore è già stato trasferito; in caso contrario, compromette le prestazioni. Il trasferimento di entrambe le reti permette di ottenere il risultato migliore, oltre a rendere l'addestramento più stabile. Inoltre, l'impiego di GAN pre-addestrate accelera la convergenza, richiedendo un numero significativamente inferiore di iterazioni necessarie a raggiungere determinate performance rispetto a un addestramento eseguito interamente da zero.
- Dimensione del dataset target: La quantità di immagini disponibili in fase di addestramento impatta il livello di realismo raggiungibile dalle immagini generate, in particolare quando si lavora su immagini ad alta risoluzione. La carenza di campioni di dati rende inevitabilmente il training set meno diversificato e meno rappresentativo della distribuzione globale dei dati. Tuttavia, si stima che modelli GAN addestrati tramite TL richiedono dalle due alle cinque volte meno immagini per eguagliare i risultati di una GAN addestrata da zero. L'inizializzazione con modelli pre-addestrati aiuta a migliorare i risultati in ogni caso, in modo più significativo quanto più limitati sono i dati di destinazione. Questo vantaggio si osserva già dalle prime epoche, dove si raggiungono forme più nitide e definite, con una miglior rappresentazione dei dettagli, accelerando così il raggiungimento della sintesi di immagini realistiche.
- Selezione del modello pre-allenato: Per i classici compiti di discriminazione, la scelta del dominio sorgente ricade semplicemente su ImageNet per gli oggetti o su Places per gli scenari. Su questi modelli dedicati alla classificazione, il FT impara a ignorare i filtri non correlati al nuovo task. Tuttavia, questa semplice regola empirica non sembra applicarsi nel contesto più complesso del trasferimento su GAN. Si è osservato, controintuitivamente, che dataset non correlati al dominio target possono ottenere prestazioni migliori rispetto ad altri apparentemente più correlati ma meno densi. Questo paradosso potrebbe essere spiegato dal fatto che un'elevata densità di campionamento delle singole classi garantisce un potenziale di trasferimento più stabile, per cui si evince che una densità maggiore del dataset sorgente è preferibile alla diversità, anche qualora i domini non siano correlati. Emerge dunque l'esigenza di adottare un parametro quantitativo per stimare quale GAN pre-addestrata sia la più idonea per lo specifico target a disposizione. Le strategie più semplici prevedono la misura della distanza tra il dominio d'origine e quello di destinazione, oppure l'utilizzo di una versione più avanzata, che

include una valutazione della qualità del modello attraverso la stima della distanza tra le immagini reali del nuovo task e gli output generati dal modello pre-allenato.

Per quantificare formalmente la somiglianza tra due dataset, ci si avvale comunemente della Fréchet Inception Distance (FID), calcolata all'interno di un *embedding space* (uno spazio vettoriale creato dalle feature estratte dalle immagini). Questa proiezione vettoriale viene estratta impiegando una specifica rete convoluzionale con parametri fissi (tipicamente un modello Inception), utilizzata per estrarre le feature da un layer specifico. Assumendo che i dati estratti seguano una distribuzione normale multivariata, ne viene estratta la media e la matrice di covarianza, utilizzate per il calcolo della FID, secondo la seguente equazione:

$$FID(\mathcal{X}_1, \mathcal{X}_2) = ||\mu_1 - \mu_2||_2^2 + Tr\left(\Sigma_1 + \Sigma_2 - 2(\Sigma_1\Sigma_2)^{\frac{1}{2}}\right), \quad (9)$$

dove X_1 è il dataset delle immagini reali, mentre X_2 è l'insieme degli output generati. I vettori μ_1 e μ_2 definiscono la media delle feature estratte rispettivamente dal dataset originale X_1 e dal set di immagini generate X_2. Σ_1 è la matrice di covarianza delle immagini reali e Σ_2 è quella delle immagini generate. Infine, l'operatore algebrico Tr (traccia) restituisce la somma degli elementi disposti lungo la diagonale principale della matrice quadrata. Il diffuso utilizzo della metrica FID è legato alla sua efficienza computazionale e alla buona correlazione con la percezione visiva umana. Tuttavia, l'applicazione tra i dati generati e i dati reali, restituisce un indicatore approssimativo di idoneità. Essa non va interpretata come valutazione assoluta ma va integrata con altri fattori (ad esempio, la qualità del modello sorgente) per selezionare il modello più adatto al dominio target [48].

3.2.2 Strategie di Transfer Learning

Tra le metodologie più diffuse per implementare il TL spiccano quattro approcci, di cui i primi due rappresentano le macro-strategie di base e i successivi due le loro declinazioni operative. Il primo è il *Frozen weights*, ovvero il trasferimento senza possibilità di aggiornamento di determinati pesi. Questo può essere utilizzato per il mantenimento della capacità di estrazione delle feature (FX, *Feature Extraction*) di un modello pre-esistente. Nella pratica si congelano i pesi dei layer di codifica del generatore (encoder), garantendo che le feature estratte saranno le stesse del modello sorgente. Il secondo approccio è l'impiego del FT, che prevede il trasferimento del modello lasciando la possibilità di aggiornare tutti i pesi durante l'addestramento [47]. Segue una rappresentazione grafica del concetto in Fig. 10.

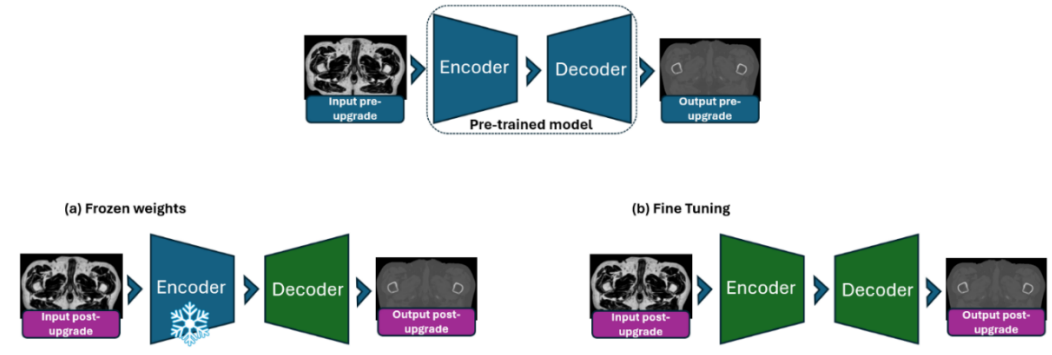


Fig. 10 Rappresentazione della strategia di apprendimento per trasferimento tramite (a) congelamento dei layer di estrazione delle feature e (b) fine tuning. In verde viene rappresentata la parte di rete soggetta all'aggiornamento dei pesi durante l'addestramento.

Un ulteriore approccio per garantire il totale mantenimento delle conoscenze apprese dai modelli pre-addestrati si fonda sull'aggiunta e messa a punto di nuovi layer neurali. In particolare, la rete sorgente viene importata congelandone i pesi. L'adattamento al nuovo task viene quindi affidato esclusivamente ai fully-connected layer aggiunti. Questi layer vengono inizializzati con valori casuali e sono accompagnati da una funzione di attivazione Softmax [49].

Un'ultima metodologia di TL prevede la copia dei primi n layer (con n che può variare da 1 al numero totale di layer presenti) da un modello pre-addestrato (i quali sono considerati layer generali). I layer rimanenti vengono di conseguenza inizializzati in modo casuale e addestrati sul task target (layer specifici per il task). In questo scenario, è possibile scegliere di retro-propagare gli errori derivanti dal nuovo compito anche alle feature copiate, al fine di ottimizzarle, oppure evitare ciò mantenendoli congelati. La scelta di ottimizzare o meno i layer trasferiti dipende dalla dimensione del dataset target a disposizione e dal numero di parametri contenuti in questi livelli. Qualora il dataset target sia di dimensioni particolarmente ridotte e il modello presenti un numero elevato di parametri, l'ottimizzazione globale innalza il rischio di incorrere nell'*overfitting*; per questa ragione, di norma si preferisce mantenere le feature congelate. Al contrario, se il dataset è sufficientemente ampio o il numero di parametri è contenuto, il problema dell'*overfitting* si riduce ed è possibile ottimizzare tutte le feature trasferite durante l'allenamento, migliorando l'adattamento del modello al nuovo task e quindi incrementando le prestazioni [45].

4 Obiettivo della tesi

Recentemente, l'azienda "TECNOLOGIE AVANZATE TA SRL" ha sviluppato un modello Pix2Pix per la generazione di sCT per i distretti addominale e pelvico, partendo da immagini RM a 3T pesate in T2 [50], al fine di supportare trattamenti radioterapeutici all'avanguardia. Nonostante la robustezza garantita dall'utilizzo di un ampio dataset multicentrico, la capacità di generalizzazione del modello non è sufficiente per sostenere l'utilizzo del modello su input acquisiti con un campo magnetico ridotto a 1.5T. Infatti, in questo campo di utilizzo, il modello va incontro a un drastico calo di prestazioni, tale da renderlo clinicamente inutilizzabile. Questo fenomeno di *data drift* è imputabile alle differenze fisiche e di imaging che intercorrono tra le acquisizioni RM effettuate a diverse intensità di campo magnetico.

Il passaggio alla RM a 1.5T comporta l'impiego di immagini caratterizzate da un minor SNR (*signal-to-noise ratio),* un ridotto CNR (*contrast-to-noise ratio*) e una peggiore risoluzione spaziale. Più in dettaglio, i sistemi a 3T sono in grado di generare segnali più puliti, con un SNR che può raddoppiare rispetto a uno scanner MR da 1.5T. L'impiego di un campo magnetico di maggiore intensità comporta la riduzione del tempo di decadimento del segnale trasversale dovuto alle interazioni spin-spin T_2^* con conseguenze dirette sul contrasto dell'immagine, sull'incremento degli effetti legati alle disomogeneità di campo e sull'aumento degli artefatti da suscettività [51], [52].

L'obiettivo della tesi consiste nel tempestivo adattamento del modello ai nuovi input di RM a 1.5T. A tal fine, si è optato per l'impiego del TL, tecnica che permette di ridurre la coorte di pazienti necessaria per allenare il modello. Tale approccio non solo accelera le tempistiche per il reperimento dei dati, ma riduce anche i tempi computazionali e migliora la convergenza della rete. Mentre il modello originale è stato allenato su 85 pazienti, qui si tenta di effettuare l'adattamento di dominio utilizzando una coorte significativamente ridotta di pazienti. Il dataset in questione, fornito dall'Ospedale Maggiore della Carità di Novara, comprende un gruppo di 20 pazienti, per i quali sono state rese disponibili le acquisizioni RM a 1.5T e le rispettive TC di pianificazione relative alla zona pelvica e addominale.

5 Materiali e Metodi

In questo capitolo si descriveranno i dati, le strategie e gli strumenti impiegati per lo sviluppo e la validazione di un modello Pix2Pix ottenuto tramite TL da un modello pre-addestrato, al fine di realizzare un adattamento di dominio, passando da una rete che riceve in input RM a 3T al nuovo modello che richiede come input RM a 1.5T.

Questa tesi è svolta in collaborazione con l'azienda Tecnologie Avanzate (TA), che ha fornito il supporto tecnico e l'accesso agli strumenti software, e con l'Ospedale Maggiore della Carità di Novara, che ha fornito i dati clinici necessari per l'addestramento della rete.

5.1 Dataset

La prima fase del lavoro ha riguardato l'analisi e il confronto delle caratteristiche dei dati a disposizione rispetto a quelli utilizzati per il modello sorgente. Come riassunto nella Tabella 2, si considerano pazienti di sesso maschile per i quali sono state rese disponibili le acquisizioni TC e RM relative al distretto pelvico-addominale. Il modello sorgente è stato acquisito con un approccio multicentrico (raccogliendo dati da tre diversi centri). Questo è un notevole vantaggio, in quanto durante il TL verrà tramandata la robustezza acquisita dalla varietà di protocolli e strumenti, consentendo l'utilizzo di un dataset target raccolto presso un solo centro, rendendo la raccolta dati più semplice e veloce. Il punto chiave per cui nasce questa tesi è la differenza nell'intensità di campo magnetico utilizzata dalla RM, nello specifico 3T per il modello originale e 1.5T per il nuovo modello. La divisione tra training-set e validation-set è stata effettuata mantenendo la medesima proporzione del modello sorgente (80%-20%); tuttavia, visto il limitato numero di pazienti target, questi sono stati suddivisi utilizzando il metodo k-fold, che verrà discusso più dettagliatamente in seguito. Si precisa che la divisione è stata effettuata sui singoli pazienti e non sulle singole slice, in modo da evitare qualsiasi possibile correlazione o somiglianza intra-paziente tra i dati proposti alla rete durante le fasi di training e validazione.

	Dataset sorgente	**Dataset target**
Numerosità	85	20
Sesso	Maschi	Maschi
Distretto anatomico	Pelvi-addome	Pelvi-addome
Centri clinici	3 centri clinici	Ospedale Maggiore della Carità di Novara
Imaging	-RM 3T pesata T2 -TC	-RM 1.5T pesata T2 -TC
Suddivisione training - validation	80% - 20%	80% - 20% (k-fold)

Tabella 2 Confronto tra le caratteristiche del dataset sorgente e del dataset target

Dalla precedente tabella emerge la mancanza di alcune informazioni di dettaglio, come ad esempio ulteriori dati demografici ed antropometrici dei pazienti (età, altezza, indice di massa corporea, ecc.), in quanto, in seguito all'anonimizzazione dei dati, questi non sono disponibili. Allo stesso modo, non risultano disponibili i tempi trascorsi fra le acquisizioni TC e RM, introducendo una potenziale incertezza sulla precisione delle successive registrazioni. Tuttavia,

grazie all'analisi dei TAG DICOM delle acquisizioni presenti nel dataset target, è stato possibile reperire e riassumere ulteriori specifiche tecniche, riportate nella Tabella 3.

	MIN	MAX	Sesso	Zona	Posizione	Dispositivo	Forza Campo / Potenza	Matrice Immagine (Rows/Columns)	Pixel Spacing (Risoluzione X,Y)	Slice Thickness	Num slice	Dose / SAR	Bits Allocated / Stored
CT Pz1	-2048	1654	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.495 x 1.495 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz1	0	518				Philips Marlin	1.5 Tesla	512 x 512 pixel	0.7826 x 0.7826 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz2	-1000	1699	M	PELVIS	HFS	Philips Marlin	1.5 Tesla	560 x 560 pixel	0.7535 x 0.7535mm	2. 0 mm (Spacing 1 mm)	300	0.608 W/kg	16/12
MR Pz2	0	877				Philips Marlin	1.5 Tesla	560 x 560 pixel	0.7535 x 0.7535mm	2.0 mm (Spacing 1 mm)	300	0.608 W/kg	16/12
CT Pz3	-2048	2061	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.084 x 1.084 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz3	0	570				Philips Marlin	1.5 Tesla	640 x 640 pixel	0.6517 x 0.6517 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz4	-2048	1650	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.228 x 1.228 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz4	0	617				Philips Marlin	1.5 Tesla	672 x 672 pixel	0.769 x 0.769 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz5	-2048	1688	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.142 x 1.142 mm	1.0 mm	441	20.9 mGy	16/16
MR Pz5	0	643				Philips Marlin	1.5 Tesla	672 x 672 pixel	0.671 x 0.671 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz6	-2048	1723	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.084 x 1.084 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz6	0	921				Philips Marlin	1.5 Tesla	640 x 640 pixel	0.659 x 0.659 mm	2.0 mm (Spacing 1 mm)	300	0.608 W/kg	16/12
CT Pz7	-2048	1727	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.554 x 1.554 mm	2.0 mm	235	20.9 mGy	16/16
MR Pz7	0	1120		PROSTATE	FFS	SIEMENS Skyra	3 Tesla	768 x 768 pixel	0.221 x 0.221 mm	3.0 mm (Spacing 3 mm)	28	1.055 W/kg	16/12
CT Pz8	-2048	1688	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.340 x 1.340 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz8	0	595				Philips Marlin	1.5 Tesla	512 x 512 pixel	0.7826 x 0.7826 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz9	-2048	1690	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	1.393 x 1.393 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz9	0	704				Philips Marlin	1.5 Tesla	720 x 720 pixel	0.6313 x 0.6313 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz10	-2048	1662	M	PELVIS	HFS	TOSHIBA Acquilion/LB	120 kV	512 x 512 pixel	0.969 x 0.969 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz10	0	622				Philips Marlin	1.5 Tesla	672 x 672 pixel	0.630 x 0.630 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz11	-2048	1712	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.074 x 1.074 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz11	0	607				Philips Marlin	1.5 Tesla	576 x 576 pixel	0.7308 x 0.7308 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz12	-2048	1933	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	0.998 x 0.998 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz12	0	680				Philips Marlin	1.5 Tesla	704 x 704 pixel	0.6495 x 0.6495 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz13	-2048	1680	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.420 x 1.420 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz13	0	699				Philips Marlin	1.5 Tesla	704 x 704 pixel	0.6880 x 0.6880 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz14	-2048	1726	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.244 x 1.244 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz14	0	790				Philips Marlin	1.5 Tesla	720 x 720 pixel	0.6218 x 0.6218 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz15	-2048	1638	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.324 x 1.324 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz15	0	580				Philips Marlin	1.5 Tesla	640 x 640 pixel	0.664 x 0.664 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz16	-2048	1704	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.351 x 1.351 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz16	0	646				Philips Marlin	1.5 Tesla	512 x 512 pixel	0.7826 x 0.7826 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz17	-2048	3801	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.372 x 1.372 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz17	0	633				Philips Marlin	1.5 Tesla	640 x 640 pixel	0.6812 x 0.6812 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz18	-2048	1695	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.426 x 1.426 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz18	0	678				Philips Marlin	1.5 Tesla	512 x 512 pixel	0.7826 x 0.7826 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz19	-2048	1691	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.580 x 1.580 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz19	0	658				Philips Marlin	1.5 Tesla	640 x 640 pixel	0.681 x 0.681 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12
CT Pz20	-2048	1712	M	PELVIS	HFS	TOSHIBA Aquilion/LB	120 kV	512 x 512 pixel	1.361 x 1.361 mm	2.0 mm	221	20.9 mGy	16/16
MR Pz20	0	650				Philips Marlin	1.5 Tesla	704 x 704 pixel	0.628 x 0.628 mm	2.0 mm (Spacing 1 mm)	300	0.4522 W/kg	16/12

Tabella 3 Informazioni ricavate dai TAG DICOM delle acquisizioni presenti nel dataset target.

Da quest'ultima tabella è possibile identificare la presenza di due outliers all'interno della coorte di pazienti:

- Paziente 2: presenta una sCT al posto della consueta TC di pianificazione.
- Paziente 7: presenta una RM a 3T, acquisita con FOV ridotto.

L’utilizzo di questi casi anomali verrà discusso in seguito.

Le ulteriori informazioni tecniche presenti in Tabella 3 sono riassunte nel seguente elenco:

- Range di intensità: valori compresi tra 0 e 921 per le RM e tra -2048 e 3801 HU per le TC.
- Dispositivi per l’imaging: Philips Marlin per le RM e TOSHIBA Aquilion/LB per le TC.
- Dimensioni delle slice: da 512x512 pixel a 720x720 pixel per le RM e fisse a 512x512 pixel per le TC.
- Risoluzione spaziale delle slice: da 0.6218x0.6218 mm a 0.7826x0.7826 mm per le RM e da 0.969x0.969 mm a 1.580x1.580 mm per le TC.
- Numero di slice per volume: 300 per le RM e 221 per la maggior parte delle TC.
- Spessore della slice: 2 mm (con spacing di 1 mm) per le RM e 2 mm per la quasi totalità delle TC.
- Dose radiogena delle TC: 20.9 mGy.
- Bit allocati: 16 bit per entrambe le modalità di imaging.

I due outliers segnalati presentano alcune deviazioni da questi dati. Per la RM a 3T (paziente 7) è stata acquisita con una strumentazione differente e presenta un range di intensità più ampio, slice di dimensioni maggiori e una migliore risoluzione spaziale, oltre a presentare uno spessore maggiore e un minor numero totale di slice. Per quanto riguarda la sCT (paziente 2), le dimensioni, la risoluzione e lo spessore delle slice risultano invece identiche a quelle della rispettiva RM, a cui si unisce la ovvia assenza del dato relativo alla dose radiogena assorbita.

5.2 Registrazione RM-CT

Per addestrare una rete Pix2Pix alla generazione di sCT è fondamentale che RM e TC siano perfettamente allineate. Per soddisfare tale requisito è stato impiegato RayStation [53], un software dedicato alla pianificazione del trattamento radioterapico e dotato di varie funzionalità, tra cui il modulo *Patient modeling* che offre tutti gli strumenti utili al caso.

Nella pratica, gli step eseguiti per ogni paziente sono stati i seguenti:

1. Creazione della ROI esterna: consiste nella segmentazione automatica del corpo del paziente, ottenuta tramite thresholding e operazioni morfologiche. Le soglie adottate corrispondono ai valori proposti dal software: una soglia fissa a -250 HU per la TC e una soglia variabile calcolata automaticamente dall’algoritmo per le RM.

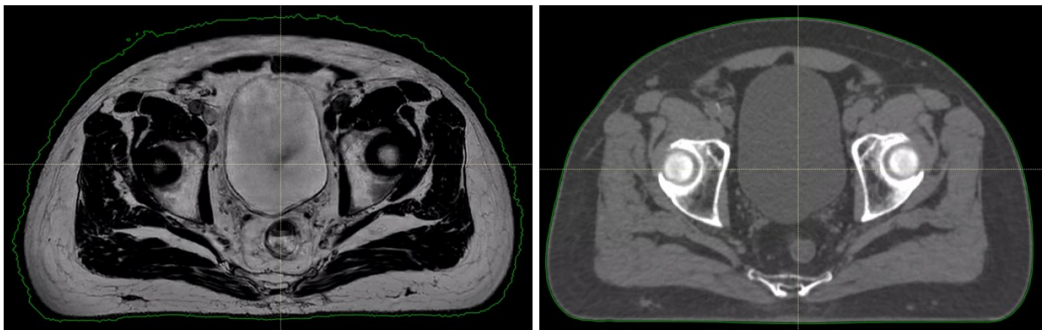


Fig. 11 A sinistra la segmentazione automatica della ROI esterna su RM, a destra su TC.

2. Registrazione rigida: il software sovrappone le due acquisizioni senza deformare i pixel. Questo avviene attraverso operazioni globali di traslazione e rotazione della RM, ottimizzate analizzando i livelli di grigio all'interno delle ROI esterne.
3. Segmentazione OAR con DL: il software integra una *fully convolutional neural network* basata su architettura U-Net che, operando su input tridimensionali, è in grado di segmentare autonomamente diversi organi [54]. Nello specifico, sono stati selezionati gli organi per i quali era disponibile la delineazione sia su RM che su TC, ovvero: tratto ano-rettale, vescica, bulbo penieno, prostata e vescicole seminali.

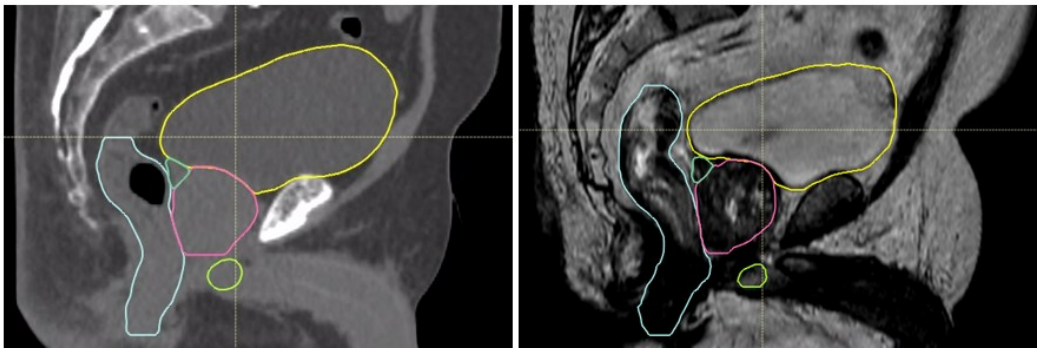


Fig. 12 Visualizzazione delle segmentazioni degli OAR su piano sagittale (paziente 1). A destra la RM e a sinistra la TC.

4. Registrazione deformabile ibrida guidata da strutture di controllo: la RM viene deformata elasticamente utilizzando una metrica di similarità. In questo caso è stata selezionata la *Mutual Information*, che rispetto al semplice coefficiente di correlazione, coglie anche le relazioni non lineari che intercorrono fra le due diverse modalità di imaging. A questo si aggiunge un termine di penalità che tiene in considerazione la sovrapposizione spaziale fra le segmentazioni selezionate. In questa fase sono state sfruttate come guida le ROI individuate dal DL per gli OAR, al fine di migliorare la corrispondenza finale dei due volumi.

5.3 Preprocessing

Al fine di ridurre la variabilità della qualità d'immagine e minimizzare la presenza di artefatti, garantendo così un modello più robusto, i dati sono stati sottoposti a una serie di operazioni di preprocessing mantenute identiche a quelle della rete di partenza, al fine di garantire maggior compatibilità dei pesi durante il TL.

Per le immagini RM:

- Ridimensionamento: si è impostato un formato finale di 512x512 pixel con una risoluzione spaziale dei pixel pari a 1.1719x1.1719 mm. Tale standardizzazione è stata ottenuta mediante un'interpolazione bi-cubica (implementata tramite la funzione *resize* della libreria *skimage* [55]).
- Normalizzazione: l'intensità di ogni slice è stata normalizzata rispetto al suo valore massimo.
- Mascheramento: per isolare l'anatomia del paziente si è applicata una prima sogliatura, considerando esclusivamente i pixel con un'intensità superiore al 10% del valore massimo della slice. In seguito, per rifinire la maschera, sono state applicate operazioni

morfologiche, partendo da una dilatazione con un disco di raggio pari a 5 pixel per poi selezionare la componente connessa più grande ed erodere con il medesimo disco. Il tutto completato dal riempimento di eventuali buchi (*hole filling*). Questa sequenza di operazioni (realizzata avvalendosi della libreria *ndimage* [56]) assicura maggiore robustezza del processo su immagini particolarmente rumorose, ottenendo sempre un'unica maschera piena.

