



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea Magistrale
in Ingegneria Biomedica

Tesi di Laurea Magistrale

Studio di algoritmi e metriche per la selezione automatica di immagini finalizzate all'addestramento di modelli di AI

Relatori:

Prof. Filippo Molinari
Prof. Massimo Salvi

Candidata:

Rossana Rita Casaluci
292199

Tutor aziendale:

Dott.ssa Stefania Zara

Anno Accademico 2023/2024

Abstract

Cone Beam Computed Tomography (CBCT) is a medical imaging method that utilizes a cone beam of X-rays. This technique finds its primary application in the field of dentistry and maxillofacial imaging due to the compact size of the acquisition systems and the reduced radiation dose provided to the patient compared to traditional Computed Tomography (CT) imaging. Another crucial application of CBCT lies within Image-Guided Radiation Therapy for oncology patients. CBCT scans play a crucial role in verifying patient positioning and monitoring anatomical variations throughout the course of treatment, ensuring precision and accuracy in delivering radiation therapy. However, compared to traditional CT, CBCT produces images with reduced spatial resolution and is susceptible to a higher occurrence of artifacts.. For these reasons, an emerging area in medical imaging research involves the use of Generative Adversarial Networks (GANs) to generate synthetic CT (sCT) images from CBCT images. Considering these developments, this thesis points to identify methods and indicators capable of automatically evaluating images from datasets aimed at training a GAN for sCT image generation from CBCT scans. Differentiating the dataset based on the manufacturer of the systems used to acquire the CBCT scan (Varian Medical Systems, Elekta), 20 metrics were calculated in 3 concentric circular Regions of Interest (ROIs) – for a total of 60 features. These metrics were employed to assess quality, similarity, and correlation between deformed CBCT

scans and the corresponding CT scans obtained from a dataset of 124 patients from different hospital centers.

The values obtained were used as “input” to train three different Machine Learning (ML) methods – Random Forest (RF), k-Nearest Neighbor (kNN), and Support Vector Machine (SVM) – to perform an image classification task. For each ML method, three models with progressively reduced dimensionality were evaluated: a model trained on all 60 features obtained (Model 1), a model trained considering the larger ROI dimensions – 110 px radius (Model 2), and a model trained with the features of the ROI with a 110 px radius extracted using a feature importance analysis with RF (Model 3). For all the method implemented the dataset was divided into a 70% training set (TRS) and a 30% test set (TSS).

The performance in the classification task of the models 1 and 2 are above 98%, both on the TRS and the TSS, with all three ML methods. In models 3 for kNN and RF the performances were similar to those of higher-dimensional datasets, while the SVM method has reported performances between 96% and 98% for both TRS and TSS.

The results demonstrate the effectiveness of the evaluated features for training RF and kNN methods. Consistency in the performances obtained in the three evaluated models was observed, even after dimensionality reduction to mitigate complexity and overfitting risks. Among the ML techniques examined, the SVM method appears to be the most sensitive to the selection of training features.

Sommario

La Cone Beam Computed Tomography (CBCT) è un metodo di imaging medico che utilizza un fascio conico di raggi X. Questa tecnica trova la sua principale applicazione in ambito odontoiatrico e maxillo-facciale per via delle dimensioni compatte dei sistemi di acquisizione e per la ridotta dose di radiazioni fornita al paziente rispetto all'imaging tramite Computed Tomography (CT) tradizionale. Un'altra applicazione cruciale della CBCT risiede nell'Image-Guided Radiation Therapy per i pazienti oncologici. Le scansioni CBCT svolgono un ruolo cruciale nella verifica del posizionamento del paziente e nel monitoraggio delle variazioni anatomiche nel corso del trattamento, garantendo precisione e accuratezza nella somministrazione della radioterapia.

Tuttavia, rispetto alla CT tradizionale, la CBCT produce immagini con una risoluzione spaziale ridotta ed è più frequentemente soggetta ad artefatti. Per queste ragioni, un ambito emergente nella ricerca dell'imaging medico riguarda l'uso di reti neurali generative (Generative Adversarial Network, GAN) per la realizzazione di CT sintetiche (sCT) a partire da immagini CBCT.