- Histogram matching: questa tecnica consiste nel modificare il contrasto del corpo in modo da far coincidere il suo istogramma con quello proveniente da un riferimento. L'obiettivo di tale operazione è quello di uniformare gli istogrammi dei vari pazienti, standardizzando la distribuzione dei livelli di grigio. Per escludere lo sfondo dal calcolo, ci si è avvalsi della maschera precedentemente discussa. Il target di riferimento è stato definito calcolando l'istogramma medio dei migliori casi clinici appartenenti alla coorte originale di acquisizioni RM a 3T. La scelta di non calcolare il target sulle acquisizioni a 1.5T è duplice: da un lato queste ultime presentano istogrammi tendenzialmente più rumorosi; dall'altro, l'allineamento ai profili di contrasto del modello sorgente permette di minimizzare il divario fra gli input sorgente e gli input target (*domain gap*), accelerando l'apprendimento.

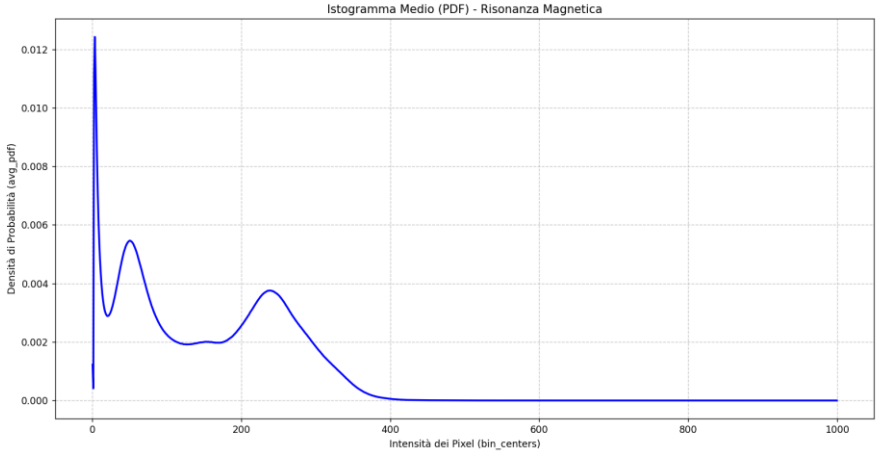


Fig. 13 Istogramma (PDF, probability density function) utilizzato per l'histogram matching

- Background: i pixel appartenenti allo sfondo, esclusi dalla maschera, sono stati uniformati ad un unico valore pari al minimo valore presente all'interno del corpo del paziente.
- Conversione a 16 bit.

Per le immagini TC:

- Ridimensionamento: è stato adottato il medesimo ridimensionamento utilizzato per le RM.
- Mascheramento: sfruttando la medesima libreria utilizzata per la RM, sono state condotte operazioni mirate a isolare il corpo del paziente. Nel dettaglio, è stata applicata una sogliatura iniziale per i valori superiori a -700 HU. Successivamente, la maschera è stata rifinita applicando un'erosione con un disco di raggio 3 pixel, seguita da un'aperura con lo stesso elemento strutturale e dalla chiusura dei buchi. Infine, per eliminare la porzione relativa al lettino, questa maschera TC è stata intersecata con la maschera della rispettiva RM previamente dilatata con un disco di raggio 3 pixel.

- Clipping: il range di intensità dell'intera immagine è stato saturato nell'intervallo compreso tra -1000 e 2000 HU, eliminando così eventuali artefatti o valori anomali generati dal rumore.
- Background: lo sfondo, isolato dalla maschera, è stato impostato a un valore di -1000 HU.
- Conversione a 16 bit.

Al termine di queste procedure di preprocessing, le singole slice di ciascuna coppia RM-TC sono state concatenate, generando un'unica immagine di dimensioni 1024x512 pixel.

Queste operazioni preparatorie vengono applicate a qualsiasi immagine fornita alla rete, sia in fase di addestramento che di inferenza.

5.4 Architettura e configurazione GAN

Al fine di poter inizializzare la rete con i pesi del modello sorgente, è necessario mantenere la stessa architettura originaria. In questo caso è stata riutilizzata l'architettura Pix2Pix resa disponibile da Isola et al. [40], una cGAN che implementa un generatore di tipo U-Net 256 e un discriminatore PatchGAN. Questa rete consente di effettuare la traduzione di immagini (*image-to-image translation*) da un dominio all'altro.

Il meccanismo di condizionamento prevede che il rumore stocastico venga introdotto attraverso l'impiego di layer di dropout all'interno del generatore, garantendo in questo modo la minima variabilità necessaria per evitare che la rete collassi su output deterministici.

La loss function complessiva da ottimizzare è definita come segue:

$$GAN_{loss} + \lambda_{L1} * \left|\left|G(A) - B\right|\right|_1 \tag{10}$$

I parametri per l'addestramento della rete sono elencati nella Tabella 4:

β_1	0.5	Termine di momentum per l'ottimizzatore Adam.
Learning rate (LR)	0.0002	Tasso di apprendimento per l'ottimizzatore Adam.
n_epochs	50	Numero di epoche eseguite con il LR iniziale.
n_epochs_decay	100	Numero di epoche impiegate per far decadere il LR a zero.
Gan mode	lsgan	Tipo di funzione obiettivo della GAN. Opzioni: [vanilla | lsgan | wgangp].
Pool size	50	Dimensione del buffer che memorizza le immagini generate in precedenza.
LR policy	linear	Politica di gestione del decadimento del LR. Opzioni: [linear | step | plateau | cosine].
LR decay iters	50	Numero di iterazioni dopo le quali il LR viene moltiplicato per un fattore gamma (utilizzato per la policy a step).
Masked loss	True	Attivazione della loss mascherata durante il training per evitare di calcolare l'errore sui pixel appartenenti allo sfondo.
λ_{L1}	100	Peso assegnato alla L1 loss.

dropout	0.5	Probabilità di azzeramento dei pesi nei layer di dropout.
norm	instance	Tipo di layer di normalizzazione utilizzato. Opzioni: [batch | instance | none]. La *Instance Normalization* normalizza i dati per ogni singola immagine in modo indipendente.

Tabella 4 Parametri di addestramento.

5.4.1 Generatore

Il generatore impiegato è basato sull'architettura U-Net 256, la quale presenta 54.403 milioni di parametri e una profondità pari a 8 livelli e si avvale di blocchi convoluzionali con *kernel* di dimensione 4x4, *stride* pari a 2 e *padding* pari a 1. Questa specifica combinazione dimezza la risoluzione spaziale a ogni passaggio lungo i livelli dell'encoder, per poi raddoppiarla progressivamente all'interno del decoder tramite operazioni di convoluzione trasposta. Il numero iniziale di filtri è pari a 64, mentre la propagazione dell'errore (*backpropagation*) è limitata esclusivamente ai pixel del corpo grazie all'utilizzo della maschera; ciò evita che la vasta preponderanza di pixel appartenenti allo sfondo sbilanci il processo di apprendimento.

L'encoder ha la funzione di estrarre feature sempre più complesse man mano che si procede verso i layer più profondi, raddoppiando il numero di filtri fino al massimo impostato pari a 512. La funzione di non-linearità adottata è una LeakyReLU con pendenza pari a 0,2. Per la riduzione di dimensionalità non si utilizzano operazioni di pooling, in quanto la contrazione spaziale viene già eseguita direttamente dalle convoluzioni con stride 2.

Segue l'elenco dettagliato dei livelli dell'encoder:

- Input: immagine RM di dimensioni 512x512 pixel su singolo canale.
- Layer Down 1 (Outermost):
  - Operazione: Conv2d(1, 64).
  - Output: 256x256 pixel, 64 canali.
- Layer Down 2:
  - Operazione: LeakyReLU -> Conv2d(64, 128) -> InstanceNorm.
  - Output: 128x128 pixel, 128 canali.
- Layer Down 3:
  - Operazione: LeakyReLU -> Conv2d(128, 256) -> InstanceNorm.
  - Output: 64x64 pixel, 256 canali.
- Layer Down 4:
  - Operazione: LeakyReLU -> Conv2d(256, 512) -> InstanceNorm.
  - Output: 32x32 pixel, 512 canali.
- Layer Down 5:
  - Operazione: LeakyReLU -> Conv2d(512, 512) -> InstanceNorm.
  - Output: 16x16 pixel, 512 canali.
- Layer Down 6:
  - Operazione: LeakyReLU -> Conv2d(512, 512) -> InstanceNorm.
  - Output: 8x8 pixel, 512 canali.
- Layer Down 7:
  - Operazione: LeakyReLU -> Conv2d(512, 512) -> InstanceNorm.

    - Output: 4x4 pixel, 512 canali.

Si prosegue con il bottleneck (Innermost layer), il livello di massima compressione caratterizzato da:

- Operazione: LeakyReLU -> Conv2d(512, 512).
- Output: 2x2 pixel, 512 canali.

Superato il collo di bottiglia, ha inizio la fase di ricostruzione dell'immagine attraverso il decoder. La peculiarità di questa architettura risiede nelle Skip Connection: prima di applicare la convoluzione trasposta, il tensore in risalita viene concatenato lungo l'asse dei canali con il tensore prodotto dal rispettivo livello dell'encoder. Di conseguenza, la successiva convoluzione trasposta riceverà in ingresso un volume con il doppio dei canali.

Segue la descrizione dei livelli del decoder:

- Layer Up 8:
    - Operazione: LeakyReLU -> ConvTranspose2d(512, 512).
    - Output: 4x4 pixel, 512 canali.
- Layer Up 7:
    - Skip Connecction: Il tensore 4x4x512 appena generato viene concatenato con il tensore 4x4x512 proveniente dal Layer Down 7.
    - Operazione: LeakyReLU -> ConvTranspose2d(1024, 512) -> InstanceNorm -> Dropout.
    - Output: 8x8 pixel, 512 canali.
- Layer Up 6:
    - Skip Connection: Concatenazione con Layer Down 6.
    - Operazione: LeakyReLU -> ConvTranspose2d(1024, 512) -> InstanceNorm ->Dropout.
    - Output: 16x16 pixel, 512 canali.
- Layer Up 5:
    - Skip Connection: Concatenazione con Layer Down 5.
    - Operazione: LeakyReLU -> ConvTranspose2d(1024, 512) -> InstanceNorm -> Dropout.
    - Output: 32x32 pixel, 512 canali.
- Layer Up 4:
    - Skip Connection: Concatenazione con Layer Down 4.
    - Operazione: LeakyReLU -> ConvTranspose2d(1024, 256) -> InstanceNorm.
    - Output: 64x64 pixel, 256 canali.
- Layer Up 3:
    - Skip Connection: Concatenazione con Layer Down 3.
    - Operazione: LeakyReLU -> ConvTranspose2d(512, 128) -> InstanceNorm.
    - Output: 128x128 pixel, 128 canali.
- Layer Up 2:
    - Skip Connection: Concatenazione con Layer Down 2.
    - Operazione: LeakyReLU -> ConvTranspose2d(256, 64) -> InstanceNorm.
    - Output: 256x256 pixel, 64 canali.
- Layer Up 1 (Outermost):
    - Skip Connection: Concatenazione con Layer Down 1.

- Operazione: LeakyReLU -> ConvTranspose2d(128, 1) -> Tanh() (la funzione di attivazione Tanh mappa il valore di ogni pixel nel range compreso tra -1 e 1).
- Output: 512x512 pixel, 1 canale (sCT).

L'output finale della rete viene mantenuto nel range compreso fra -1 e 1 per il successivo invio al discriminatore. Parallelamente, ai fini del salvataggio su disco, l'immagine viene sottoposta ad una trasformazione lineare per essere riportata nel formato a 16 bit.

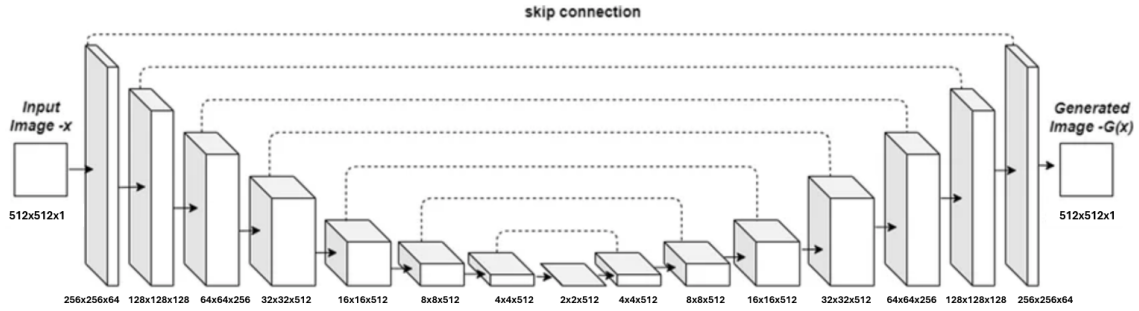


Fig. 14 Schema dell'architettura del generatore.

5.4.2 Discriminatore

Il modulo di discriminazione adotta un'architettura PatchGAN con una profondità di cinque layer convoluzionali, caratterizzati da kernel 4x4, stride pari a 2 o 1, e padding a 1. Il compito di questa rete, composta da 11.032 milioni di parametri, è valutare la veridicità dell'immagine proposta, fornendo un feedback al generatore. Per bilanciare la tendenza del generatore a produrre output eccessivamente sfocati, il discriminatore è progettato per focalizzarsi sui dettagli ad alta frequenza. A tale scopo, la valutazione non avviene sull'immagine globale, ma su un campo ricettivo ristretto ad una matrice di 70x70 pixel, che viene fatta scorrere sull'intera immagine.

L'input fornito è rappresentato dalle coppie TC-RM e sCT-RM. Il numero iniziale di filtri è impostato a 128. In questo modulo non viene applicato alcun mascheramento per escludere le patch in cui vi è solo sfondo, in modo da prevenire la comparsa di allucinazioni o artefatti generativi all'esterno dei confini corporei.

Segue la descrizione dei vari livelli della rete:

- Input: concatenazione (lungo l'asse dei canali) dell'immagine RM e dell'immagine TC (reale o sintetica). La dimensione iniziale è pari a 512x512 pixel per 2 canali.
- Layer 1:
  - Operazione: Conv2d(2, 128) con stride=2 -> LeakyReLU.
  - Output: 256x256 pixel, 128 canali.
- Layer 2:
  - Operazione: Conv2d(128, 256) con stride=2 -> InstanceNorm -> LeakyReLU.
  - Output: 128x128 pixel, 256 canali.
- Layer 3:
  - Operazione: Conv2d(256, 512) con stride=2 -> InstanceNorm -> LeakyReLU.
  - Output: 64x64 pixel, 512 canali.
- Layer 4:
  - Operazione: Conv2d(512, 1024) con stride=1 -> InstanceNorm -> LeakyReLU.

    - Output: 63x63 pixel, 1024 canali.
- Layer 5 (Output Map):
    - Operazione: Conv2d(1024, 1) con stride=1.
    - Output: 62x62 pixel, 1 canale.

L'output finale contiene quindi 3844 predizioni appartenenti all'analisi di ciascuna patch. Mediando questi valori si ottiene la valutazione generale dell'immagine come reale o falsa.

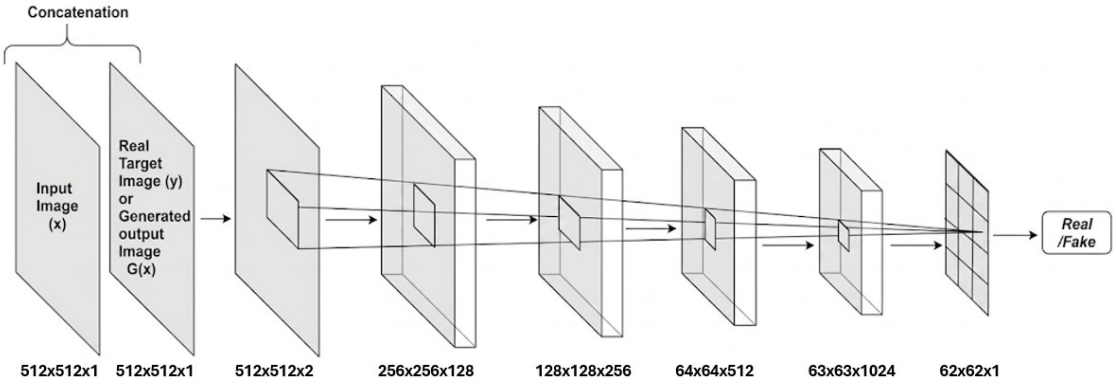


Fig. 15 Schema dell'architettura del discriminatore.

5.5 Metriche

Per ottenere una valutazione quantitativa delle diverse configurazioni e prove effettuate, sono state impiegate le medesime metriche adottate nel modello sorgente, data la vicinanza dei due task, questo approccio consente di disporre di un solido riferimento con cui confrontare le prestazioni ottenute. Nello specifico, la valutazione si basa su quattro delle metriche precedentemente introdotte nella Tabella 1:

- MAE: fondamentale per quantificare l'errore medio assoluto tra le TC reali e quelle sintetiche.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |sCT_i - CT_i| \tag{11}$$

- RMSE: rispetto al MAE, questa metrica penalizza maggiormente gli scarti più elevati, in quanto gli errori vengono elevati al quadrato prima di calcolarne la media. Ciò permette di tenere in considerazione discrepanze significative localizzate in pochi voxel, le quali rischierebbero di essere mascherate all'interno di un basso errore medio complessivo.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (sCT_i - CT_i)^2} \tag{12}$$

- PSNR: misura il rapporto tra l'intensità massima del segnale e il rumore introdotto; valori elevati indicano una migliore qualità dell'immagine e una minore distorsione [52]. Questa metrica permette quindi di monitorare il rumore introdotto dalla rete a causa di incertezze e bias.

$$PSNR = 10\log_{10}\left(\frac{Q^2}{MSE}\right) \tag{13}$$

- SSIM: rappresenta un indice maggiormente allineato alla percezione visiva umana, in quanto valuta la somiglianza delle immagini tenendo conto della sensibilità del nostro sistema visivo alle caratteristiche strutturali, alla luminosità e al contrasto [52]. Il vantaggio del suo impiego si traduce quindi nella possibilità di quantificare le distorsioni introdotte durante la generazione delle immagini [33].

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

5.6 Data cleaning

La pulizia dei dati rappresenta uno dei passaggi fondamentali per massimizzare le prestazioni raggiungibili. L'utilizzo di dati corrotti, affetti da artefatti o che presentano anomalie distanti dalle reali casistiche cliniche, introduce bias nella rete, determinando un conseguente calo delle prestazioni. Spesso, un dataset completo e non revisionato risulta meno performante rispetto alla sua versione ripulita.

Come prima operazione di *data cleaning,* è stata effettuata la ricerca e l'eliminazione delle slice poste agli estremi del volume, le quali presentavano solo una porzione del corpo o dei vuoti nell'immagine a causa della deformazione subita durante la registrazione (Esempio in Fig. 16).



Fig. 16 Esempio di slice scartata per via delle deformazioni alle estremità del volume.

Per quanto riguarda i pazienti outlier identificati in precedenza, è stato ipotizzato che il paziente 2 (sCT) avrebbe introdotto dei bias nella rete a causa degli artefatti presenti. Al contrario, si è supposto che il paziente 7 (RM a 3T con FOV ridotto) potesse arricchire il training con una maggiore variabilità anatomica e di qualità d'immagine, introducendo inoltre un maggior focus sui tessuti molli circostanti la prostata. Al fine di confermare tali aspettative, sono state eseguite tre prove sperimentali mirate a quantificare le variazioni prestazionali includendo o escludendo questi specifici outlier dalla coorte di addestramento.

Le tre prove sono state condotte mantenendo invariati i parametri utilizzati nel task sorgente e adottando il medesimo validation-set per garantire un confronto oggettivo. Per il training-set, invece, sono stati esplorati i seguenti approcci:

- Prova A: utilizzo di tutti i pazienti non inclusi nel validation-set.
- Prova B: medesima configurazione della prova A, ma con l'esclusione del paziente 2 dal training-set.

- Prova C: medesima configurazione della prova B, ma con l'ulteriore esclusione del paziente 7.

Il validation-set è stato estratto in modo casuale (escludendo a priori gli outlier) e risulta composto dai pazienti 1, 4, 8, 9.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova A	42,2	108,2	**0,891**	29,2
Prova B	**41,7**	**106,3**	**0,891**	**29,4**
Prova C	42,3	107,2	0,889	29,3

Tabella 5 Risultati di prova 1, prova 2 e prova 3

Come evincibile dalla Tabella 5, le ipotesi iniziali sono state confermate. Il setup più performante è risultato quello della prova B, che prevede l'inclusione del paziente 7 nel training-set ma non quella del paziente 2. Quest'ultimo, evidentemente, introduceva bias nella rete, favorendo la produzione artefatti; per tale ragione, il paziente 2 è stato eliminato dal dataset.

5.7 Divisione in Fold

In assenza di un validation-set sufficientemente ampio da poter rappresentare in modo esaustivo la popolazione d'interesse, si è ricorsi alla tecnica della *k-fold cross-validation*. Questa strategia permette di ottenere una stima affidabile delle metriche, iterando l'addestramento su diverse suddivisioni del dataset. Nella pratica, l'approccio prevede di ripartire casualmente i dati in *k* gruppi, o fold, di dimensioni approssimativamente uguali. A turno, un singolo fold viene isolato e impiegato come set di validazione, mentre il modello viene addestrato sui restanti *k-1* fold. La procedura viene ripetuta *k* volte, garantendo che ciascun sottoinsieme venga utilizzato esattamente una volta per la validazione [57].

Sebbene questo metodo comporti un maggiore onere computazionale, reso necessario dall'addestramento e dalla validazione di k modelli indipendenti, esso garantisce l'ottenimento di metriche più robuste e rappresentative dell'intera popolazione target, calcolate come media delle prestazioni registrate sui singoli validation-set. L'addestramento di modelli multipli apre inoltre la strada all'implementazione di tecniche di *ensemble* per la successiva fase di inferenza, ispirandosi al concetto di *bagging* (*Bootstrap Aggregating*). Similmente a quanto avviene con i campionamenti *bootstrap*, in cui si addestrano più modelli su sottoinsiemi di dati per poi aggregarne i risultati, è possibile combinare gli output delle *k* reti addestrate durante la *cross-validation*. Calcolare la media o la mediana delle predizioni restituite da ciascun modello come output finale consente di compensare le imperfezioni e gli artefatti tipici generati da una singola rete, riducendo la varianza complessiva e incrementando così la robustezza e la performance della generazione [57].

Nel presente lavoro di tesi, si è optato per una divisione casuale del dataset in 5 fold (k = 5). Un accorgimento metodologico fondamentale ha riguardato il paziente 7 (acquisizione RM a 3T) in quanto, si è scelto di escluderlo dai set di validazione e di includerlo unicamente nelle fasi di training. La motivazione risiede nel fatto che, sebbene le immagini 3T favoriscano l'addestramento, la loro presenza in fase di validazione rischierebbe di falsare le metriche, producendo stime non rappresentative del dominio target reale (1.5T).

La Tabella 6 descrive in dettaglio la composizione dei singoli fold.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Pazienti nel training-set	3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	1, 4, 5, 7, 8, 9, 10, 11, 12, 13, 15, 17, 18, 19, 20	1, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20	1, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 18, 19	1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 17, 18, 19, 20
Pazienti nel validation-set	1, 4, 8, 9	3, 6, 14, 16	5, 18, 19	10, 12, 17, 20	11, 13, 15

Tabella 6 Divisione in 5 fold.

6 Risultati

In questo capitolo vengono discusse le prove sperimentali condotte per l'implementazione del TL. Data l'elevata quantità di configurazione e strategie esplorabili, si è scelto di ottimizzare i tempi testando queste ultime esclusivamente sul primo fold, per poi validare solo le prove più rilevanti sui restanti fold. Le epoche ottimali sono state selezionate ricercando il miglior compromesso tra le metriche di valutazione. Nella maggior parte dei casi, il picco prestazionale per ciascuna metrica si verifica in corrispondenza della medesima epoca; tuttavia, nei casi di parità, la valutazione è stata estesa includendo il 95° percentile per MAE e RMSE e il 5° percentile per SSIM e PSNR. I grafici relativi alle varie prove sono riportati nell'Appendice B.

6.1 Studio d'ablazione

Al fine di stabilire una baseline di riferimento per la selezione delle prove più rilevanti, si è inizialmente quantificato l'entità del *data drift* misurando il calo prestazionale del modello sorgente. Questo è stato ottenuto confrontando le prestazioni del modello sul validation-set sorgente con quelle sul validation-set target.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Validation-set sorgente	**43,3**	**102,0**	**0,871**	**29,7**
Validation-set target (fold 1)	46,1	111,8	0,865	28,9

Tabella 7 Valutazione crollo prestazionale del modello sorgente passando da input RM 3T a RM 1.5T.

I risultati riportati in Tabella 7 mostrano un evidente peggioramento che coinvolge tutte le metriche, definendo così la soglia di partenza.

La prova successiva è stata impiegata per quantificare i miglioramenti introdotti dal TL (prova 1) rispetto a un addestramento eseguito partendo da un'inizializzazione casuale dei pesi (prova 2). Per questa prima indagine è stata adottata la tecnica di TL più semplice, ossia il trasferimento di tutti i pesi e il relativo FT dell'intera rete.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 1	**41,7**	**106,3**	**0,891**	**29,4**
Prova 2	44,4	112,3	0,885	28,9

Tabella 8 Risultati delle prove 1 e 2 sul primo fold.

Dimostrata l'effettiva utilità del TL, lo studio è proseguito analizzando gli effetti del LR. È stata eseguita la prova 3 con un LR dimezzato e la prova 4 introducendo un *warm-up* del LR nelle prime 50 epoche. L'obiettivo era valutare se un approccio con una forza d'apprendimento meno aggressiva potesse giovare all'addestramento rendendolo più stabile. Come visibile nella seguente tabella, queste due strategie non risultano vantaggiose rispetto alla configurazione standard della prova 1.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 3	43,8	111,3	0,885	28,9
Prova 4	42,2	107,5	0,890	29,3

Tabella 9 Risultati delle prove 3 e 4 sul primo fold.

A partire da questo punto, è stato rimosso il decadimento del LR programmato per le ultime 100 epoche, poiché le prestazioni ottimali venivano sistematicamente raggiunte all'interno delle prime 50. Questo dato dimostra il vantaggio intrinseco del TL nell'accelerare la convergenza della rete e ridurre i tempi di allenamento.

La prova 5 si è concentrata sull'iperparametro λ_{L1}, il quale assegna un peso alla L1 loss rispetto alla GAN loss. Ipotizzando che la rete fosse già in grado di riconoscere e ricostruire le varie strutture anatomiche, si è supposto che il compito principale del nuovo task fosse l'adattamento al diverso rumore e contrasto dei nuovi input. Per questo motivo, si è tentato di aumentare il valore di λ_{L1} a 150, conferendo così maggiore importanza agli errori di intensità a livello del singolo pixel. Anche in questo caso, la modifica ha alterato l'equilibrio della rete, portando a un peggioramento dei risultati.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 5	42,3	107,5	**0,892**	29,3

Tabella 10 Risultati della prova 5 sul primo fold.

Entrando più nel dettaglio nelle strategie di TL, si è analizzato l'effetto del congelamento dei pesi a diverse profondità architetturali. In particolare, il freezing è stato effettuato in modo simmetrico su encoder e decoder tramite le seguenti prove:

- Prova 6: congelamento del livello più esterno del generatore.
- Prova 7: congelamento dei due livelli più esterni del generatore.
- Prova 8: congelamento dei tre livelli più esterni del generatore.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 6	41,7	106,2	**0,892**	**29,4**
Prova 7	**41,6**	**105,8**	0,890	**29,4**
Prova 8	43,3	109,5	0,884	29,1

Tabella 11 Risultati delle prove 6, 7 e 8 sul primo fold.

Come si può notare dalla Tabella 11, per il primo fold, la soluzione più conveniente è risultata quella di congelare i primi due layer.

Successivamente, si è voluto indagare se la ripresa dell'addestramento a partire dall'epoca migliore, accompagnata dallo sblocco dei primi layer, potesse favorire un addestramento più fine. L'obiettivo era forzare la rete a concentrarsi globalmente sull'ottimizzazione delle feature di alto livello, dedicando al contempo alcune epoche a un leggero riadattamento dei livelli superficiali. A tale scopo, l'addestramento della prova 7 è stato ripreso dall'epoca ottimale, sbloccando i layer secondo quattro diverse strategie:

- Prova 9: sblocco di tutti i layer e LR mantenuto costante.
- Prova 10: sblocco del secondo layer e avvio del decadimento del LR.
- Prova 11: sblocco di tutti i layer e avvio del decadimento del LR.
- Prova 12: sblocco del secondo layer e LR mantenuto costante.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 9	41,9	106.9	**0,891**	29,3
Prova 10	45,6	108,6	0,868	29,2
Prova 11	42,3	106,9	0,889	29,3
Prova 12	42,3	107,6	0,888	29,3

Tabella 12 Risultati delle prove 9, 10, 11 e 12 sul primo fold.

Non avendo prodotto alcun miglioramento tangibile, la strategia di sblocco progressivo dei layer è stata accantonata.

Un'ulteriore strategia di TL ha previsto il congelamento dell'intero encoder, al fine di preservare integralmente la capacità di estrazione delle feature apprese sul dominio sorgente. Questa tecnica, indagata tramite la prova 13, non è tuttavia riuscita a superare i risultati ottenuti con il semplice FT.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 13	42,8	108,7	0,888	29,1

Tabella 13 Risultati della prova 13 sul primo fold.

Fino a questo punto, il TL era stato applicato congiuntamente al generatore e al discriminatore. Replicando la metodologia di Wang et al. [48], si è proceduto a valutare l'efficacia di un trasferimento asimmetrico. Utilizzando la prova 1 come riferimento per il TL totale, sono stati configurati il trasferimento esclusivo del discriminatore con la prova 14 e quello esclusivo del generatore con la prova 15.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 14	45,9	115,4	0,880	28,6
Prova 15	**41,3**	**106,1**	**0,892**	**29,4**

Tabella 14 Risultati delle prove 14 e 15 sul primo fold.

Come prevedibile trasferire unicamente i pesi del discriminatore ha causato un peggioramento delle prestazioni. Mentre, al contrario delle aspettative, il trasferimento dei soli pesi del generatore si è rilevato più vantaggioso rispetto all'inizializzazione di entrambe le reti.

Come ultimo test, si sono volute combinare le due strategie che si erano rivelate maggiormente efficaci. Quindi la prova 16 è stata configurata applicando contemporaneamente il congelamento dei primi due layer e il TL sul solo generatore.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Prova 16	**41,2**	**105,6**	0,889	**29,5**

Tabella 15 Risultati della prova 16 sul primo fold.

Come visibile nella Tabella 15, la combinazione delle strategie ottimali si dimostra l'approccio più conveniente.

A questo punto dello studio, al fine di validare i risultati ottenuti su una coorte più ampia di pazienti e ottimizzare i tempi di implementazione, è stata evitata la ripetizione di tutte le 16 configurazioni. Sono state selezionate esclusivamente le prove che si erano distinte nel primo fold, procedendo con la loro validazione sui restanti fold. Sono state quindi estese la prova di baseline (modello sorgente su validation-set target) e le prove 1, 2, 6, 7, 8, 13, 14 e 15. Nel caso in cui si dovessero confermare i medesimi trend prestazionali, si valuterà l'inclusione della prova 16. Nelle immagini seguenti vengono riportati i risultati della *k-validation* a confronto con le prestazioni del modello sorgente.



Fig. 17 Risultati del k-validation. SM -> 3T indica le prestazioni del modello sorgente sul validation-set sorgente. SM -> 1.5T indica le prestazioni del modello sorgente mediate sui 5 fold del task target. SM -> 3T indica la soglia di prestazioni ideali mentre il distacco fra SM -> 3T e SM -> 1.5T indica il crollo delle prestazioni per il data drift.

Come si evince dai grafici in Fig. 17, calcolando la media delle prestazioni su tutti e cinque i fold, il congelamento dei primi layer non risulta più vantaggioso (confronto tra la prova 1 e le prove 6, 7 e 8). Mentre, si conferma l'efficacia del trasferimento limitato al solo generatore (confronto tra la prova 1 e le prove 14 e 15).

6.2 Conversione a 2.5D

Al fine di ottenere una maggiore continuità spaziale fra le slice, si è proceduto alla conversione dell'architettura da 2D a 2.5D mediante l'introduzione di due canali aggiuntivi. In questo modo, il modello non analizza una singola slice isolata, ma prende in considerazione contemporaneamente anche la slice adiacente superiore e quella inferiore, integrando così informazioni contestuali lungo l'asse longitudinale del paziente.

L'approccio basato su TL è stato mantenuto, inizializzando la rete con i pesi corrispondenti all'epoca ottimale della prova 15 per ciascun fold. Per adattare il trasferimento dei pesi alla nuova dimensionalità dell'input, si è preso ispirazione da due metodologie utilizzate in letteratura per la transizione da architetture 2D a 3D [58]:

- Centering inflation: i pesi pre-addestrati vengono trasferiti esclusivamente sul canale centrale, mentre i canali laterali vengono inizializzati a zero.
- Average inflation: i pesi pre-addestrati vengono replicati su tutti i canali, dividendo il loro valore originario per il numero totale di canali (in questo caso 3).

Oltre a testare queste due tecniche, si è voluto verificare se la conversione al 2.5D risultasse vantaggiosa anche per il discriminatore, per il quale si continua a non applicare il TL. Sono state pertanto configurate quattro prove sul primo fold per determinare la strategia migliore; i risultati sono riportati nella tabella seguente.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
Centering inflation + discriminatore 2D	**41,4**	106,4	**0,893**	**29,4**
Centering inflation + discriminatore 2.5D	**41,4**	**105,8**	**0,893**	**29,4**
Average inflation + discriminatore 2D	42,2	107,2	0.891	29,3
Average inflation + discriminatore 2.5D	41,9	106,3	0.892	**29,4**

Tabella 16 Risultati delle prove per la selezione della miglior strategia per la conversione della rete a 2.5D.

I dati confermano che l'adozione della centering inflation, unita all'estensione dell'architettura 2.5D anche al discriminatore, rappresenta la soluzione più robusta. Infatti, indipendentemente dalla natura del discriminatore, la centering inflation si attesta come la strategia di inizializzazione più efficace; mentre, a parità di metodo di trasferimento, l'impiego di un discriminatore 2.5D garantisce un incremento delle prestazioni. La transizione verso un generatore e un discriminatore 2.5D ha comportato un lieve aumento del numero di parametri addestrabili, i quali hanno raggiunto rispettivamente quota 54.406 milioni e 11.037 milioni.

In questa fase avanzata dello studio, per consentire un'ispezione più fine del processo di convergenza, la frequenza di salvataggio dei pesi è stata incrementata, passando da un salvataggio ogni due epoche ad ogni epoca. La configurazione ottimale è stata quindi applicata a tutti i fold. La Tabella 17 illustra i risultati finali mediati, includendo anche le metriche relative al training-set, al fine di poter iniziare a valutare l'*overfitting* della rete.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
validation-set	**39,0**	**100,9**	**0,896**	**29,8**
training-set	23,1	52,5	0,935	35,5

Tabella 17 Risultati del k-validation dopo conversione a rete 2.5D.

Sebbene, in media, i risultati siano migliorati per tutte le metriche, un'analisi più dettagliata dei singoli fold rivela che, contrariamente alle aspettative, non tutte le suddivisioni hanno tratto pieno beneficio dalla conversione al 2.5D. La Tabella 18 riporta la variazione relativa delle metriche nel passaggio dall'architettura 2D a quella 2.5D per ciascun fold.

	MAE (%)	RMSE (%)	SSIM (%)	PSNR (%)
Fold 1	0,00	0,56	-0,11	0,00
Fold 2	0,00	-0,09	0,00	0,00
Fold 3	1,15	0,86	0,00	-0,33
Fold 4	1,20	1,24	0,00	-0,34
Fold 5	0,54	-0,52	-0,11	0,00

Tabella 18 Confronto in termini relativi fra rete 2D e 2.5D per ogni fold.

Questa ambiguità prestazionale può essere giustificata osservando la dinamica dell'addestramento: per tre fold su cinque, l'epoca con i risultati migliori si colloca entro le primissime tre iterazioni, rendendo improbabile il corretto assestamento dei pesi sui canali laterali. Ciò suggerisce che, in questi specifici casi, la rete rimanga bloccata in un punto di ottimo

locale, raggiunto dalla precedente rete 2D, dal quale non riesce a smuoversi, probabilmente a causa della limitata quantità di dati a disposizione.

6.3 Data augmentation

L'ultima fase di ottimizzazione ha previsto l'applicazione di tecniche di *data augmentation*, impiegate per incrementare artificialmente la variabilità del dataset, migliorare le performance e la capacità di generalizzazione dei modelli, nonché per superare lo stallo prematuro dell'apprendimento della rete riscontrato in seguito alla conversione in 2.5D. Le trasformazioni geometriche applicate alle immagini di ciascun paziente sono le seguenti:

- Flip orizzontale: consente di migliorare la riproduzione delle strutture anatomiche che risultano orizzontalmente speculari all'interno delle acquisizioni sul piano assiale del paziente.
- Stretching verticale combinato a rotazione casuale di ± 10°: incrementa la variabilità delle conformazioni anatomiche e conferisce alla rete maggiore flessibilità e robustezza rispetto a lievi disallineamenti del paziente sul lettino di trattamento.
- Stretching orizzontale combinato a rotazione casuale di ± 10°: persegue il medesimo scopo del punto precedente, applicando tuttavia la deformazione lungo l'asse ortogonale.
- Estrazione casuale di patch: consiste nell'estrarre una tra le 6 possibili patch di dimensioni 128x128 pixel illustrate in Fig. 18. Questa operazione costringe il discriminatore ad adottare un approccio multiscala, inducendolo a fornire valutazioni basate anche su porzioni ridotte dell'immagine. Ciò impedisce che la valutazione finale trascuri errori generativi localizzati compensandoli con l'elevata accuratezza del resto dell'immagine.

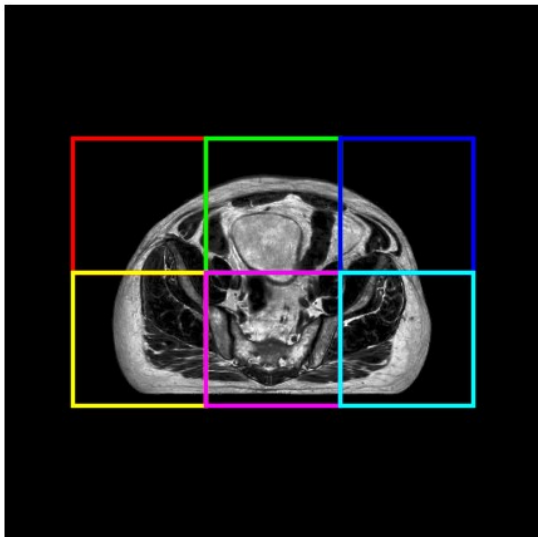


Fig. 18 Patch estraibili durante il data augmentation.

Replicando la prova di conversione a 2.5D sull'intero dataset aumentato, si riscontrano risultati nettamente superiori all'interno dei singoli fold. In particolare, i processi di addestramento non si arrestano prematuramente durante le primissime epoche, ma progrediscono in maniera funzionale. A conferma di ciò, la Tabella 19 ripropone il confronto prestazionale in termini relativi rispetto alla rete 2D di partenza.

	MAE (%)	RMSE (%)	SSIM (%)	PSNR (%)
Fold 1	2,91	3,58	-0,56	-1,02
Fold 2	3,19	4,33	-0,45	-1,37
Fold 3	2,58	1,29	-0,33	-0,33
Fold 4	5,30	6,37	-0,45	-2,38
Fold 5	4,02	3,23	-0,55	-0,99

Tabella 19 Confronto in termini relativi fra rete 2D e 2.5D combinata al data augmentation per ogni fold.

Infine, la Tabella 20 riporta le prestazioni finali ottenute mediando le metriche restituite dai cinque modelli.

	MAE (HU)	RMSE (HU)	SSIM	PSNR (dB)
validation-set	**37,8**	**97,4**	**0,900**	**30,1**
training-set	19,8	42,0	0,946	37,4

Tabella 20 Risultati del k-validation dopo conversione a rete 2.5D combinata al data augmentation.

7 Discussione dei risultati

Analizzando i risultati ottenuti, è possibile trarre diverse osservazioni rilevanti. In primo luogo, dai grafici delle prove 1 e 2 (riportati in Appendice B), si nota come il TL permetta di raggiungere l'epoca ottimale in un minor numero di iterazioni, ottenendo risultati superiori e riducendo il divario prestazionale tra il training-set e il validation-set. Inoltre, osservando l'andamento della L_1 loss e i risultati delle prime epoche, emerge chiaramente come il punto di partenza dell'allenamento con il TL sia nettamente più vantaggioso rispetto all'inizializzazione con pesi casuali.

Per quanto concerne la modifica degli iperparametri, l'analisi indica che, per TL su domini molto simili, è preferibile mantenere inalterati i parametri di addestramento della rete originale. Eventuali variazioni hanno prodotto non solo un peggioramento delle metriche, ma anche una marcata instabilità del processo, con violente fluttuazioni visibili anche sul training-set. Fa eccezione la rimozione del decadimento del LR; infatti, il suo mantenimento in questo nuovo task provoca un peggioramento delle performance persino sul training-set.

L'estensione dei test tramite *k-fold cross-validation* (Fig. 17), ha permesso di rivalutare in modo robusto le conclusioni emerse sul singolo fold. Sebbene inizialmente apparisse conveniente congelare i primi due layer del generatore, tale approccio non si è confermato valido per gli altri fold. Si evince, pertanto, che per questo specifico scenario è preferibile non congelare alcun livello della rete. In accordo con lo studio di Yosinski et al. [45], questo risultato suggerisce che, data la quantità di dati a disposizione, il rischio di *overfitting* è ampiamente compensato dal vantaggio derivante dal totale adattamento dei pesi al nuovo dominio. Per ragioni analoghe, e in linea con quanto riscontrato da Galapon Jr et al. [47], il congelamento dell'intero encoder non ha apportato miglioramenti ai risultati finali.

Sempre dai grafici di Fig. 17 si deduce che la strategia migliore consiste nell'applicare un TL totale sul solo generatore seguito da un FT completo dei suoi pesi, inizializzando invece il discriminatore in modo casuale. Quest'ultima evidenza contrasta con le conclusioni di Wang et al. [48], si ipotizza che la giustificazione risieda nell'utilizzo di un discriminatore caratterizzato da pochi layer, applicato a un dominio target estremamente affine a quello sorgente. In questo scenario, trasferire i pesi pre-addestrati su una rete avversaria già strutturalmente "semplice" le conferisce un vantaggio sproporzionato rispetto al generatore, essendo già capace di classificare le immagini reali e sintetiche con troppa facilità. Tale sbilanciamento impedirebbe al generatore di ingannare il discriminatore, privandolo dei feedback fondamentali per guidare la propria ottimizzazione.

Un'ulteriore osservazione di rilievo (Fig. 17) riguarda il fatto che il nuovo modello, pur elaborando acquisizioni RM a 1.5T (qualitativamente inferiori alle RM a 3T), non solo ha compensato il calo prestazionale indotto dal data drift, ma ha addirittura superato le performance del modello sorgente. Questo inaspettato incremento prestazionale è verosimilmente attribuito a due fattori:

- Esposizione a nuovi pazienti: operando su domini simili, il modello ha di fatto prolungato il proprio addestramento, beneficiando di nuove morfologie anatomiche e di un'ulteriore componente di varianza indotta dalla diversa intensità del campo magnetico. A differenza di un TL applicato su domini eterogenei, in questo caso la rete ha continuato ad assimilare informazioni riguardanti lo stesso compito e sul medesimo distretto corporeo.

- Maggiore accuratezza nella registrazione: un notevole vantaggio derivante dall'adozione del TL in questo scenario risiede nel ridotto numero di pazienti necessari, il che ha consentito di dedicare maggiore tempo ed attenzione al processo di registrazione di ciascun soggetto. Nel modello sorgente, basato su un'ampia coorte, si era optato per lo scarto delle slice in cui la sovrapposizione degli OAR a fine registrazione risultava insoddisfacente, impoverendo il training-set di importanti informazioni sui tessuti molli. In questo studio, grazie al campione ridotto a 20 pazienti, è stato possibile revisionare manualmente le segmentazioni imprecise degli OAR. Ciò ha garantito un netto miglioramento degli allineamenti spaziali e ha permesso di preservare integralmente le fette contenenti i tessuti molli, un prerequisito rivelatosi cruciale anche per la successiva conversione dell'architettura in 2.5D. Poiché tale processo di correzione manuale avrebbe richiesto tempistiche proibitive se applicato al dataset sorgente di 85 pazienti, il TL ha indirettamente permesso di utilizzare meno dati per l'addestramento, rendendo di fatto sostenibile l'impiego di metodologie più laboriose e meticolose per l'ottimizzazione e la cura dei dati.

Analizzando il passaggio alla configurazione 2.5D (Tabella 18), sebbene si registri un miglioramento medio su tutte le metriche, non tutti i fold hanno tratto pieno beneficio da questa tecnica. Questo fenomeno, in apparenza controintuitivo considerando il valore aggiunto dalle informazioni spaziali tridimensionali, trova una plausibile spiegazione nella dinamica di addestramento: per tre fold su cinque, le prestazioni massime si sono verificate entro le prime tre iterazioni, indicando che la rete era rimasta intrappolata in un punto di ottimo locale (ereditato dai pesi della rete 2D), dal quale faceva fatica ad allontanarsi a causa della limitata mole di dati. Dunque, si ipotizza che la rete tendeva ad accontentarsi delle predizioni basate sul canale centrale (inizializzato con centering inflation), fallendo nel calibrare i pesi sui canali laterali.

La *data augmentation* ha permesso di superare questa inerzia negli esperimenti condotti. Le trasformazioni geometriche applicate hanno garantito un beneficio netto su tutti i fold (Tabella 19), introducendo un miglioramento relativo medio rispetto alla rete 2D pari a 3,8% per l'RMSE e al 3,6% per il MAE. Confrontando la Tabella 17 e la Tabella 20, si osserva che la *data augmentation*, in questo caso particolare, più che diminuire il distacco prestazionale tra training-set e validation-set, ha permesso alla rete di adattarsi più efficacemente alla nuova configurazione con 3 canali di input. Ciò ha consentito un addestramento più efficiente, scongiurando l'arresto prematuro della convergenza alle primissime epoche, e si è tradotto nel raggiungimento di iterazioni ottimali caratterizzate da prestazioni decisamente superiori sia sul training-set che sul validation-set.

L'analisi quantitativa si conclude con il raggiungimento e superamento delle performance originarie del modello sorgente per ogni metrica valutata. In termini relativi l'architettura finale ha conseguito una riduzione dell'errore pari al 12,8% per il MAE e al 4,5% per l'RMSE, congiuntamente ad un incremento del 3,4% per l'indice SSIM e del 1,3% per il PSNR.

A completamento dello studio, si propone un'analisi qualitativa delle predizioni. In Appendice C sono riportati i risultati generati dalle reti 2D, 2.5D e 2.5D combinata alla *data augmentation* per i pazienti appartenenti ai validation-set. Nello specifico, vengono illustrati sia i volumi sintetizzati a partire dalle RM originali, sia quelli prodotti dalle RM deformate; questa duplice visualizzazione consente di confrontare gli output sia con le RM non deformate (simulando lo scenario della pratica clinica reale), sia con le TC di riferimento. Valutando visivamente l'evoluzione delle immagini attraverso le tre configurazioni esplorate, si osserva un progressivo miglioramento nella definizione dei margini ossei, con particolar evidenza a carico dei femori e

delle creste iliache (es. pazienti 1 e 8). Contestualmente, si rileva un incremento della risoluzione strutturale delle vertebre (es. pazienti 3, 10 e 17), una morfologia della vescica più conforme alla realtà anatomica (es. paziente 5) e una migliore continuità spaziale dei tessuti (es. pazienti 18 e 20). È opportuno specificare che tali considerazioni finali sul miglioramento della morfologia e della continuità delle strutture derivano esclusivamente da un'analisi qualitativa, e non da una valutazione formalizzata da parte di esperti.

Tra i limiti residui dell'algoritmo, si segnala la classica difficoltà nella riproduzione delle sacche d'aria intestinali (es. paziente 13), verosimilmente imputabile al divario temporale intercorso tra le acquisizioni originali TC e RM. Un ulteriore difetto è rappresentato dalla comparsa occasionale, in tutti i modelli, di artefatti puntiformi ipointensi non fisiologici all'interno di alcune slice (particolarmente visibile nel paziente 6). Ciononostante, la genesi stocastica e la localizzazione casuale di tali artefatti attraverso le varie reti addestrate rendono questa criticità del tutto trascurabile, in quanto calcolando la mediana voxel per voxel degli output restituiti dai cinque modelli, si ottiene un'inferenza finale priva di questi valori outlier. A dimostrazione di ciò, si utilizza il paziente 2, il quale, come discusso in precedenza, era stato escluso dai training-set e dai validation-set in quanto sprovvisto di una TC reale di riferimento. Utilizzando la sua acquisizione RM come input per i cinque modelli dell'ensemble e calcolando la mediana, si ottengono i risultati riportati nella figura seguente (immagini più dettagliate sono riportate a fondo dell'Appendice C).

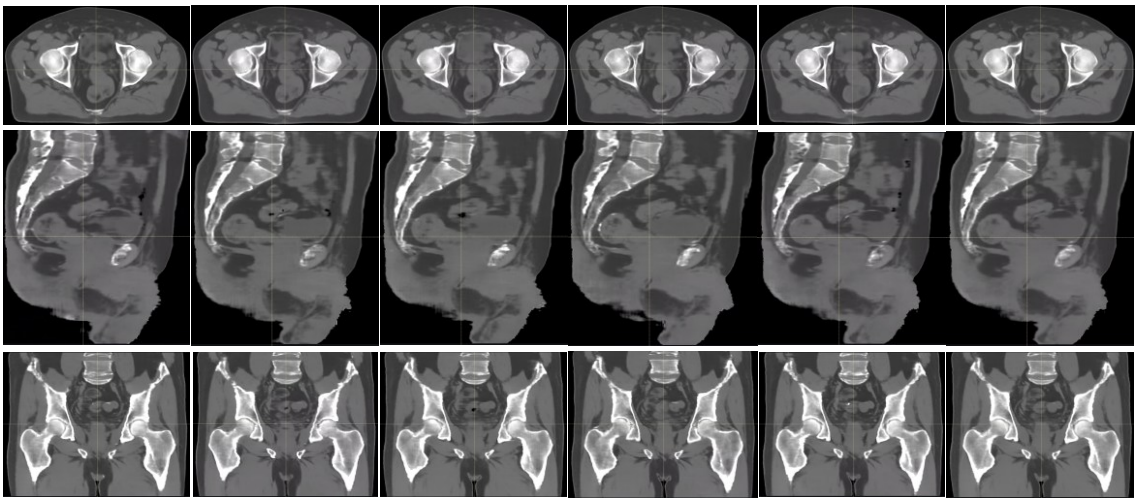


Fig. 19 Risultati della generazione di sCT per il paziente 2. Le colonne indicano rispettivamente i risultati di: modello 1, modello 2, modello 3, modello 4, modello 5 e mediana di tutti i risultati prodotti. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

Si osserva come il *bagging* tramite mediana sopprima efficacemente i valori anomali prodotti dalle singole reti, garantendo inoltre la riduzione del rumore di fondo. Tale evidenza lascia presumere che la futura valutazione su un test-set esterno indipendente possa esibire metriche persino migliori di quelle riportate dall'attuale *k-fold cross-validation*.

8 Conclusione

Il presente lavoro di tesi si conclude con il raggiungimento dell'obiettivo prefissato, attraverso l'individuazione di una strategia di TL potenzialmente riutilizzabile per i futuri adattamenti della medesima rete su differenti distretti anatomici o su input provenienti da strumentazioni differenti. Rispetto al classico addestramento con inizializzazione casuale dei pesi, l'approccio TL ha consentito di addestrare il modello impiegando una coorte ridotta di pazienti, mitigando il rischio di *overfitting* e riducendo i tempi di realizzazione della rete.

Per compensare il ridotto numero di pazienti su cui validare il modello, l'adozione della *k-fold cross-validation* si è confermata una scelta adatta, consentendo una stima più robusta delle prestazioni sulla coorte disponibile. L'addestramento multiplo generato dal protocollo k-fold ha inoltre reso possibile l'impiego di tecniche di *bagging*, ottenendo un'inferenza finale robusta e meno sensibile ai difetti dei singoli modelli, grazie all'aggregazione dei rispettivi output.