Nell'ottica di questi sviluppi, il presente lavoro di tesi si pone come obiettivo l'identificazione di metodi e indicatori in grado di valutare automaticamente le immagini di dataset indirizzati all'allenamento di una GAN per la generazione di immagini sCT a partire da scansioni CBCT.

Distinguendo il dataset in base alla casa produttrice dei sistemi con cui è stata acquisita la scansione CBCT (Varian Medical Systems, Elekta), sono state calcolate 20 metriche in 3 regioni di interesse (ROI) circolari concentriche – per un totale di 60 features. Queste metriche sono state impiegate per valutare qualità, similarità e correlazione tra scansioni di CBCT deformate e le corrispettive CT ottenute da un dataset di 124 pazienti provenienti da differenti centri ospedalieri.

I valori ricavati sono stati utilizzati come “input” per addestrare tre diversi metodi di Machine Learning (ML) – Random Forest (RF), k-Nearest Neighbor (kNN) e Support Vector Machine (SVM) – per effettuare un task di classificazione delle immagini. Per ogni metodo ML, sono stati valutati tre modelli a dimensionalità progressivamente ridotta: un modello allenato sulla totalità delle 60 features ottenute (Modello 1), un modello allenato considerando la ROI di dimensioni maggiori – 110 px di raggio (Modello 2), e un modello allenato con le features della ROI con raggio 110 px estratte a seguito di un’analisi di feature importance con RF (Modello 3). Per tutti i metodi implementati, il dataset è stato suddiviso in Training Set e Test Set con proporzioni 70:30.

Le performance nel task di classificazione dei modelli 1 e 2 si attestano al di sopra del 98%, sia su TRS che sul TSS, con tutti i 3 metodi ML considerati; nel modello 3, per RF e kNN si ottengono performance analoghe ai modelli di dimensionalità maggiore, mentre il metodo SVM ha riportato performance tra il 96% e il 98% sia per il TRS che per il TSS.

I risultati ottenuti dimostrano l’efficacia delle features valutate per l’allenamento dei metodi RF e kNN. In particolare si è osservata consistenza nelle performance ottenute nei tre modelli valutati, anche a seguito della riduzione di dimensionalità per ridurre la complessità e mitigare i rischi di overfitting. Tra le tecniche di ML esaminate, il metodo SVM risulta essere il più sensibile alla selezione delle features di allenamento.

Indice

Abstract

Sommario

Introduzione	10
Capitolo 1 Cone-Beam Computed Tomography	13
1.1. Hardware della kV-CBCT	14
1.2. Acquisizione dell'immagine	16
1.3. Ricostruzione dell'immagine	17
1.4. Qualità dell'immagine.....	18
1.5. Ottimizzazione dell'esposizione	21
Capitolo 2 Machine Learning e GAN	22
2.1. Random Forest	24
2.2. k-Nearest Neighbors	26
2.3. Support Vector Machine.....	28
2.4. Generative Adversarial Network.....	29
Capitolo 3 Materiali e procedura sperimentale	31
3.1. Preparazione dataset.....	31
3.2. Valutazioni preparatorie	31
3.3. Estrazione delle immagini.....	33
3.4. Suddivisione del dataset.....	34
3.5. Determinazione delle ROI	35
3.6. Calcolo delle metriche	38
3.7. Step di calcolo.....	39
3.8. Classificazione delle immagini	40