Sebbene l'utilizzo di molteplici reti in fase di inferenza comporti un inevitabile incremento dell'onere computazionale, l'impiego di architetture hardware adeguate (in questo caso una GPU NVIDIA Quadro P6000 con 24 GB di VRAM), unito all'automatizzazione e parallelizzazione delle inferenze seguita dalla loro aggregazione tramite mediana, ha permesso di ottenere una *pipeline* che restituisca le sCT in tempistiche clinicamente sostenibili, stimabili attorno ai 2 minuti e 56 secondi.

Tuttavia, bisogna riconoscere e analizzare i limiti del sistema elaborato. Dalle informazioni disponibili sui pazienti impiegati nello sviluppo del modello, si può affermare che questo garantisce prestazioni ottimali su pazienti di sesso maschile, le cui immagini siano state acquisite con scanner Philips Marlin a 1.5T (sequenze pesate T2), limitatamente ai distretti pelvico e addominale, esenti da protesi o impianti. Ne consegue la raccomandazione di circoscrivere l'utilizzo dell'algoritmo a pazienti che presentino le caratteristiche indicate e che vengano sottoposti a identici protocolli di imaging, avvalendosi di una verifica finale del risultato da parte di un esperto del settore. Tale validazione visiva dovrà inoltre adottare un approccio critico verso la problematica relativa alla mappatura delle sacche d'aria intestinali e alla possibilità di generazione di allucinazioni o artefatti.

9 Sviluppi futuri

Sebbene i risultati ottenuti nella fase di validazione siano ampiamente promettenti, il sistema sviluppato presenta ulteriori margini di ottimizzazione e necessita di successive fasi di verifica. In primo luogo, risulta indispensabile procedere con una valutazione su test set esterni indipendenti. Questo passaggio è cruciale per rilevare le reali capacità di generalizzazione del modello e per quantificare il beneficio prestazionale indotto dal *bagging* degli output. A tal fine, la validazione dovrebbe articolarsi su tre direttrici:

- L'impiego di una nuova coorte clinica proveniente dal medesimo centro (Ospedale Maggiore della Carità di Novara) e con i medesimi protocolli per osservare eventuali cali prestazionali su dati mai processati in fase di addestramento o validazione;
- L'utilizzo del dataset del task sorgente per verificare se la rete adattata abbia preservato la conoscenza iniziale, traendo peraltro vantaggio dell'architettura 2.5D;
- La costruzione di un test-set multicentrico (con differenti dispositivi e protocolli) per misurare la robustezza dell'algoritmo rispetto ad eventuali variazioni delle modalità di acquisizione delle immagini.

Il secondo step consisterà nella valutazione dosimetrica delle sCT generate, al fine di accertare che i piani radioterapeutici elaborati su di esse risultino clinicamente accettabili. Tale analisi potrà basarsi sulle metriche di dose riportate in Tabella 1, ad esempio valutando le differenze sul DVH e utilizzando il Gamma pass-rate.

Superate anche le verifiche dosimetriche, il sistema dovrà essere sottoposto al giudizio di esperti del settore cercando uno scambio di pareri volto a raccogliere spunti di miglioramento per l'impiego operativo. In aggiunta, l'esecuzione di verifiche sperimentali su fantocci antropomorfi consentirebbe di attestare i reali traguardi raggiunti dalla tecnologia, escludendo dall'analisi eventuali deviazioni imputabili esclusivamente all'imperfezione delle registrazioni elastiche tra RM e TC in vivo.

Passando invece alle possibilità di miglioramento tecnico e algoritmico del modello, i futuri sviluppi potranno concentrarsi sui seguenti aspetti:

- Correzione del bias field: L'algoritmo N4 Bias Field Correction, comunemente impiegato per correggere la non-uniformità di intensità a bassa frequenza tipica delle RM, è stato escluso già nel modello sorgente a causa di tempistiche computazionali eccessivamente onerose [59]. Sebbene le buone metriche finali dimostrino che il modello operi già in modo robusto (suggerendo una parziale capacità della GAN di compensare intrinsecamente questo artefatto), ciò non esclude la possibilità di ulteriori miglioramenti. Tra gli sviluppi futuri, si potrebbe pertanto valutare la ricerca di implementazioni ottimizzate per integrare questa correzione in fase di preprocessing, verificando così se la pulizia del segnale a monte consenta di innalzare ulteriormente le performance.
- Estensione ad ulteriori dispositivi e intensità di campo: Al fine di incrementare la robustezza generale del modello, si delinea la potenziale opportunità di ampliare il dataset con acquisizioni provenienti da strumentazioni eterogenee. Questa prospettiva includerebbe la possibilità di estendere l'applicabilità della rete anche a immagini RM a basso campo magnetico (es. 0.55 T e 0.35 T).
- Estensione ad altri distretti anatomici: Come noto, esistono già modelli commerciali per altri distretti, come ad esempio il torace o il distretto testa-collo. Si potrebbe esplorare

l'impiego del TL per adattare la rete attuale ad altre sedi anatomiche, valutando l'utilizzo di modelli altamente specifici per i diversi distretti o lo sviluppo di un'unica architettura generalizzata capace di comprenderli tutti.

- Aumento della dimensionalità dell'input: L'analisi dei risultati ha evidenziato come la riproduzione della vescica abbia beneficiato del passaggio alla configurazione 2.5D a tre slice, ma non risulta ancora estremamente fedele alla realtà. Un potenziale sviluppo potrebbe consistere nell'aumentare ulteriormente il numero di slice in input per inglobare l'intero volume dell'organo, oppure la conversione ad una rete 3D. Tuttavia, considerata la crescita rilevante del numero di parametri e il conseguente onere computazionale, la fattibilità di una simile transizione andrebbe valutata considerando l'hardware a disposizione e le tempistiche di allenamento e di inferenza.
- Ottimizzazione computazionale: Al fine di supportare un eventuale passaggio a modelli 3D e ridurre i tempi di inferenza, si potrebbero indagare diverse strategie di ottimizzazione. In fase di addestramento, l'utilizzo della *Mixed Precision* promette di migliorare l'efficienza di memoria del 40% circa, agevolando il calcolo su architetture complesse. Nel dettaglio, la tecnica combina la precisione ridotta a 16 bit (FP16) per l'allocazione di pesi, attivazioni e gradienti con la precisione standard a 32 bit (FP32) per gli aggiornamenti dell'ottimizzatore [60]. Per velocizzare invece la fase di inferenza, si potrebbe ricorrere alla quantizzazione, comprimendo i calcoli in virgola mobile in operazioni su interi. Dato che le reti GAN risultano particolarmente sensibili al degrado generato dalla semplice quantizzazione post-training, l'approccio più indicato consisterebbe nell'implementare il Quantization-Aware Training, che utilizza la quantizzazione già durante l'addestramento, agevolando il mantenimento dell'accuratezza della rete [61].
- Uncertainty quantification: Come ulteriore livello di garanzia per l'affidabilità del sistema, si potrebbe prospettare l'integrazione di tecniche volte a stimare l'incertezza predittiva del modello. L'incertezza aleatoria, legata al rumore e alla variabilità dei dati in ingresso, potrebbe essere mappata mediante la *Test-Time Augmentation*, che prevede l'applicazione di manipolazioni spaziali reversibili ai dati prima dell'inferenza e la loro successiva trasformazione contraria sugli output. Invece, l'incertezza epistemica, intrinseca invece ai parametri della rete, potrebbe essere indagata ricorrendo al Monte-Carlo Dropout, procedendo con la generazione di molteplici predizioni dal medesimo input mantenendo il dropout attivo. In entrambi gli scenari, il calcolo della deviazione standard per singolo voxel e l'estrazione del relativo logaritmo naturale consentirebbero di ricavare mappe di incertezza, capaci di segnalare visivamente le aree a minore confidenza predittiva [62].

Appendice A

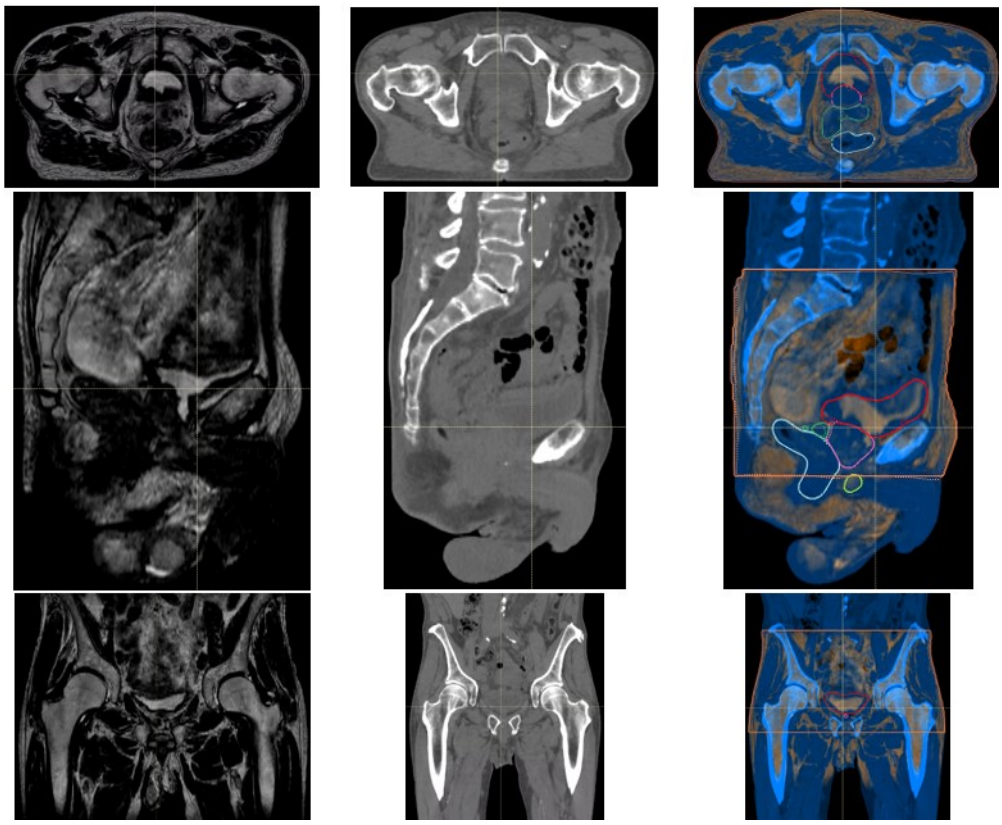


Figure Paziente 1; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

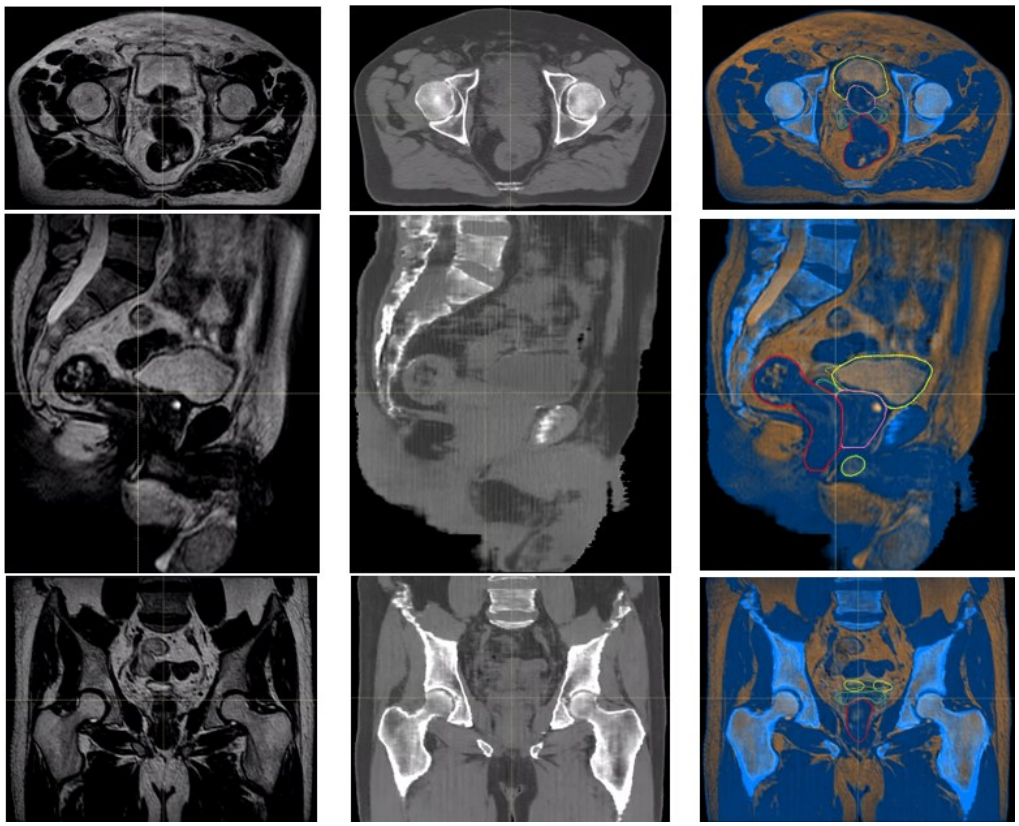


Figure Paziente 2; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

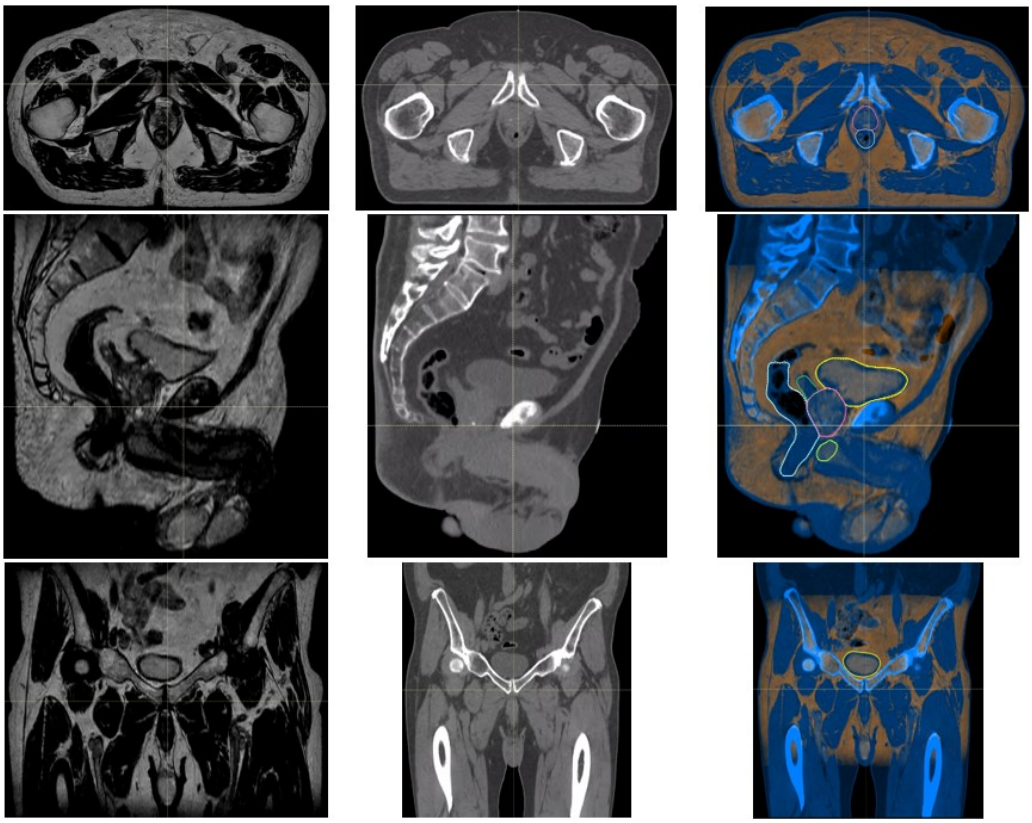


Figure Paziente 3; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

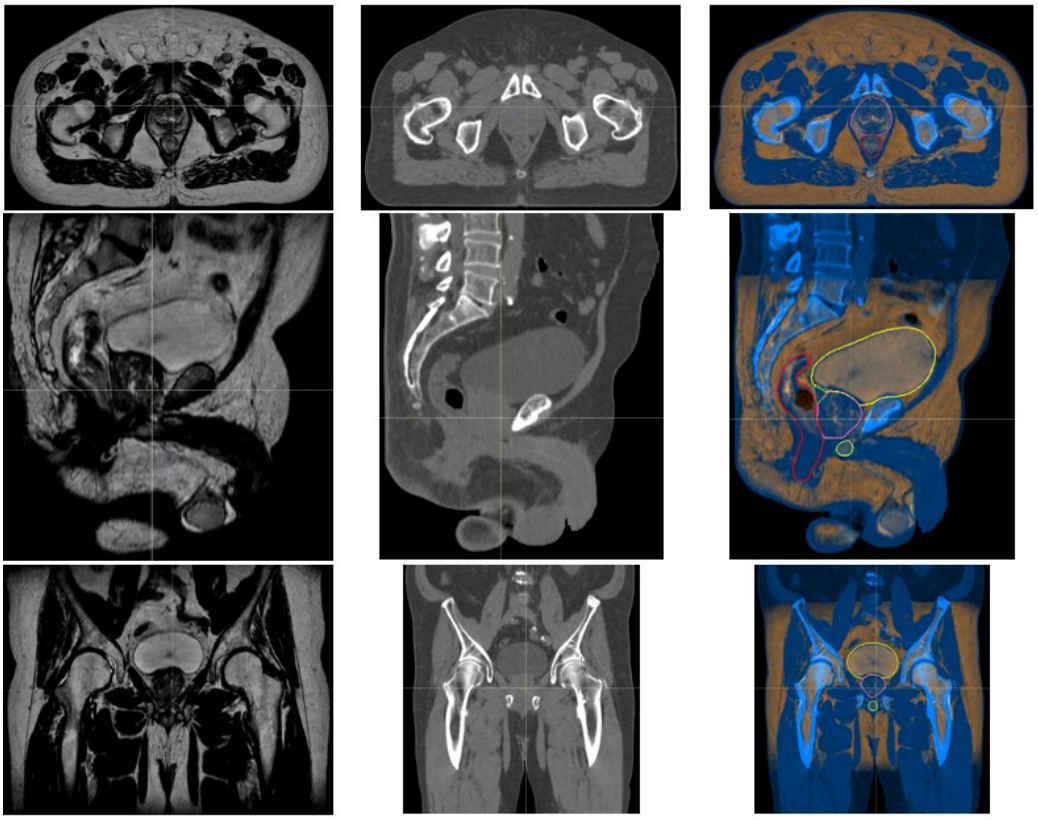


Figure Paziente 4; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

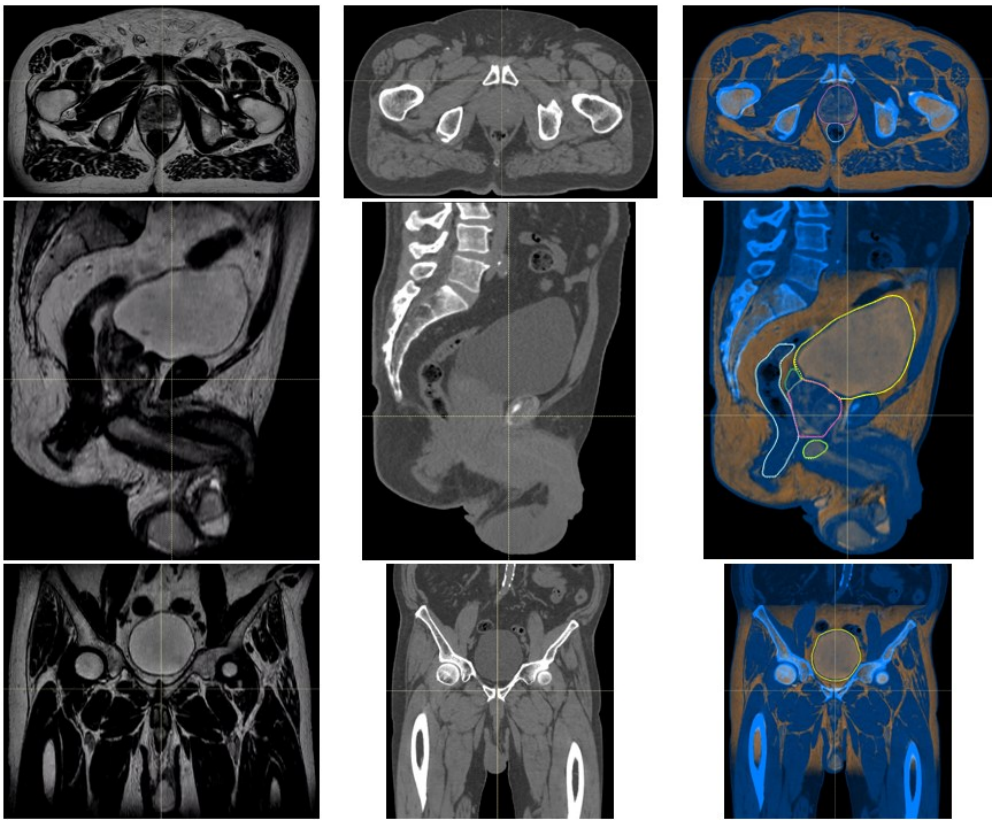


Figure 1 Paziente 5; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

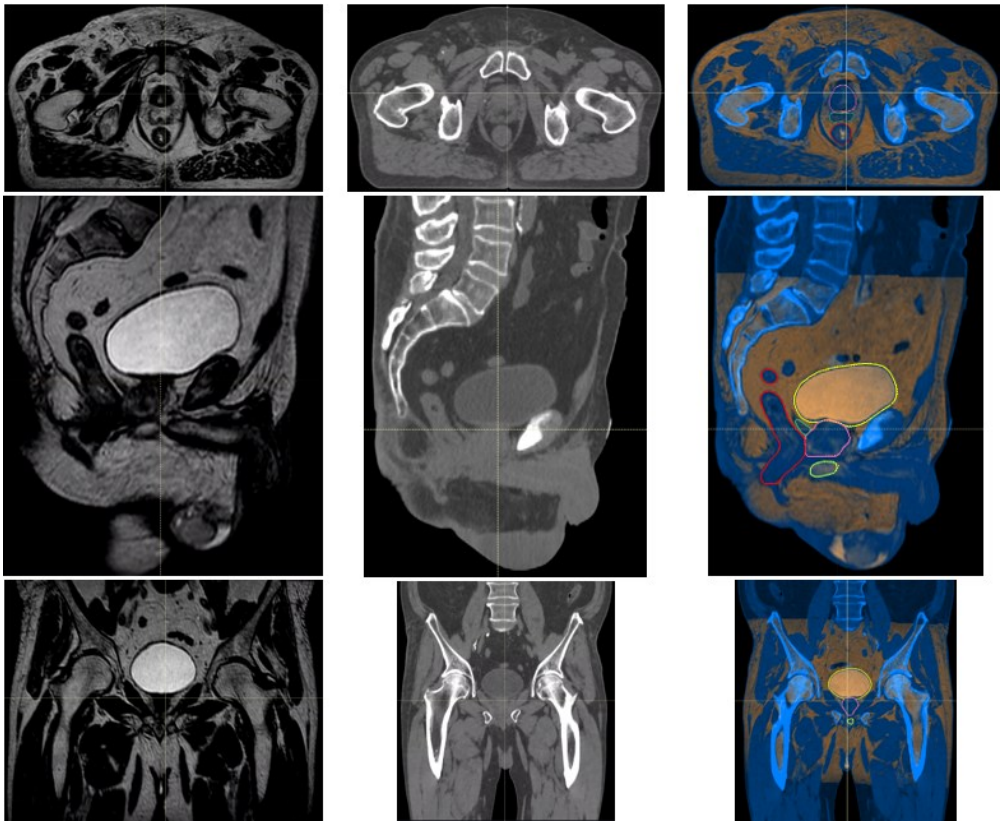


Figure Paziente 6; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

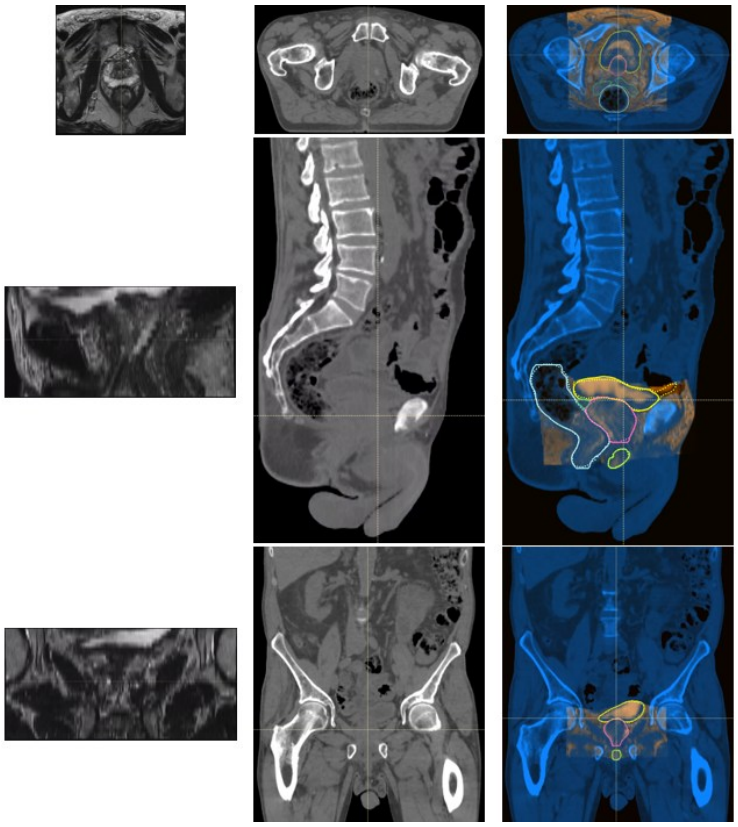


Figure Paziente 7; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

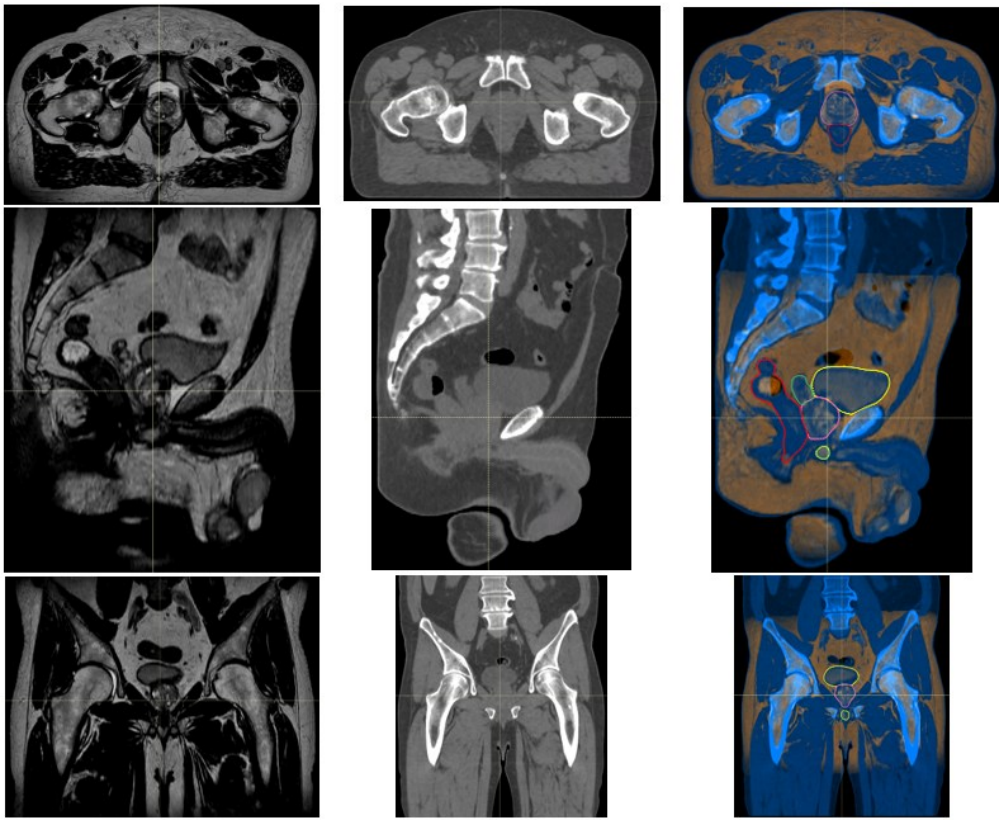


Figure Paziente 8; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

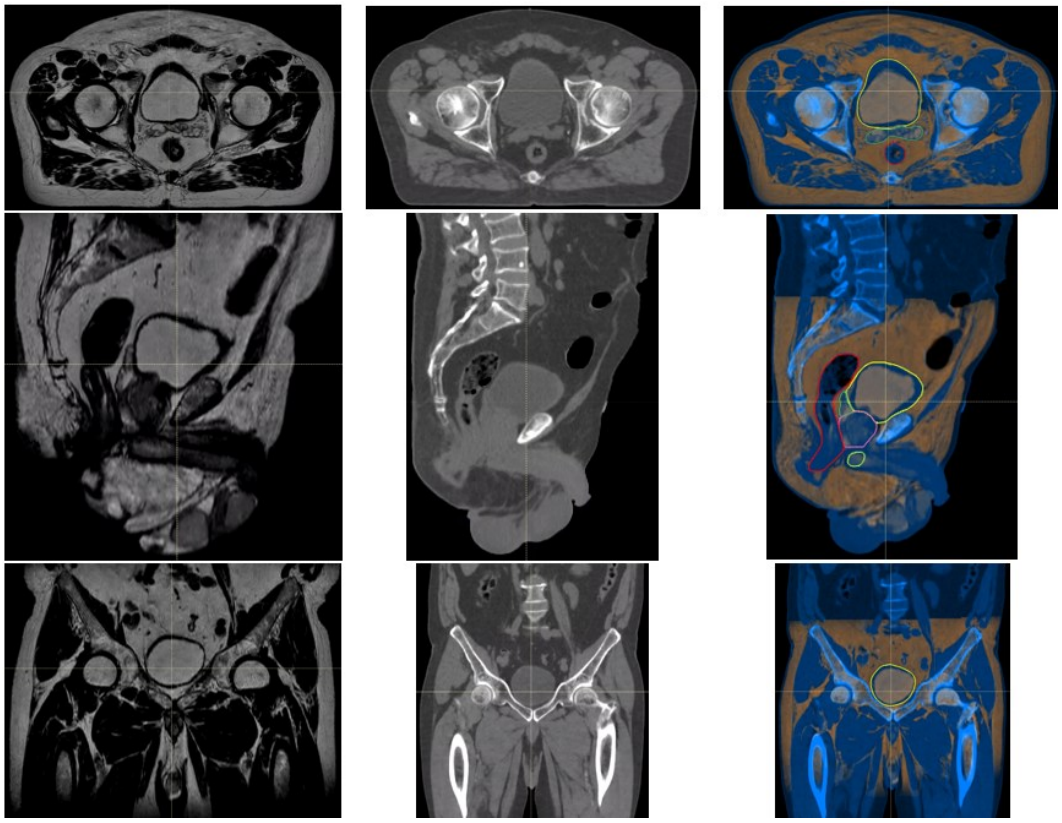


Figure Paziente 9; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

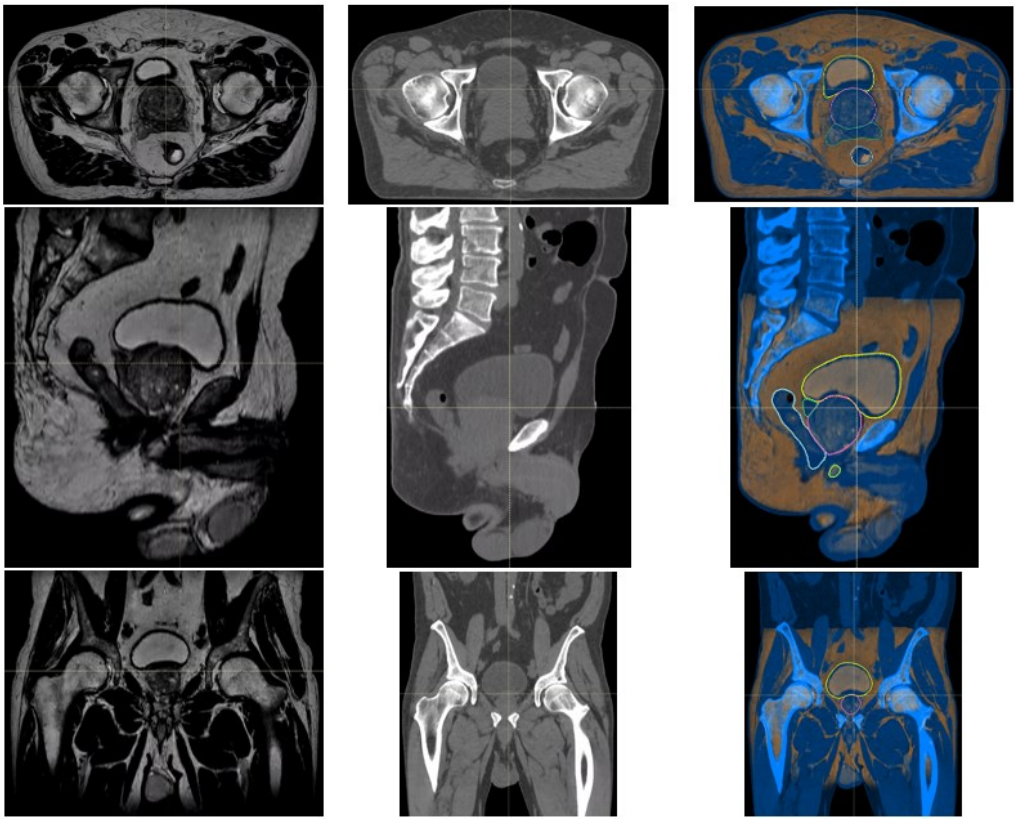


Figure Paziente 10; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

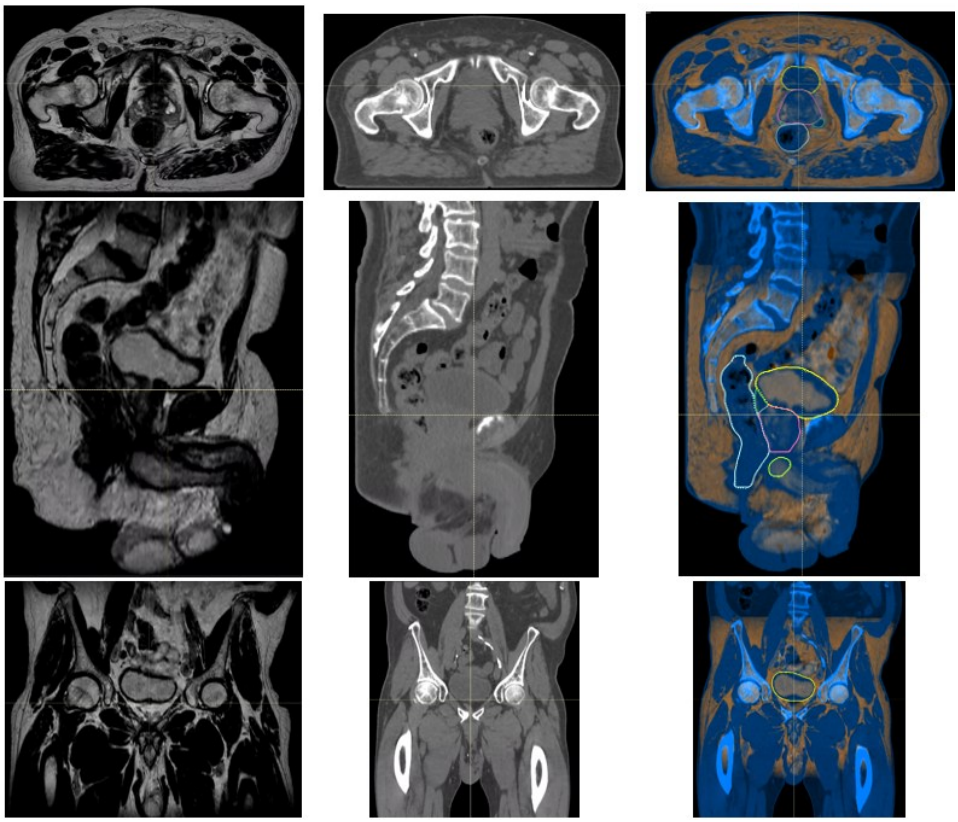


Figure Paziente 11; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

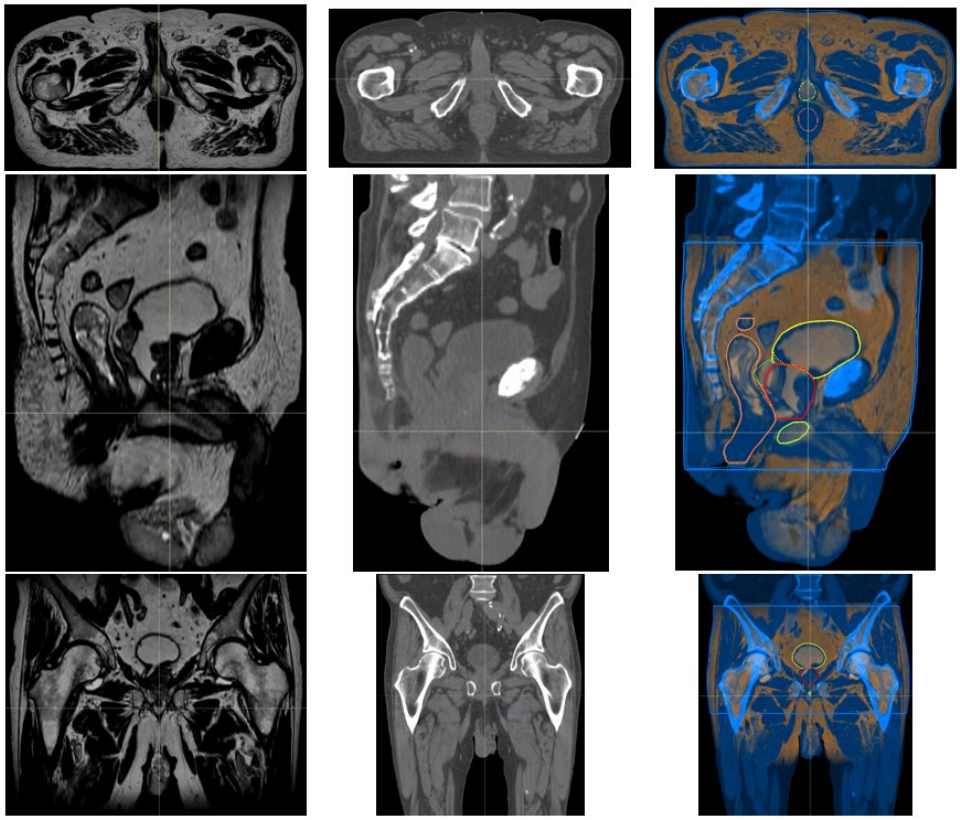


Figure Paziente 12; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

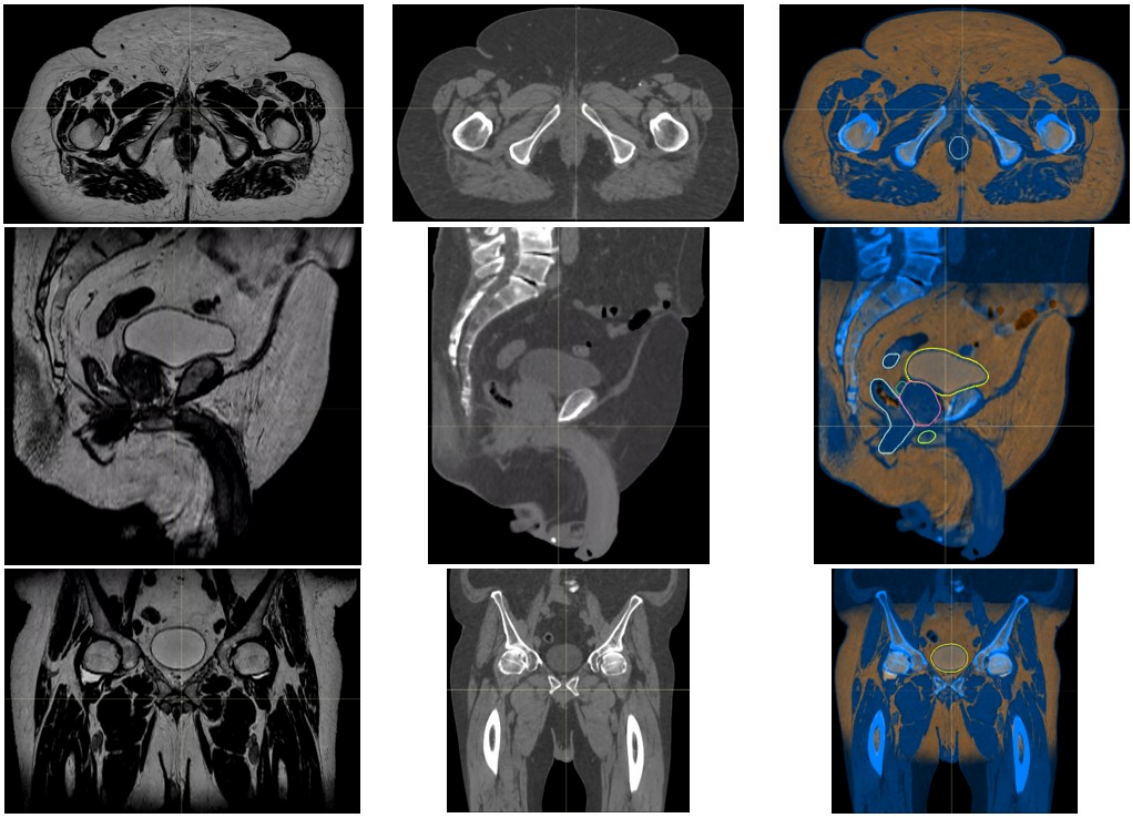


Figure Paziente 13; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

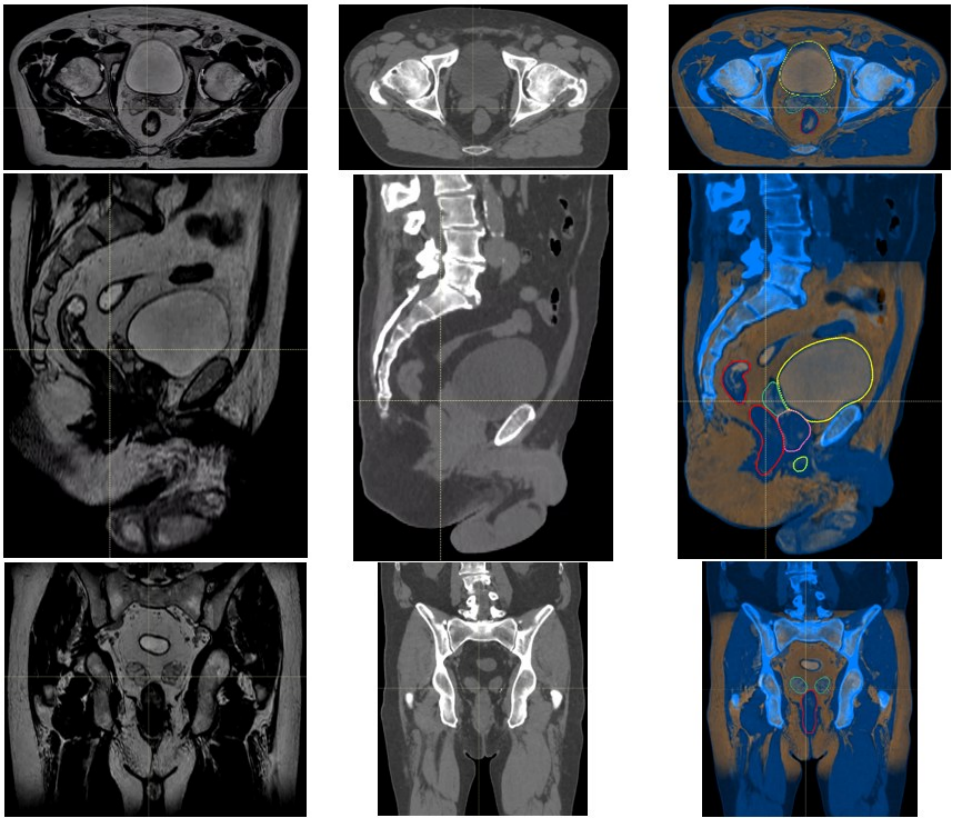


Figure Paziente 14; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

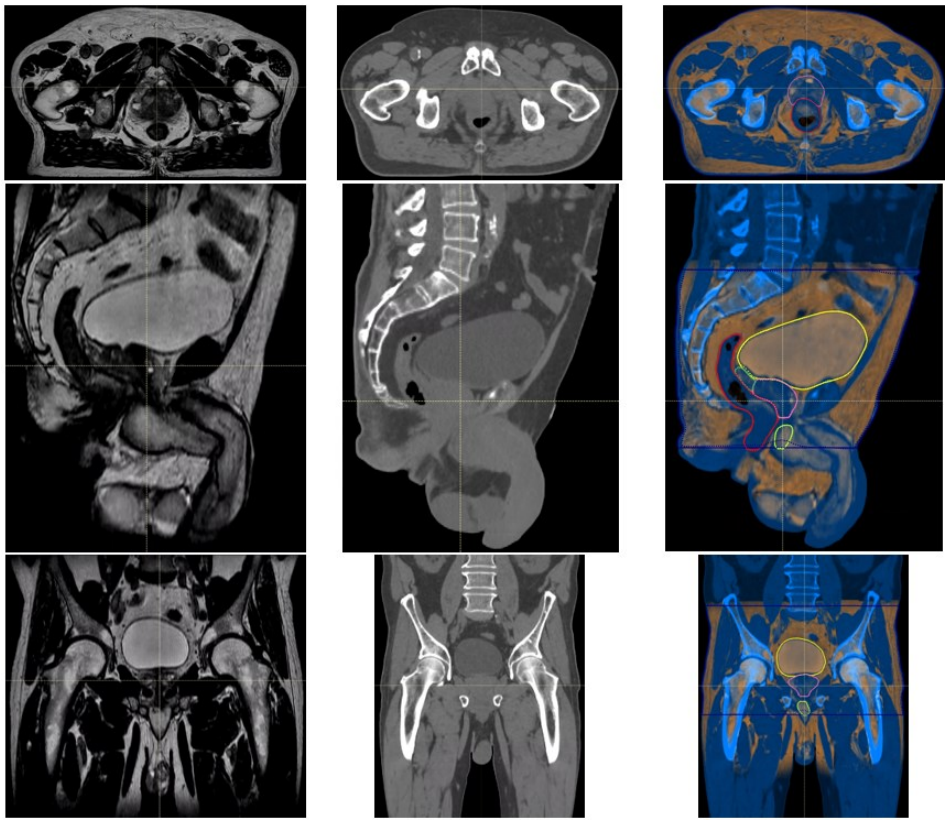


Figure Paziente 15; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

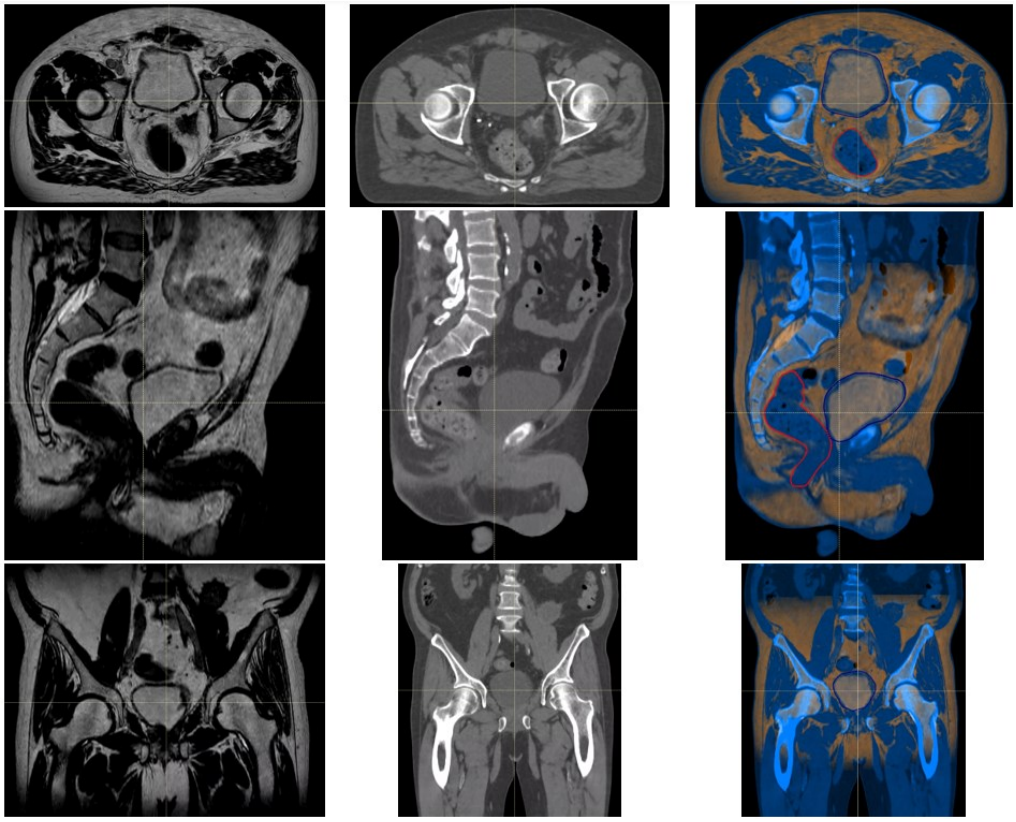


Figure Paziente 16; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

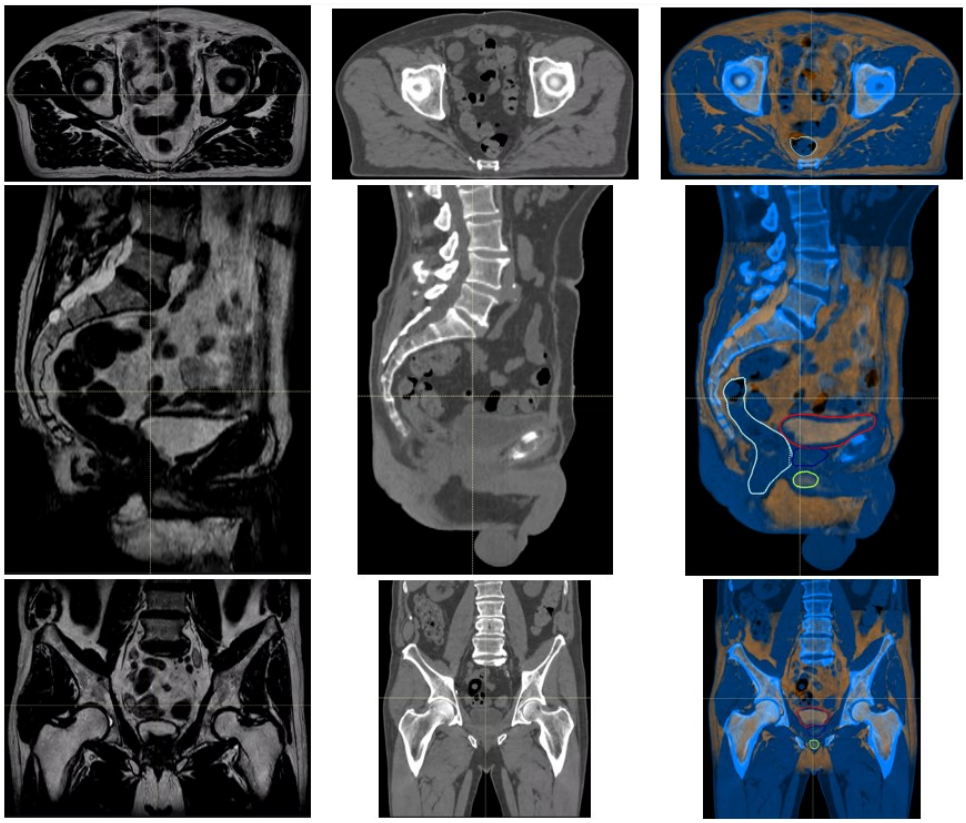


Figure Paziente 17; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