3.9.	Descrizione dei modelli	42
3.9.1.	Modello 1. 60 features: ROI 110, ROI 85, ROI 60.....	42
3.9.2.	Modello 2. 20 features: ROI 110.....	42
3.9.3.	Modello 3. Features selezionate: ROI 110.....	43
Capitolo 4 Analisi dei dati e dei risultati		44
4.1.	Valutazione preliminare dei dati	44
4.2.	Valutazioni sulle metriche	46
4.3.	Performance dei modelli di Machine Learning.....	49
4.3.1.	Modello a 60 features: ROI 110, ROI 85, ROI 60	49
4.3.1.	Modello a 20 features, ROI 110	51
4.3.2.	Modello con features selezionate	52
4.4.	Metodo delle soglie.....	55
4.5.	Valutazioni su altro dataset	56
4.5.1.	Valutazione con i modelli Varian	57
4.5.2.	Valutazione con i modelli Elekta	58
4.5.3.	Nuovo allenamento	58
Capitolo 5 Conclusioni		62
Appendice A.....		64
Appendice B.....		66
Bibliografia		80

Introduzione

Contesto clinico e stato dell'arte

Il cancro prostatico è il secondo cancro più comunemente diagnosticato e la sesta causa principale di morte per cancro tra gli uomini, con un'incidenza stimata di 1 467 854 di nuovi casi e 397 430 morti al mondo nel 2022. In Italia nel 2022 si registrano 38 180 nuovi casi e una mortalità del 4.2% [1]. A livello globale si prevede un aumento dell'incidenza e delle relative morti, con proiezioni che indicano 2.3 milioni di nuovi casi e 740 000 morti entro il 2040, principalmente attribuibile all'aumento demografico e all'invecchiamento della popolazione [2]. Tuttavia, è stata osservata una diminuzione della mortalità correlata al carcinoma prostatico nella maggior parte dei Paesi ad alto sviluppo umano, suggerendo un'alta percentuale di sopravvivenza a lungo termine, che si riflette nella maggiore accessibilità a trattamenti quali chirurgia, terapia ormonale e radioterapia, insieme a miglioramenti nei trattamenti stessi [2,3]. Sebbene la scelta del metodo di trattamento si basi su diversi fattori, quali caratteristiche prognostiche, preferenze del paziente, stato di salute generale, presenza di altre patologie e aspettativa di vita [4], l'uso della radioterapia costituisce uno strumento efficace nella limitazione dell'incidenza di metastasi, della progressione locale e della somministrazione a lungo termine della terapia di deprivazione androgenica [3].

Fino agli anni '90 la radioterapia convenzionale era la tecnica principale nel trattamento del cancro alla prostata, utilizzando immagini di fluoroscopia e campi di trattamento standard per ridurre gli errori di localizzazione. Tuttavia, le limitazioni associate alla tecnica di imaging adottata comportavano l'inclusione di ampi volumi di organi critici adiacenti, come il retto e la vescica [4]. Oggi la pratica standard si avvale della Image-Guided Radiation Therapy (IGRT) per definire l'area da trattare, considerando le variazioni anatomiche e il posizionamento del paziente durante frazioni di trattamento successive, primariamente attraverso l'uso di imaging basato su tomografia computerizzata (CT) [5] e dell'imaging a risonanza magnetica (MRI), che offre un miglior contrasto dei tessuti molli [4,6–8].

Nella regione pelvica, i cambiamenti anatomici tra le frazioni di trattamento risultano particolarmente critici a causa delle considerevoli variazioni nel riempimento di vescica, tratto intestinale e retto, che, includendo fattori quali posizionamento del paziente e un potenziale errore umano nella fase di setup, causano rotazioni, traslazioni e deformazioni della prostata [9–11]. Dunque, risulta evidente l'importanza di adottare metodi image-guided per ottimizzare la precisione nel trattamento e minimizzare il rischio di errori nella definizione dei margini del volume da trattare [12].

Sebbene l'MRI offra una miglior definizione dei tessuti nella zona pelvica rispetto all'imaging basato su CT, l'integrazione delle informazioni MRI nella pianificazione del trattamento necessita la registrazione dei volumi dall'MRI alla CT, comportando potenziali errori durante il processo di fusione delle immagini [5].