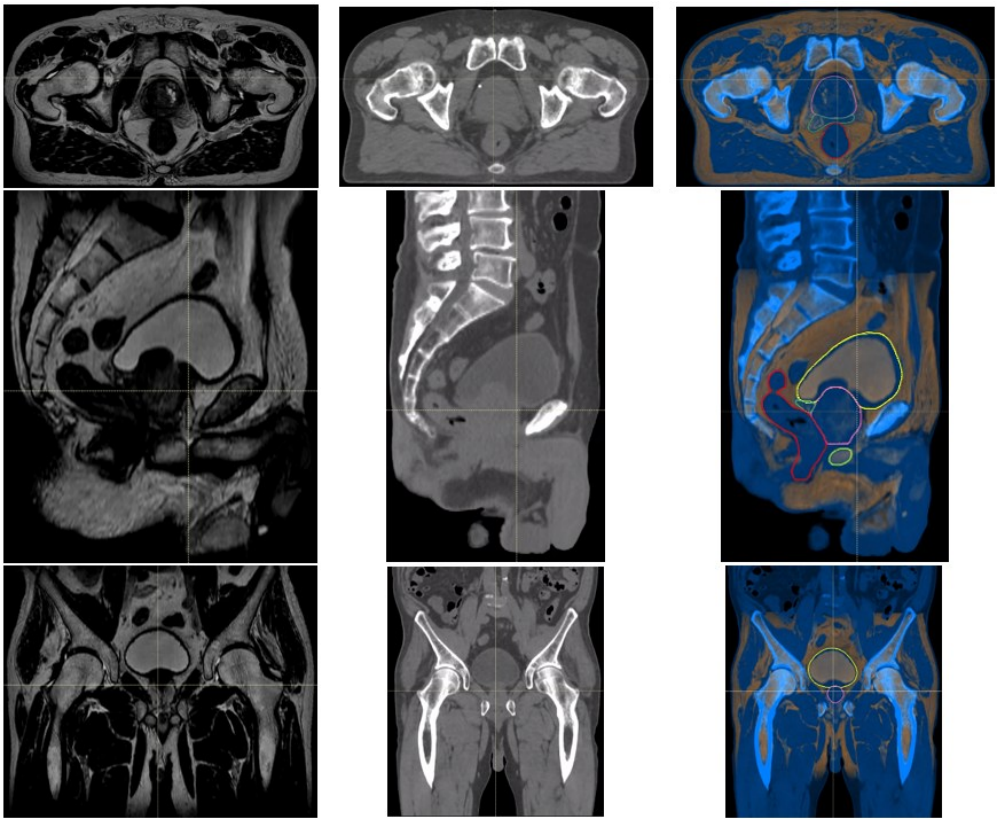


Figure Paziente 18; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

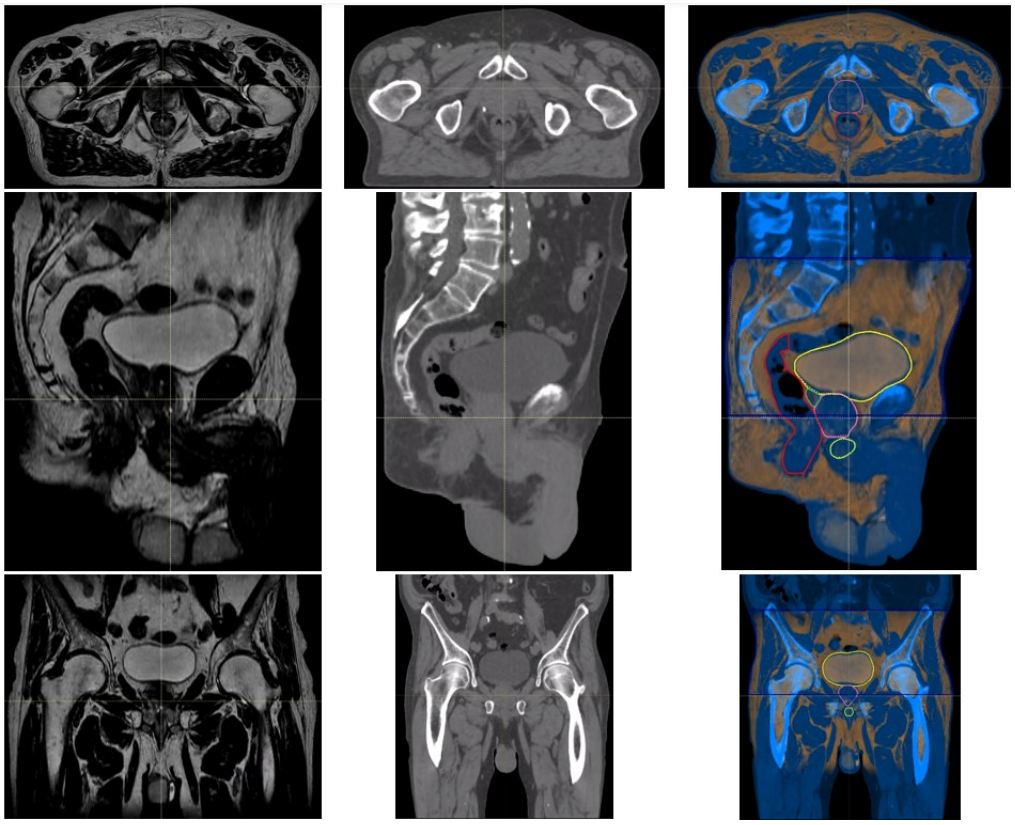


Figure Paziente 19; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

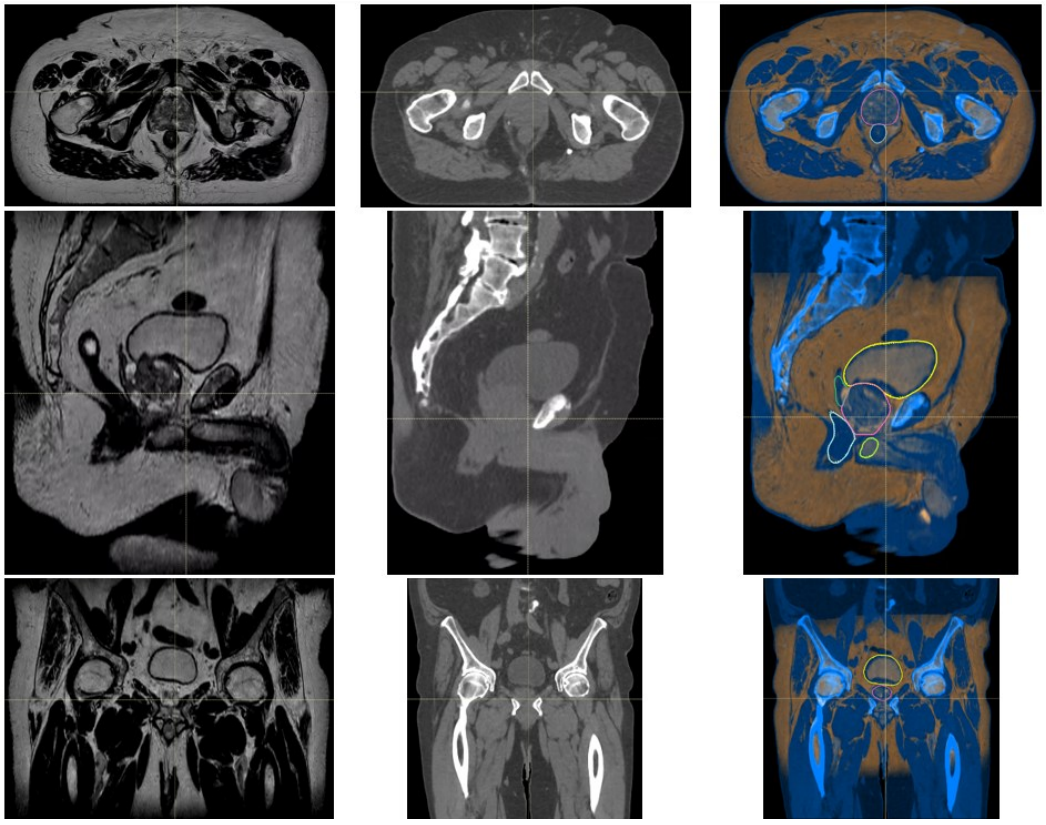


Figure Paziente 20; RM – TC – Registrazione. Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

Appendice B

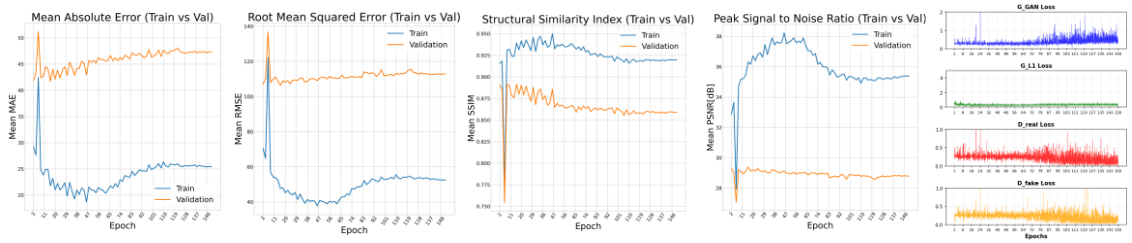


Figure Grafici prova 1 su primo fold.

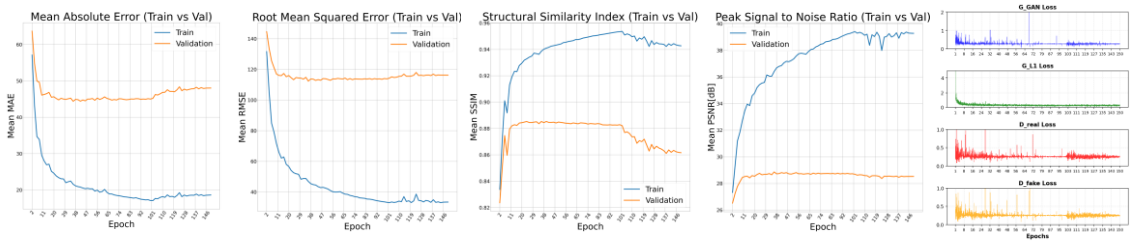


Figure Grafici prova 2 su primo fold.

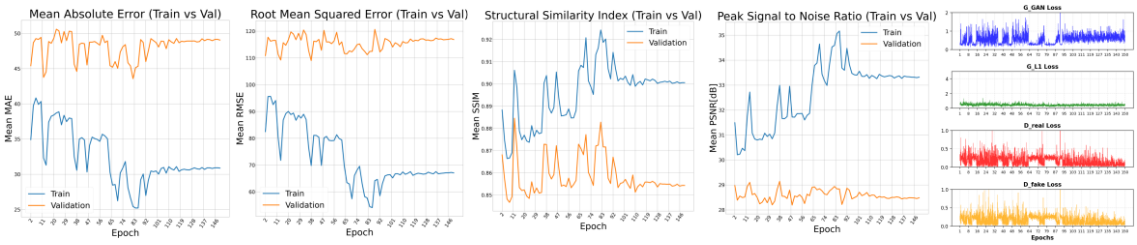


Figure Grafici prova 3 su primo fold.

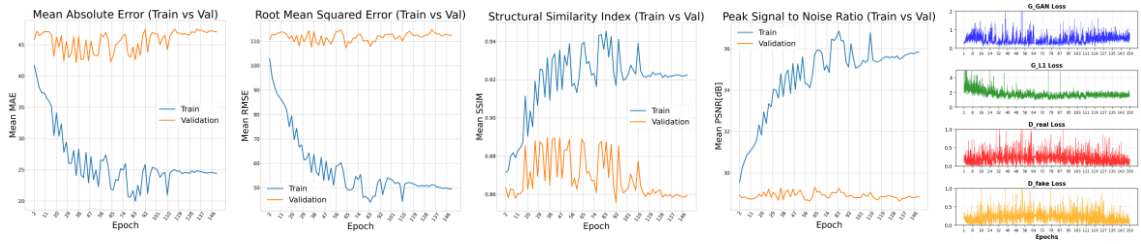


Figure Grafici prova 4 su primo fold.

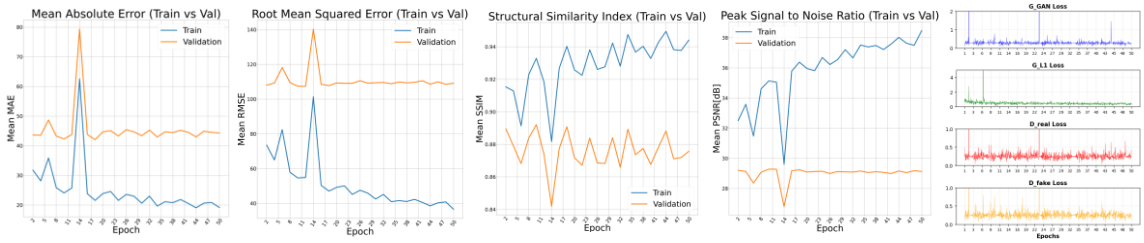


Figure Grafici prova 5 su primo fold.

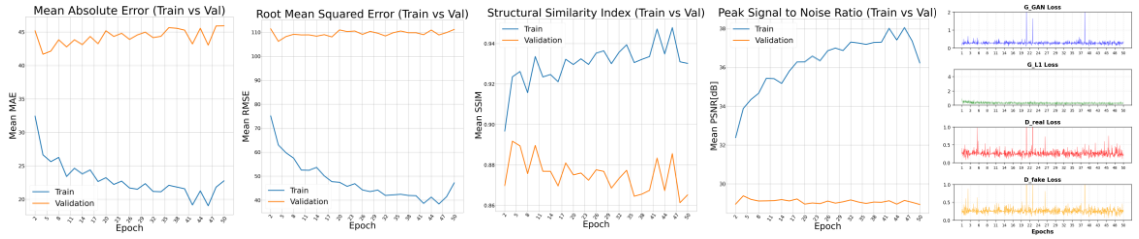


Figure Grafici prova 6 su primo fold.

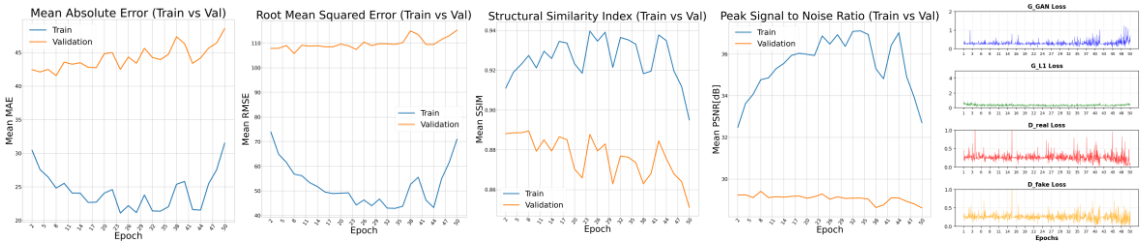


Figure Grafici prova 7 su primo fold.

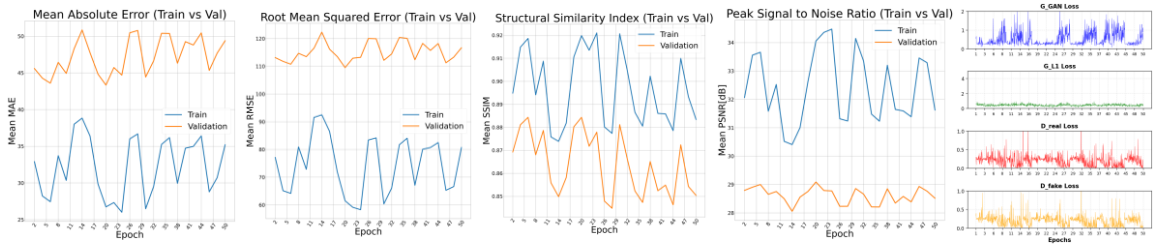


Figure Grafici prova 8 su primo fold.

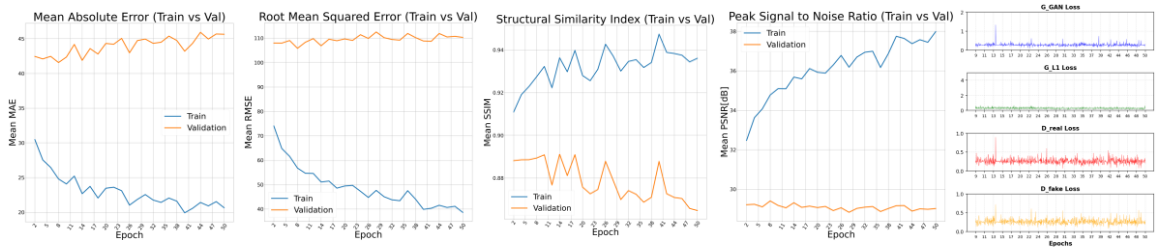


Figure Grafici prova 9 su primo fold.

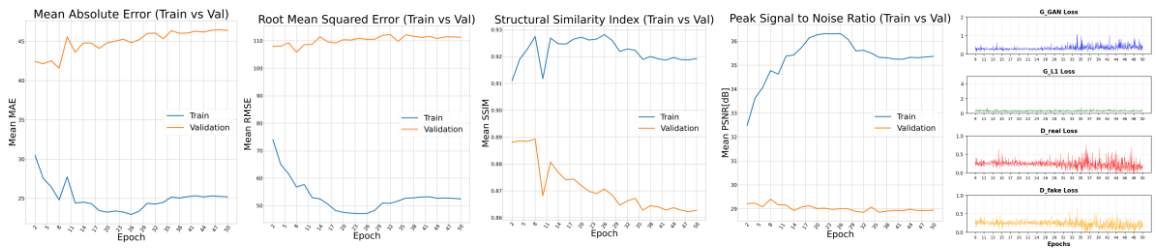


Figure Grafici prova 10 su primo fold.

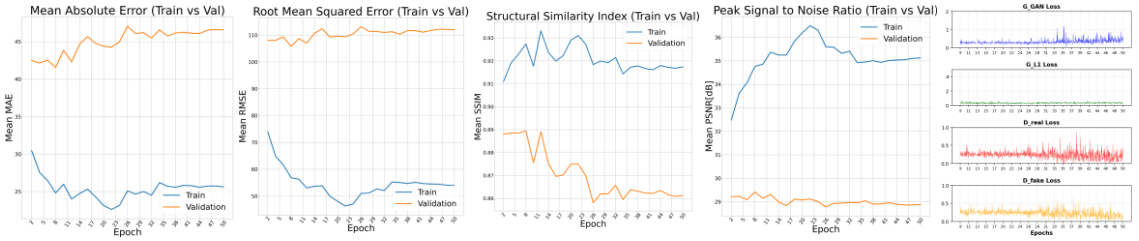


Figure Grafici prova 11 su primo fold.

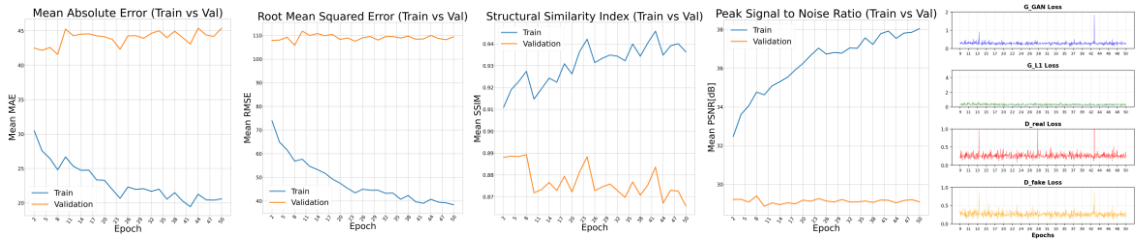


Figure Grafici prova 12 su primo fold.

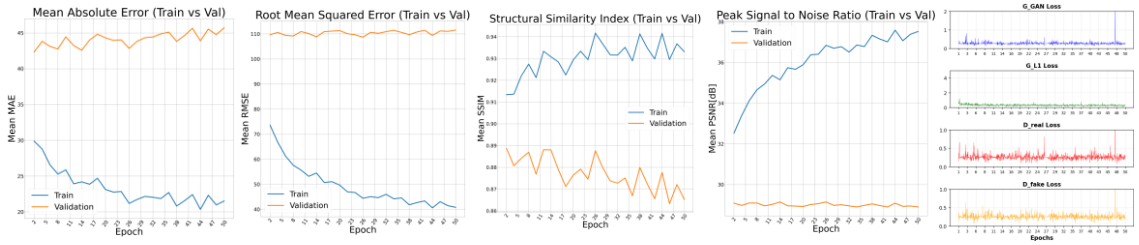


Figure Grafici prova 13 su primo fold.

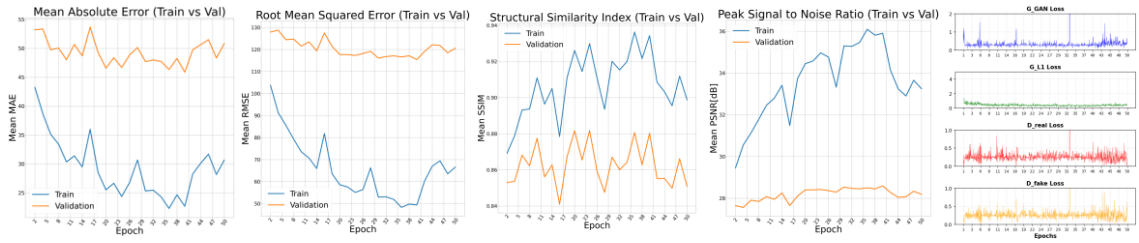


Figure Grafici prova 14 su primo fold.

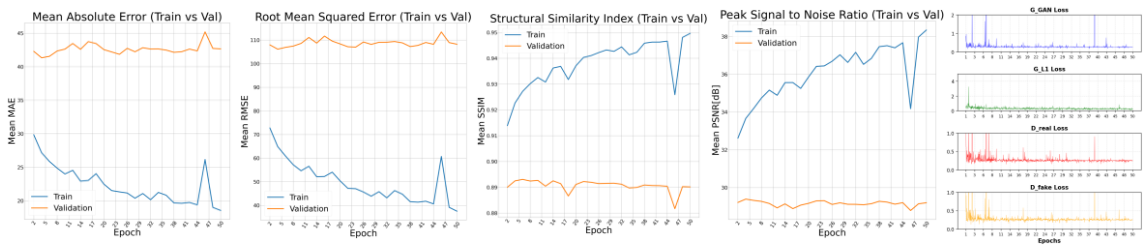


Figure Grafici prova 15 su primo fold.

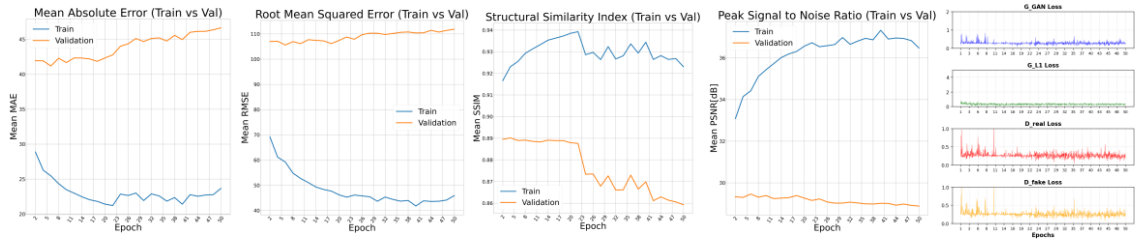


Figure Grafici prova 16 su primo fold.

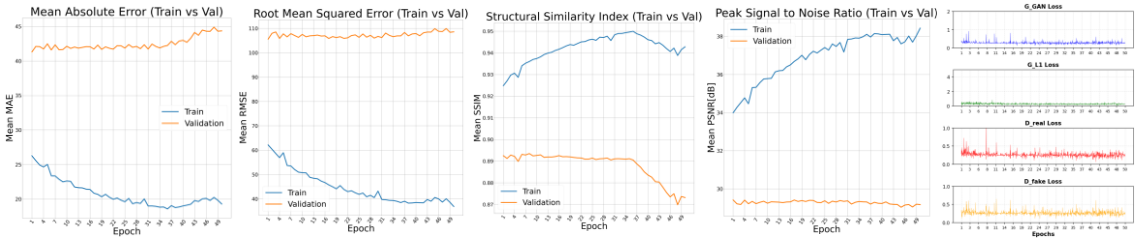


Figure Grafici 2.5D su primo fold.

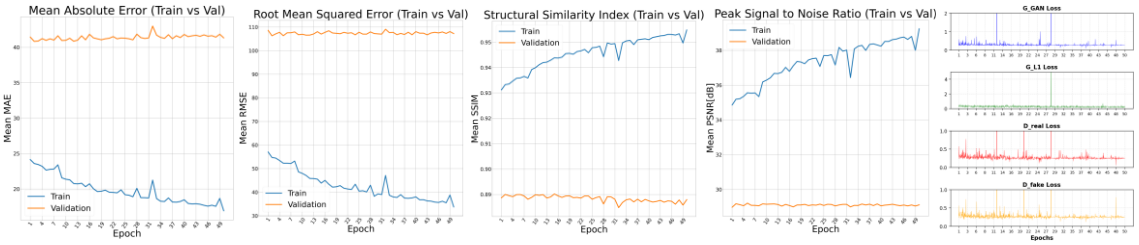


Figure Grafici 2.5D su secondo fold.

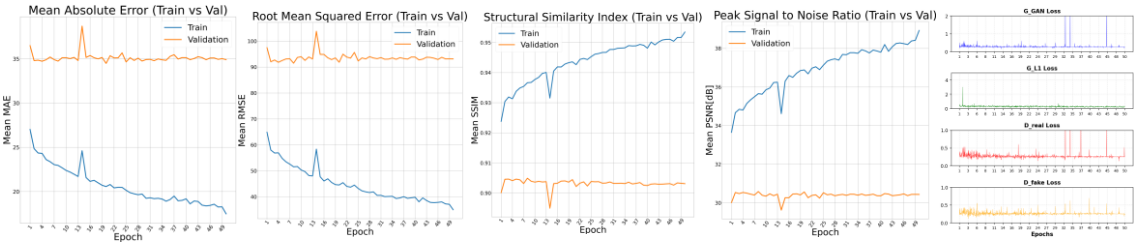


Figure Grafici 2.5D su terzo fold.

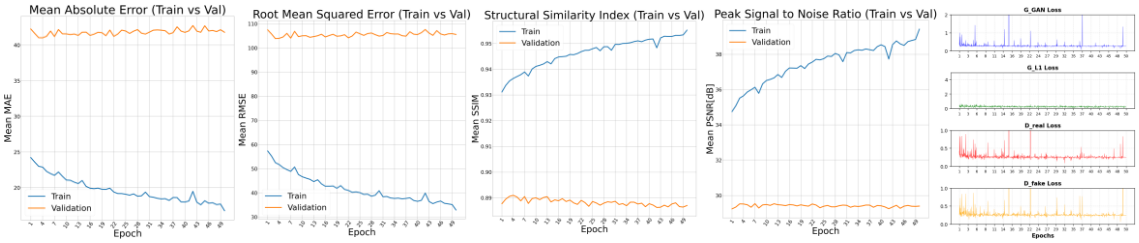


Figure Grafici 2.5D su quarto fold.

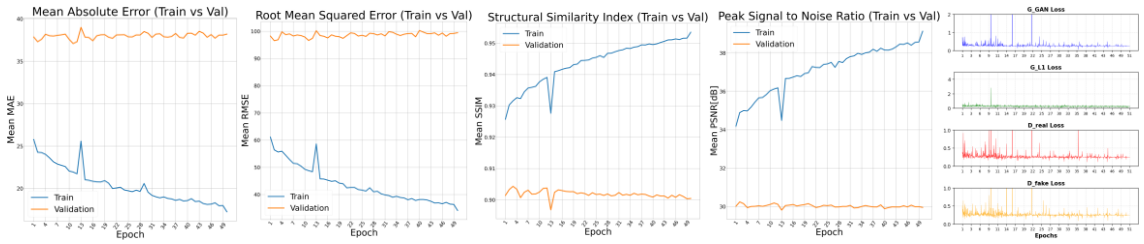


Figure Grafici 2.5D su quinto fold.

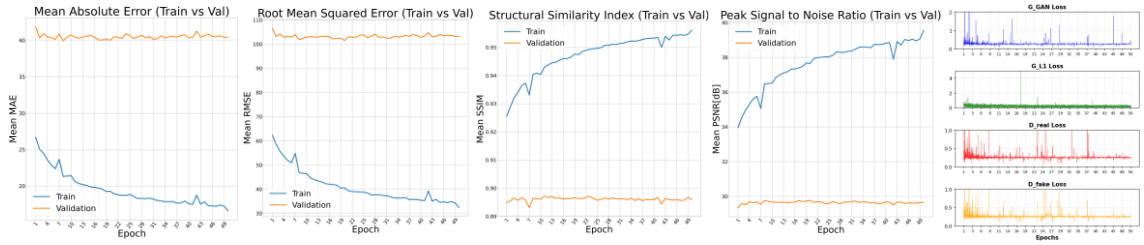


Figure Grafici 2.5D + data augmentation su primo fold.

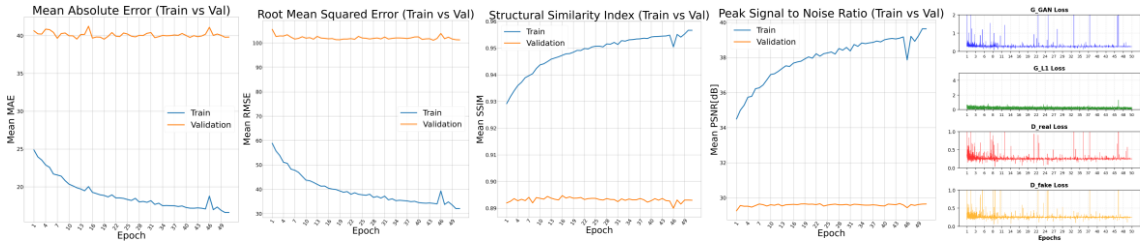


Figure Grafici 2.5D + data augmentation su secondo fold.

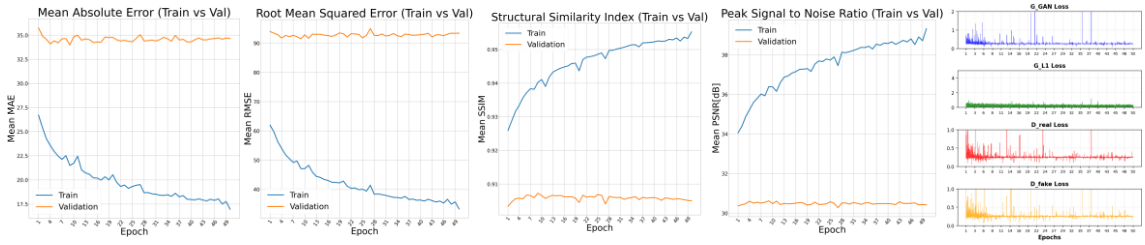


Figure Grafici 2.5D + data augmentation su terzo fold.

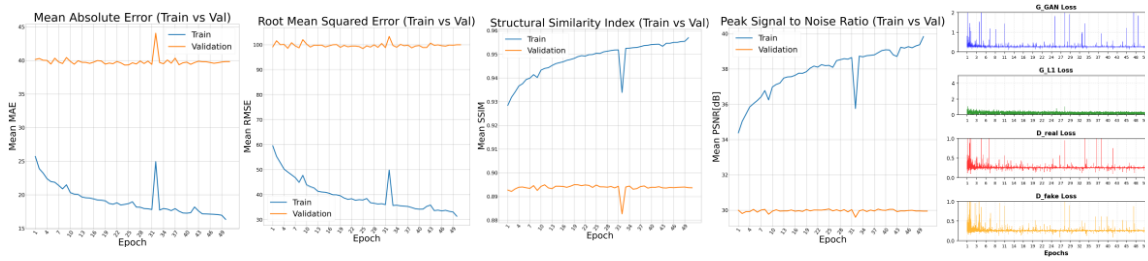


Figure Grafici 2.5D + data augmentation su quarto fold.

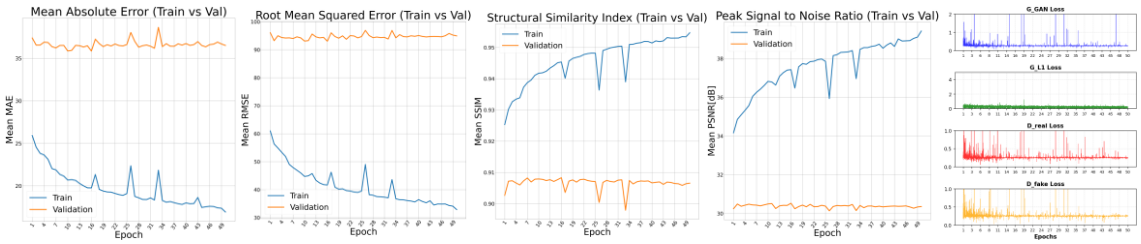


Figure Grafici 2.5 + data augmentation su quinto fold.

Appendice C

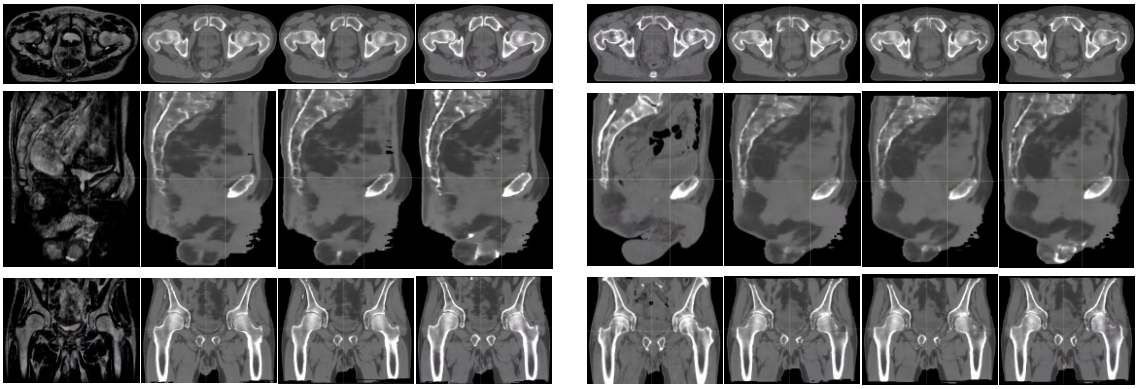


Figure Paziente 1; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

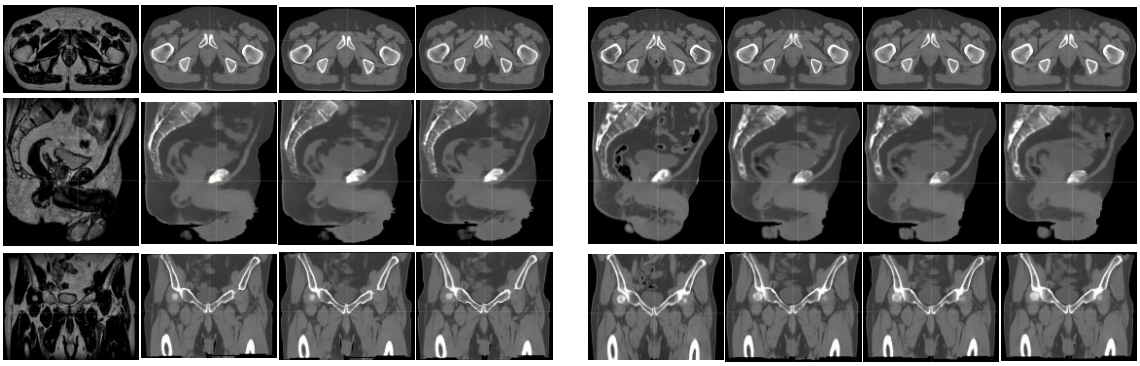


Figure Paziente 3; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

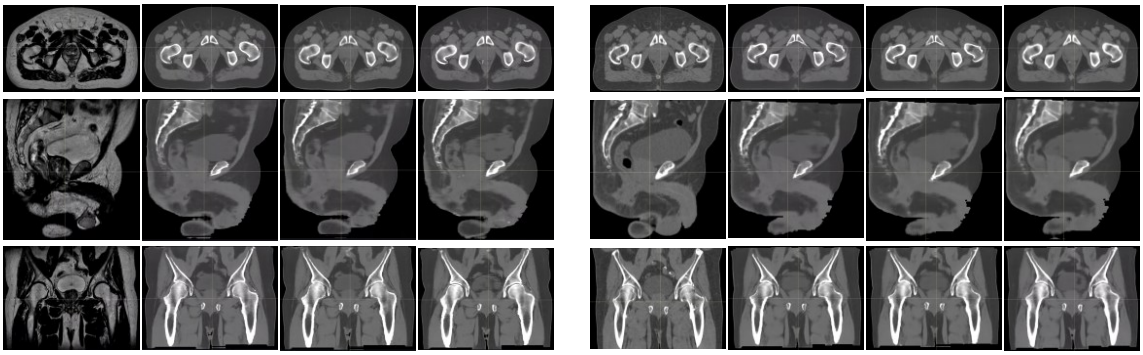
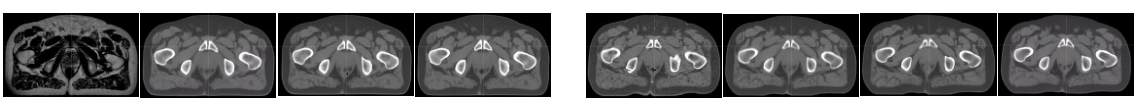


Figure Paziente 4; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.



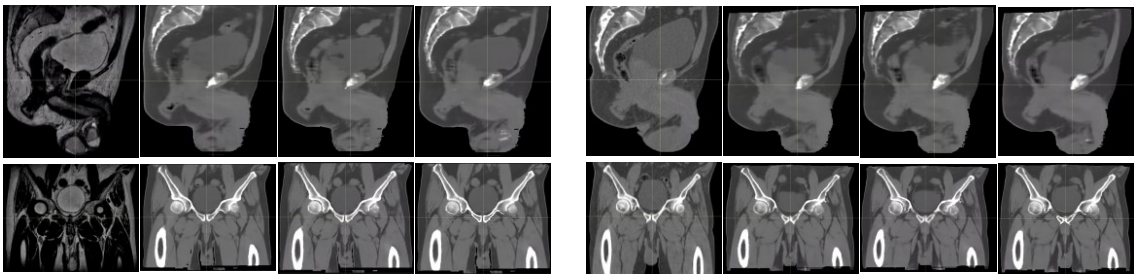


Figure Paziente 5; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

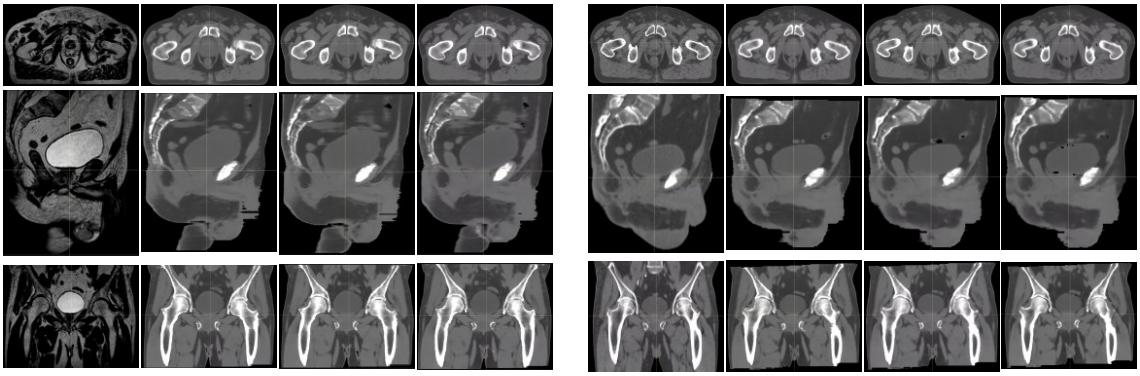


Figure Paziente 6; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

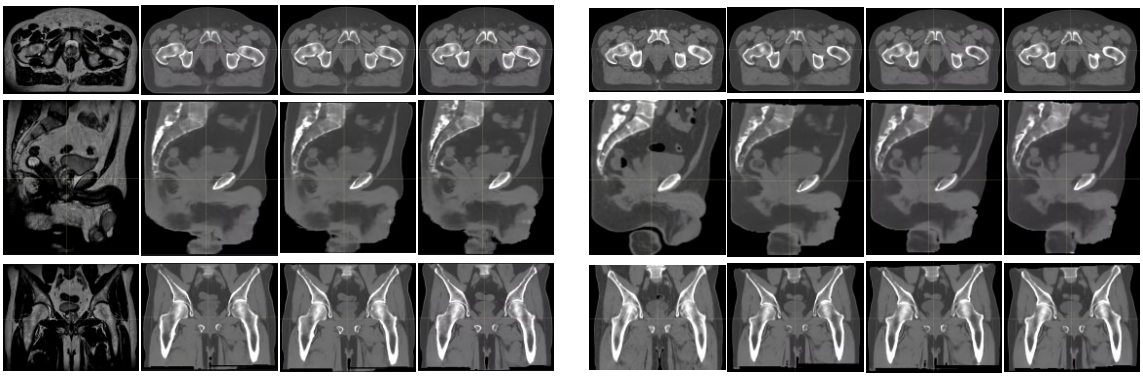


Figure Paziente 8; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

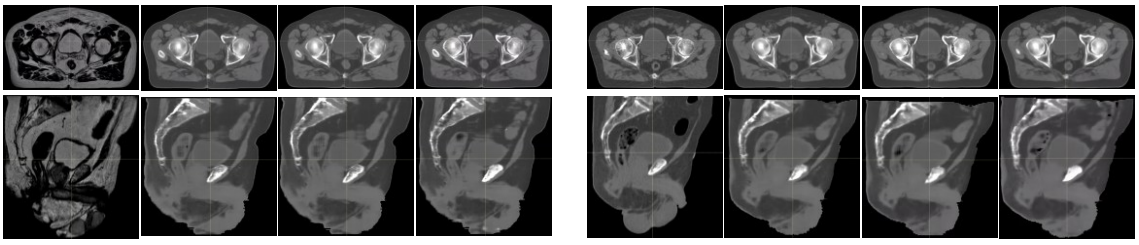




Figure Paziente 9; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

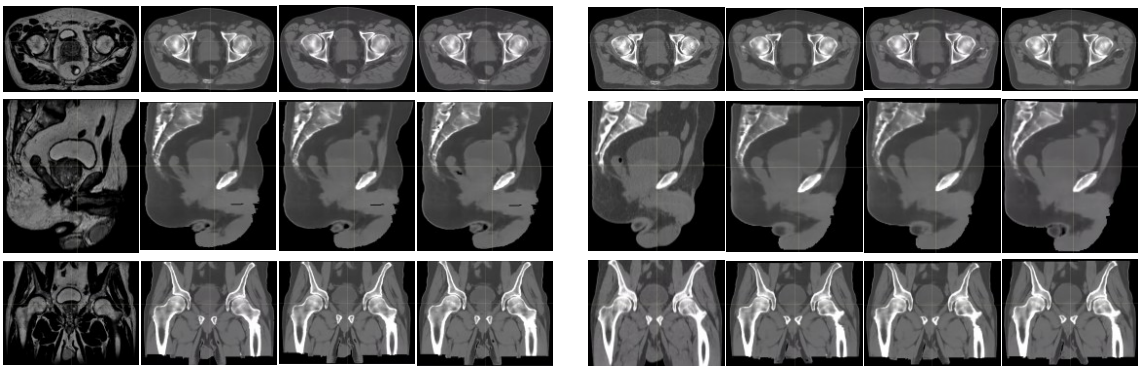


Figure Paziente 10; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

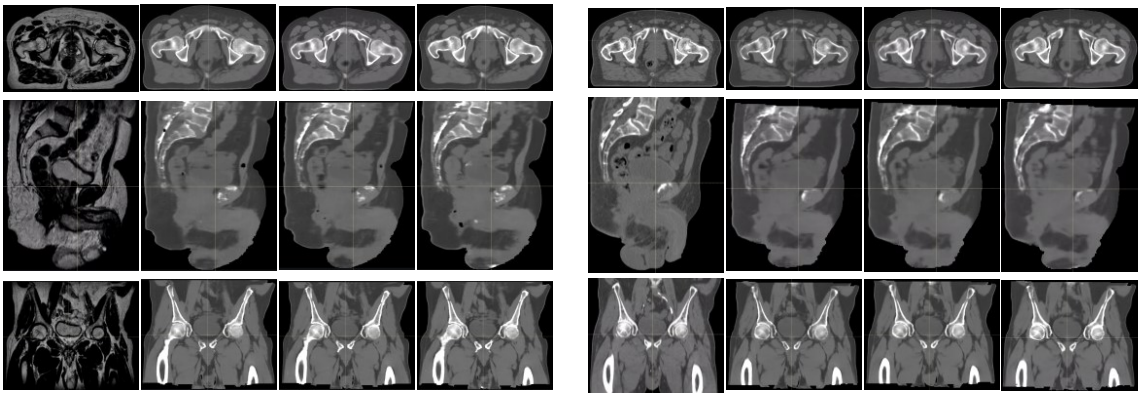


Figure Paziente 11; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

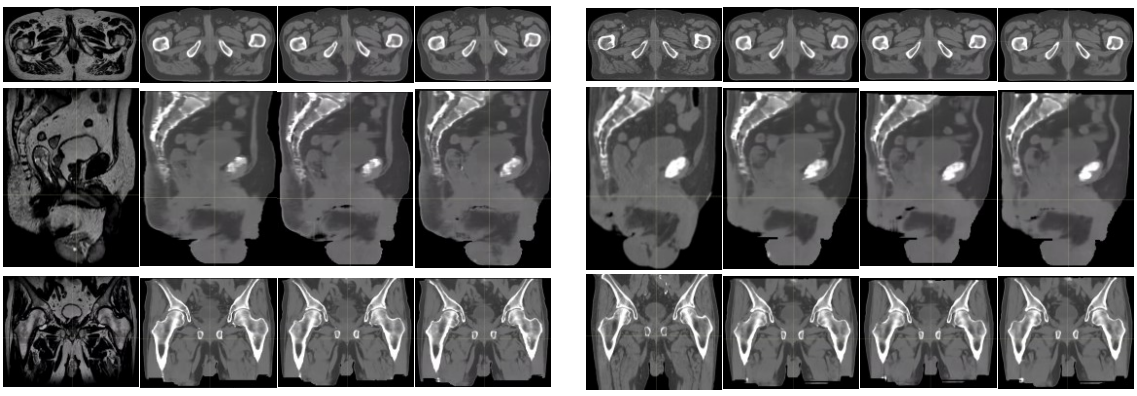


Figure Paziente 12; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

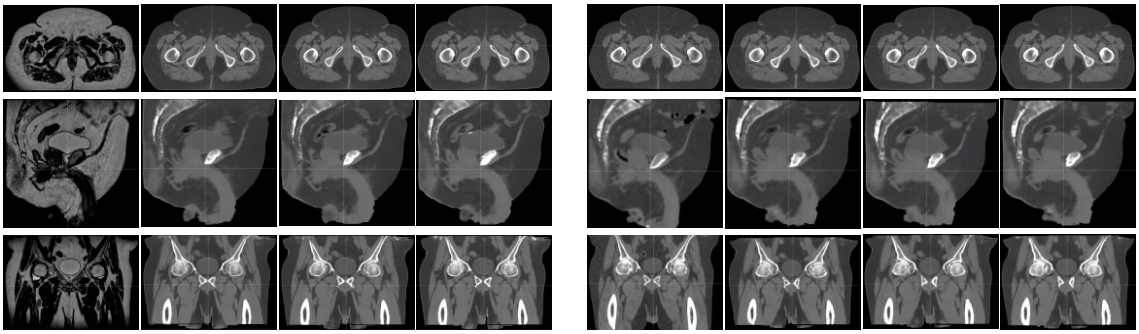


Figure Paziente 13; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

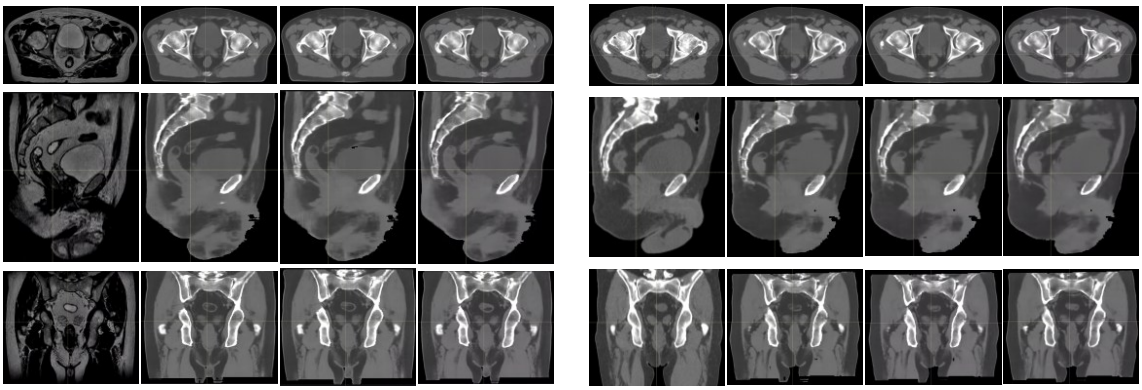


Figure Paziente 14; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

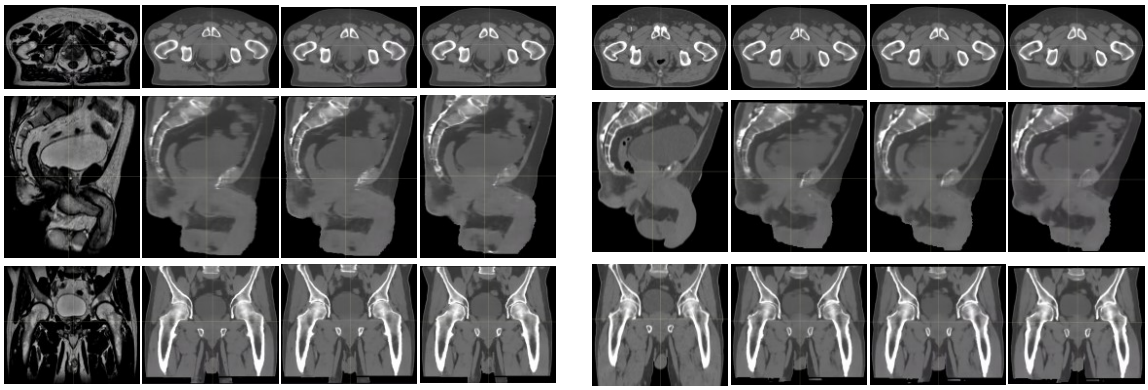


Figure Paziente 15; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

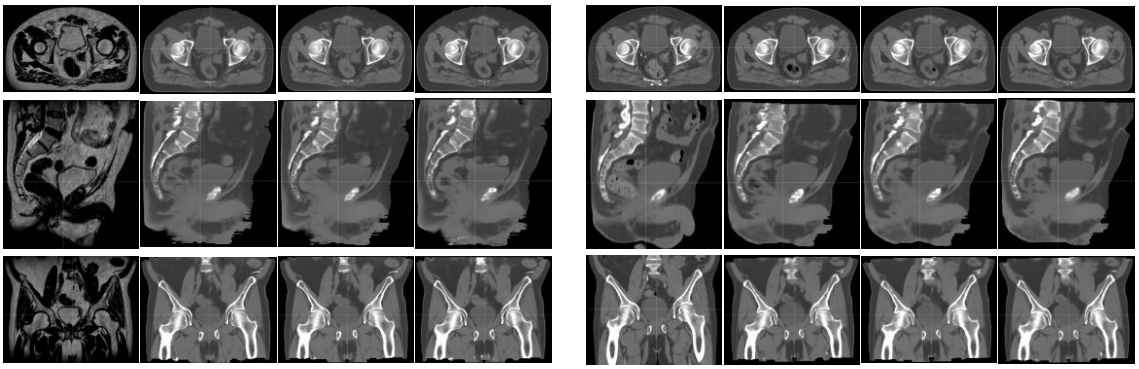


Figure Paziente 16; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

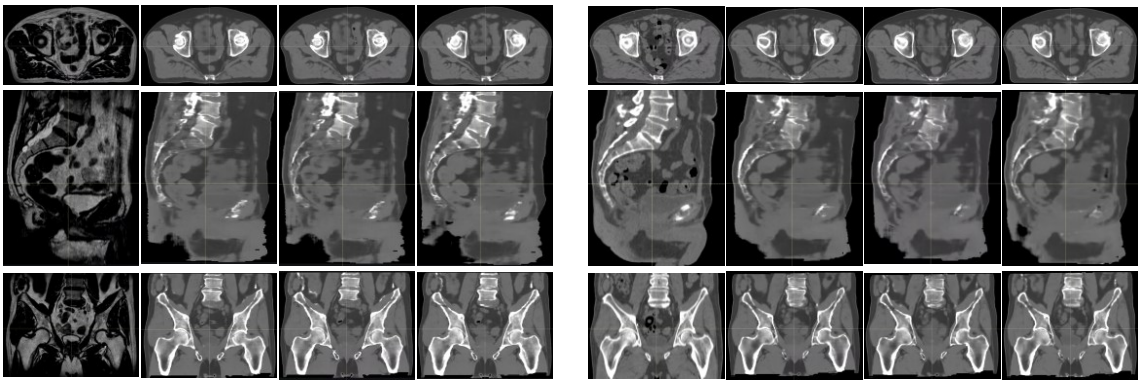


Figure Paziente 17; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

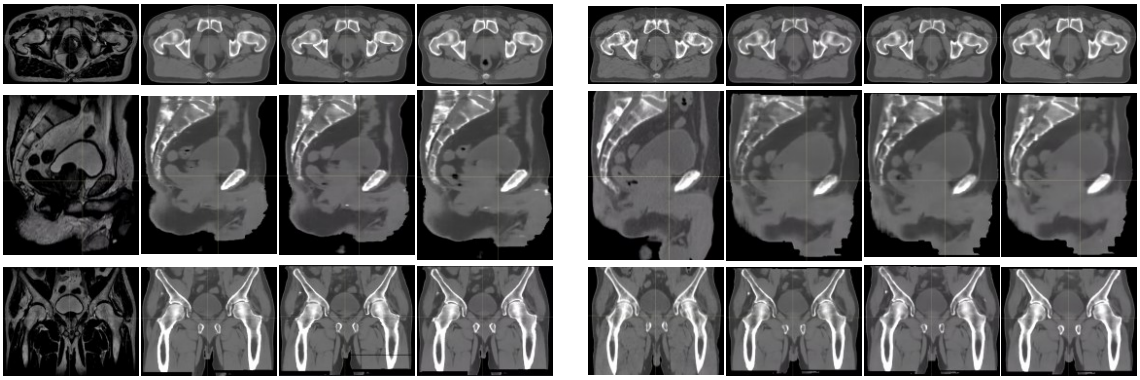
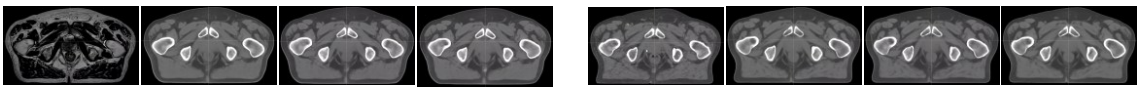


Figure Paziente 18; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.



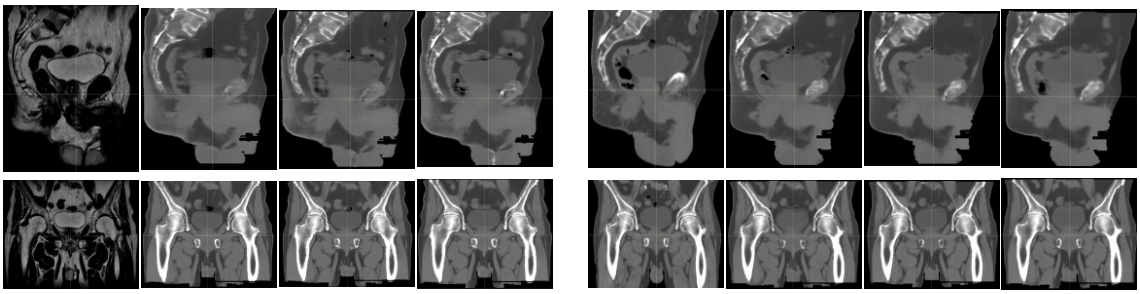


Figure Paziente 19; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

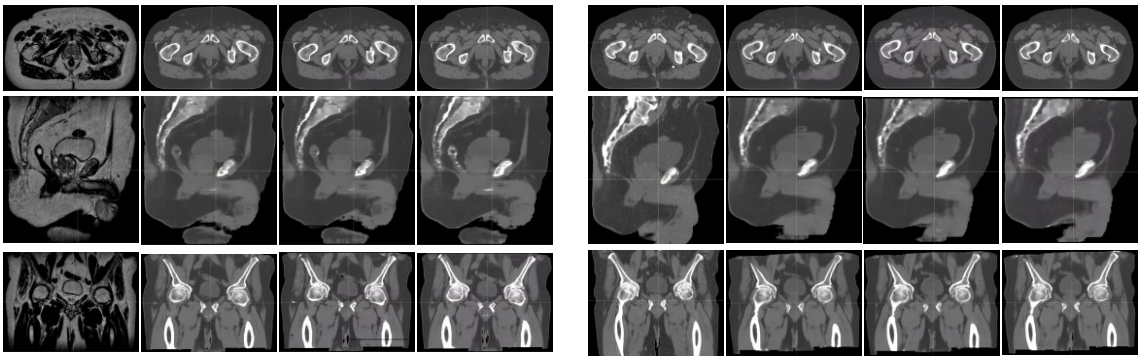


Figure Paziente 20; MR originale - sCT 2D su MR originale - sCT 2.5D su MR originale - sCT 2.5D + data augmentation su MR originale – TC vera - sCT 2D su MR deformata - sCT 2.5D su MR deformata - sCT 2.5D + data augmentation su MR deformata; Visualizzazione sui 3 piani: trasversale, sagittale e coronale.

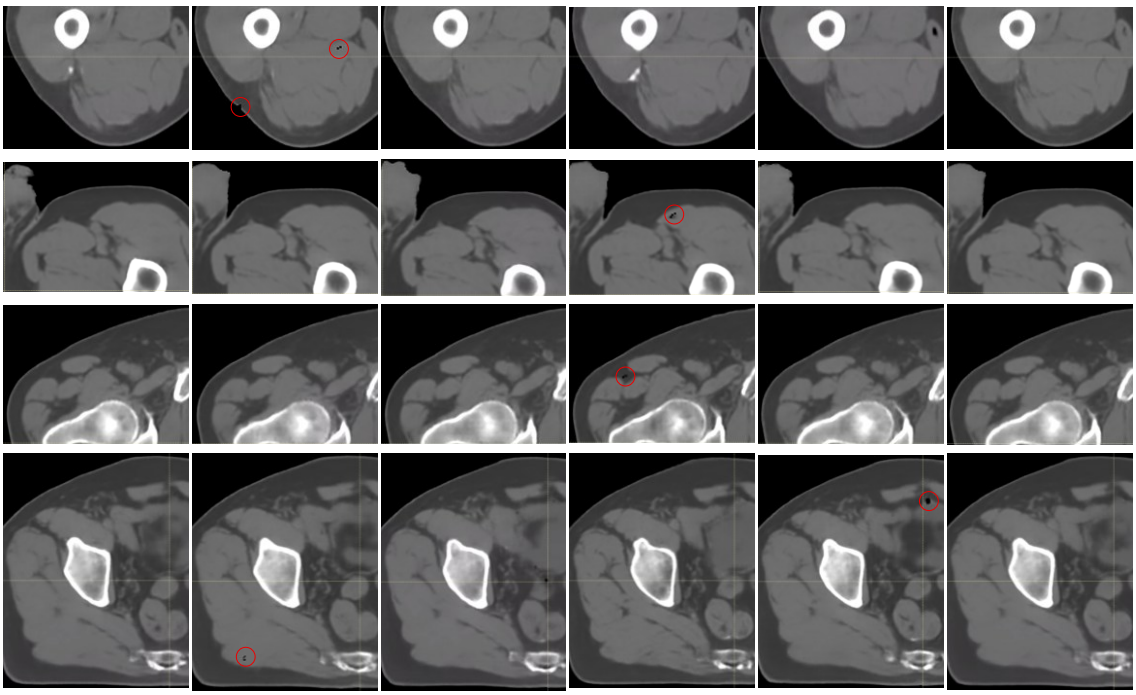


Figure I cerchi in rosso evidenziano gli artefatti trovati nella slice delle sCT generate per il paziente 2 dai 5 modelli. Le colonne mostrano: modello 1 - modello 2 - modello 3 - modello 4 - modello 5 - mediana delle 5 sCT generate.

Bibliografia

[1] R. Baskar, K. A. Lee, R. Yeo e K.-W. Yeoh, «Cancer and Radiation Therapy: Current Advances and Future Directions,» *International Journal of Medical Sciences,* vol. 9, n. 3, pp. 193-199, 2012, doi: 10.7150/ijms.3635.

[2] S. R. Mehta, V. Suhag, M. Semwal e N. Sharma, «Radiotherapy: Basic Concepts and Recent Advances,» *Medical Journal Armed Forces India,* vol. 66, n. 2, pp. 158-162, 2010, doi: 10.1016/S0377-1237(10)80132-7.

[3] A. Beddok, R. Lim, J. Thariat, H. A. Shih e G. E. Fakhri, «A Comprehensive Primer on Radiation Oncology for Non-Radiation Oncologists,» *Cancers,* vol. 15, p. 4906, 2023, doi: 10.3390/cancers15204906.

[4] S. J. Hoogcarspel, J. M. V. d. Velden, J. J. W. Lagendijk, M. v. Vulpen e B. W. Raaymakers, «The feasibility of utilizing pseudo CT-data for online MRI based treatment plan adaptation for a stereotactic radiotherapy treatment of spinal bone metastases,» *Physics in Medicine & Biology,* vol. 59, n. 23, pp. 7383-7391, 2014, doi: 10.1088/0031-9155/59/23/7383.

[5] M. Fusella, E. Alvarez Andres, F. Villegas, L. Milan, T. M. Janssen, R. Dal Bello, C. Garibaldi, L. Placidi, D. Cusumano, «Results of 2023 survey on the use of synthetic computed tomography for magnetic resonance Imaging-only radiotherapy: Current status and future steps,» *Physics and Imaging in Radiation Oncology,* vol. 32, p. 100652, 2024, doi: 10.1016/j.phro.2024.100652.

[6] R. Huang e P.-K. Zhou, «DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy,» *Signal Transduction and Targeted Therapy,* vol. 6, p. 254, 2021, doi: 10.1038/s41392-021-00648-z.

[7] L.-y. Li, Y.-d. Guan, X.-s. Chen, J.-m. Yang e Y. Cheng, «DNA Repair Pathways in Cancer Therapy and Resistance,» *Frontiers in Pharmacology,* vol. 11, p. 629266, 2021, doi: 10.3389/fphar.2020.629266.

[8] G. D. Wilson, «Cell Kinetics,» *Clinical Oncology,* vol. 19, n. 3, pp. 370-384, 2007, doi: 10.1016/j.clon.2007.02.015.

[9] S. Wasim, J. Park, S. Nam e J. Kim, «Review of Current Treatment Intensification Strategies for Prostate Cancer Patients,» *Cancers,* vol. 15, p. 5615, 2023, doi: 10.3390/cancers15235615.

[10] C. Chargari, E. Deutsch, P. Blanchard, S. Gouy, H. Martelli, F. Guérin, I. Dumas, A. Bossi, P. Morice, A. N. Viswanathan, C. Haie-Meder, «Brachytherapy: An Overview for Clinicians,» *CA: A Cancer Journal for Clinicians,* vol. 69, n. 5, pp. 386-401, 2019, doi: 10.3322/caac.21578.

[11] B. A. Fraass, «The development of conformal radiation therapy,» *Medical Physics,* vol. 22, n. 11, pp. 1911-1921, 1995, doi: 10.1118/1.597607.

[12] B. Cho, «Intensity-modulated radiation therapy: a review with a physics perspective,» *Radiation Oncology Journal,* vol. 36, n. 1, pp. 1-10, 2018, doi: 10.3857/roj.2018.00122.

[13] D. Winkel et al., «Adaptive radiotherapy: The Elekta Unity MR-linac concept,» *Clinical and Translational Radiation Oncology,* vol. 18, pp. 54-59, 2019, doi: 10.1016/j.ctro.2019.04.001.

[14] Princess Margaret Cancer Centre, «Radiation Therapy on the Magnetic, https://youtube.com/shorts/MLI0AFHNUZs?si=07aeucoYX_Md2Nnu,» [Online].

[15] Bruno Vieira, Derya Demirtas, Jeroen B. van de Kamer, Erwin W. Hans, Wim van Harten, «Improving workflow control in radiotherapy using discrete-event simulation,» *BMC Medical Informatics and Decision Making,* vol. 19, p. 199, 2019, doi: 10.1186/s12911-019-0910-0.

[16] International Commission on Radiation Units and Measurements, «Prescribing, Recording, and Reporting Photon Beam Therapy (ICRU Report 50),» *ICRU Publications,* 1993.

[17] J. C. L. Alfonso, M. A. Herrero e L. Núñez, «A dose-volume histogram based decision-support system for dosimetric comparison of radiotherapy treatment plans,» *Radiation Oncology,* vol. 10, p. 263, 2015, doi: 10.1186/s13014-015-0569-3.

[18] W. Schneider, T. Bortfeld e W. Schlegel, «Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions,» *Physics in Medicine & Biology,* vol. 45, n. 2, pp. 459-478, 2000, doi: 10.1088/0031-9155/45/2/306.

[19] Z. Ji, B. Qiu, Y. Jiang, H. Sun, T. Zhao, Y. Chen, M. Li e J. Wang, «Evaluation of cone-beam computed tomography (CBCT) correction magnitudes for patient positioning in radiotherapy,» *Journal of Radiation Research and Applied Sciences,* p. 102008, 2025, doi: 10.1016/j.jrras.2025.102008.

[20] S. S. R. Langrodi, F. Goudarzi, and D. Stanbouly, «Etiology of Tinnitus on CT and CBCT: A Narrative Review,» *International Tinnitus Journal,* vol. 26, n. 2, pp. 95-100, 2022, doi: 10.5935/0946-5448.20220014.

[21] C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, J. V. Lebesque, «Definition of the prostate in CT and MRI: A multi-observer study,» *International Journal of Radiation OncologyBiologyPhysics,* vol. 43, n. 1, pp. 57-66, 1999, doi: 10.1016/S0360-3016(98)00351-4.

[22] Z. Nigogosyan, J. E. Ippolito, S. P. Collins e E. C. Wang, «Prostate MRI in Stereotactic Body Radiation Treatment Planning and Delivery for Localized Prostate Cancer,» *RadioGraphics,* vol. 42, n. 4, pp. 1251-1264, 2022, doi: 10.1148/rg.210114.

[23] O. L. Green, L. E. Henke, G. D. Hugo, «Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy,» *Seminars in Radiation Oncology,* vol. 29, n. 3, pp. 219-227, 2019, doi: 10.1016/j.semradonc.2019.02.004.

[24] J. Willmann, B. Sidiqi, C. Wang, C. Czmielewski, H. J. Li, R. Dick-Godfrey, M. Chawla, R. P. Lee, E. Gelb, A. J. Wu, M. Lovelock, Z. Zhang, E. D. Yorke, A. Rimner, «Four-Dimensional Computed Tomography-Based Correlation of Respiratory Motion of Lung Tumors With Implanted Fiducials and an External Surrogate,» *Advances in Radiation Oncology,* vol. 7, p. 100885, 2022, doi: 10.1016/j.adro.2021.100885.

[25] K. M. M. Touabti, F. Kharfi, K. Benkahila e S.-A. Merouane, «Computed Tomography/Magnetic Resonance Imaging (CT/MRI) Image Registration and Fusion Assessment for Accurate Glioblastoma Radiotherapy Treatment Planning,» *International Journal of Cancer Management,* vol. 13, n. 9, p. e103160, 2020, doi: 10.5812/ijcm.103160.

[26] RayStation, «https://www.raysearchlabs.com».

[27] RaySearch Laboratories, «Deformable registration in RayStation,» *RaySearch Laboratories,* 2017, https://www.raysearchlabs.com/media/publications/white-papers/deformable-registration-in-raystation/.

[28] R. Pestana, K. Seidensaal, C. Beyer, J. Debus, S. Klüter e J. Bauer, «Evaluation of Pelvic MRI-to-CT Deformable Registration for Adaptive MR-Guided Particle Therapy,» *International Journal of Particle Therapy,* vol. 14, p. 100636, 2024, doi: 10.1016/j.ijpt.2024.100636.

[29] J. Jonsson, T. Nyholm e K. Söderkvist, «The rationale for MR-only treatment planning for external radiotherapy,» *Clinical and Translational Radiation Oncology,* vol. 18, pp. 60-65, 2019, doi: 10.1016/j.ctro.2019.03.005.

[30] D. J. Brenner, E. J. Hall, «Computed Tomography: An Increasing Source of Radiation Exposure,» *The New England Journal of Medicine,* vol. 357, n. 22, pp. 2277-2284, 2007, doi: 10.1056/NEJMra072149.

[31] D. Pereira, L. Pereira, J. Barbosa e M. Caetano, «Applicability of MRI-only technique in external beam radiotherapy: Dosimetric evaluation, IGRT, and quality assurance A systematic review,» *Radiography,* vol. 32, p. 103202, 2026, doi: 10.1016/j.radi.2025.103202.

[32] A. Largent, A. Barateau, J. C. Nunes, C. Lafond, P. B. Greer, J. A. Dowling, H. Saint-Jalmes, O. Acosta, R. de Crevoisier, «Pseudo-CT Generation for MRI-Only Radiation Therapy Treatment Planning: Comparison Among Patch-Based, Atlas-Based, and Bulk Density Methods,» *International Journal of Radiation OncologyBiologyPhysics,* vol. 103, n. 2, pp. 479-490, 2019, doi: 10.1016/j.ijrobp.2018.10.002.

[33] M. Boulanger, J.-C. Nunes, H. Chouraka, A. Largent, S. Tahri, O. Acosta, R. De Crevoisier, C. Lafond, A. Barateau, «Deep learning methods to generate synthetic

CT from MRI in radiotherapy: A literature review,» *Physica Medica,* vol. 89, pp. 265-281, 2021, doi: 10.1016/j.ejmp.2021.07.027.

[34] L. Vandewinckele, M. Claessens, A. Dinkla, C. Brouwer, W. Crijns, D. Verellen, W. van Elmpt, «Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance,» *Radiotherapy and Oncology,* vol. 153, pp. 55-66, 2020, doi: 10.1016/j.radonc.2020.09.008.

[35] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie e L. Farhan, «Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,» *Journal of Big Data,* vol. 8, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[36] F. Villegas, R. Dal Bello, E. Alvarez-Andres, J. Dhont, T. Janssen, L. Milan, C. Robert, G. A. M. Salagean, N. Tejedor, P. Trnková, M. Fusella, L. Placidi, D. Cusumano, «Challenges and opportunities in the development and clinical implementation of artificial intelligence based synthetic computed tomography for magnetic resonance only radiotherapy,» *Radiotherapy and Oncology,* vol. 198, p. 110387, 2024, doi: 10.1016/j.radonc.2024.110387.

[37] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, Medical Image Synthesis with Context-Aware Generative Adversarial Networks - editors. MICCAI, Medical Image Computing and Computer Assisted Intervention, vol. 10435, Cham: Springer International Publishing, 2017, pp. 417-425, doi: 10.1007/978-3-319-66179-7_48.

[38] H. Emami, M. Dong, S. P. Nejad-Davarani, C. K. Glide-Hurst, «Generating synthetic CTs from magnetic resonance images using generative adversarial networks,» *Medical Physics,* vol. 45, pp. 3627-3636, 2018, doi: 10.1002/mp.13047.

[39] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, et al., «Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,» *Computer Vision and Pattern Recognition (CVPR),* 2017, doi: 10.48550/arXiv.1609.04802.

[40] P. Isola, J.-Y. Zhu, T. Zhou e A. A. Efros, «Image-to-Image Translation with Conditional Adversarial Networks,» *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 5967-5976, 2017, doi: 10.1109/CVPR.2017.632.

[41] U. Demir, G. Unal, «Patch-Based Image Inpainting with Generative Adversarial Networks,» *arXiv,* p. arXiv:1803.07422, 2018, doi: 10.48550/arXiv.1803.07422.

[42] I. J. Goodfellow; J. Pouget-Abadie; M. Mirza; B. Xu; D. Warde-Farley; S. Ozair; A. Courville; Y. Bengio, «Generative Adversarial Nets,» *arXiv,* p. arXiv:1406.2661, 2014, doi: 10.48550/arXiv.1406.2661.

[43] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang e S. P. Smolley, «Least Squares Generative Adversarial Networks,» *IEEE International Conference on Computer Vision (ICCV),* pp. 2794-2802, 2017, doi: 10.1109/ICCV.2017.302.

[44] J.-Y. Zhu; T. Park; P. Isola; A. A. Efros, «Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,» in *IEEE International Conference on Computer Vision (ICCV)*, 2017, doi: 10.1109/ICCV.2017.244.

[45] J. Yosinski, J. Clune, Y. Bengio e H. Lipson, «How transferable are features in deep neural networks?,» *Advances in Neural Information Processing Systems (NIPS),* pp. 1-9, 2014, doi: 10.48550/arXiv.1411.1792.

[46] M. Shetty, K. Raghavendra e P. N. Sarappadi, «Transfer Learning with Pix2Pix GAN for Generating Realistic Photographs from Viewed Sketch Arts,» *Journal of Southwest Jiaotong University,* vol. 57, n. 4, pp. 198-207, 2022, doi: 10.35741/issn.0258-2724.57.4.17.

[47] A. Galapon Jr, D. Wagenaar, J. A. Langendijk, S. Both, «Transfer learning for DL-based Synthetic CT after reconstruction algorithm upgrade in a proton therapy clinic,» *Medical Physics,* vol. 52, p. e17937, 2025, doi: 10.1002/mp.17937.

[48] Y. Wang, C. Wu, L. Herranz, J. van de Weijer, A. Gonzalez-Garcia, B. Raducanu, «Transferring GANs: generating images from limited data,» *arXiv,* p. arXiv:1805.01677, 2018, doi: 10.48550/arXiv.1805.01677.

[49] F. Haimour, R. Al-Sayyed, W. Mahafza e O. S. Al-Kadi, «Bidirectional brain image translation using transfer learning from generic pre-trained models,» *Computer Vision and Image Understanding,* vol. 248, p. 104100, 2024, doi: 10.1016/j.cviu.2024.104100.

[50] «https://tecnologieavanzate.com/casi-studio/ai-deep-learning-sct/».

[51] B. J. Soher, B. M. Dale e E. M. Merkle, «A Review of MR Physics: 3T versus 1.5T,» *Magnetic Resonance Imaging Clinics of North America,* vol. 15, n. 3, pp. 277-290, 2007, doi: 10.1016/j.mric.2007.06.002.

[52] S. R. Jannat, K. Lynch, M. Fotouhi, S. Cen, J. Choupan, N. Sheikh-Bahaei, G. Pandey e B. A. Varghese, «Advancing 1.5T MR imaging: toward achieving 3T quality through deep learning super-resolution techniques,» *Frontiers in Human Neuroscience ,* vol. 19, p. 1532395, 2025, doi: 10.3389/fnhum.2025.1532395.

[53] «https://www.raysearchlabs.com/raystation/?gad_source=1&gad_campaignid=20151694501&gbraid=0AAAApn42h-qxkxBpMKv487lvBIIm4qsH&gclid=Cj0KCQjwo_PRBhDNARIsAEcVALXeyHQ0eLZVBVAQAeZxNN1Q8PUW7da_pOVkf9uGbB2ew4QYjiNN0ucaApsMEALw_wcB».

[54] RaySearch Laboratories, «Deep Learning Segmentation - White Paper, RaySearch Laboratories,» 2024. [Online]. Available: https://www.raysearchlabs.com/media/publications/white-papers/deep-learning-segmentation/](https://www.raysearchlabs.com/media/publications/white-papers/deep-learning-segmentation/.

[55] «skimage resize - https://scikit-image.org/docs/0.20.x/api/skimage.transform.html#skimage.transform.resize,» [Online].

[56] «SciPy ndimage - https://docs.scipy.org/doc/scipy/reference/ndimage.html,» [Online].

[57] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: with Applications in R, New York: Springer, 2013, doi: 10.1007/978-1-4614-7138-7.

[58] Y. Zhang, S. C. Huang, Z. Zhou, M. P. Lungren, S. Yeung, «Adapting Pre-trained Vision Transformers from 2D to 3D through Weight Inflation Improves Medical Image Segmentation,» *Proceedings of Machine Learning Research (ML4H 2022),* vol. 193, pp. 1-15, 2022, doi: 10.48550/arXiv.2302.04303.

[59] X. Zhu, H. Pan, B. Gundogdu, D. Jha, Y. Velichko, A. B. Murphy, A. Ross, B. Turkbey, A. E. Cetin, U. Bagci, «A Probabilistic Hadamard U-Net for MRI Bias Field Correction,» *arXiv,* p. arXiv:2403.05024, 2024, doi: 10.48550/arXiv.2403.05024.

[60] P. Subramaniam, T. Kossen, K. Ritter, A. Hennemuth, K. Hildebrand, A. Hilbert, J. Sobesky, M. Livne, I. Galinovic, A. A. Khalil, J. B. Fiebach, D. Frey, V. I. Madai, «Generating 3D TOF-MRA volumes and segmentation labels using generative adversarial networks,» *Medical Image Analysis,* vol. 78, p. 102396, 2022, doi: 10.1016/j.media.2022.102396.

[61] P. Andreev, A. Fritzler, D. Vetrov, «Quantization of Generative Adversarial Networks for Efficient Inference: A Methodological Study,» *arXiv,* p. arXiv:2108.13996, 2021, doi: 10.48550/arXiv.2108.13996.

[62] B. Rusanov, G. M. Hassan, M. Reynolds, M. Sabet, P. Rowshanfarzad, N. Bucknell, S. Gill, J. Dass, M. Ebert, «Transformer CycleGAN with uncertainty estimation for CBCT based synthetic CT in adaptive radiotherapy,» *Physics in Medicine & Biology,* vol. 69, n. 3, p. 035014, 2024, doi: 10.1088/1361-6560/ad1cfc.