Attualmente, nel contesto della IGRT, lo stato dell'arte prevede l'utilizzo della kV Cone-Beam Computed Tomography (CBCT), tecnologia di imaging medico sviluppata verso la fine degli anni '90 e adottata in radioterapia a partire dai primi anni 2000 [13], che si avvale di un fascio conico di raggi X [5,14,15]. In questi sistemi, la sorgente dei raggi X e il rivelatore ruotano attorno un fulcro fissato nel centro della regione di interesse, acquisendo proiezioni planari sequenziali del campo di vista (FOV). Con questa configurazione, una singola rotazione del gantry consente di acquisire dati sufficienti per la ricostruzione di un'immagine tridimensionale [5,15,16].

Questa tecnica nasce come alternativa alla CT tradizionale, che utilizza un fascio a ventaglio e una scansione a spirale, e, rispetto ai metodi di imaging precedentemente menzionati, si distingue per la capacità di fornire un'acquisizione più rapida con ridotte dosi di radiazione fornite al paziente, in un dispositivo più compatto e complessivamente meno costoso [16], che permette l'ottimizzazione del flusso di lavoro durante il trattamento [15].

I vantaggi associati alla CBCT hanno condotto ad un utilizzo principalmente nell'ambito odontoiatrico e nel distretto maxillo-facciale [17], ciononostante costituisce uno strumento essenziale nella IGRT per la sua capacità di visualizzare le variazioni anatomiche dei tessuti molli, consentendo un corretto posizionamento del paziente durante i trattamenti [9,13,18].

Sebbene i benefici della CBCT nella pratica clinica siano evidenti, è importante tener conto del fatto che, rispetto alla CT tradizionale, la CBCT produce immagini con risoluzione spaziale inferiore e con una qualità complessivamente più bassa a causa degli effetti di scattering dei raggi X dovuti alla forma a cono del fascio di radiazione [16,19,20]. La qualità dell'immagine, inoltre, viene ulteriormente ridotta dalla generazione di artefatti intrinseci allo strumento d'acquisizione, di cui si menzionano quelli relativi al fascio di raggi X e alla sua geometria (beam hardening, bassa risoluzione di contrasto, effetto di volume parziale, cone-beam effect) e allo scanner (artefatto ad anello) [16,21,22].

La qualità inferiore delle immagini CBCT ha condotto la ricerca nell'imaging medico verso l'uso dell'intelligenza artificiale per generare CT sintetiche (sCT), in particolare impiegando il Deep Learning con le Generative Adversarial Network (GAN) allenate con dataset multimodali, composti da immagini CT e immagini CBCT, [18,23,24].

In letteratura si osserva come le GAN siano già state già impiegate nella generazione di sCT partendo da acquisizioni MRI [8,24,25], ma recentemente la stessa strategia è stata adottata per migliorare la correttezza anatomica delle immagini CBCT [23] e il mapping impreciso delle unità Hounsfield e consentire, dunque, valutazioni di dose accurate [24,26–33], con risultati già promettenti per il distretto pelvi [18,27].

Dal momento che le architetture di Deep Learning necessitano di una vasta quantità di dati per ottenere performance ottimali [34], considerando che la valutazione manuale di una tale entità di dati può risultare complessa, il presente lavoro di tesi si pone come obiettivo l'identificazione di metodi e indicatori in grado di effettuare un'analisi automatica delle immagini CT e CBCT all'interno di dataset destinati all'allenamento di una GAN per la generazione di sCT a partire da scansioni CBCT.

Capitolo 1

Cone-Beam Computed Tomography

La Cone-Beam Computed Tomography è una tecnica di imaging radiografico che impiega un fascio di raggi X divergente di forma conica o piramidale per l'acquisizione di immagini tridimensionali delle strutture anatomiche [15,35,36]; nei dispositivi CBCT è presente un detector ad area che cattura un intero volume in una singola rotazione senza richiedere movimentazione del paziente [35]. Queste sono le differenze sostanziali rispetto alla CT spirale che sfrutta un fascio di raggi X a ventaglio e dei rilevatori lineari, per cui è necessario spostare continuamente il paziente lungo il supporto mentre il fascio di raggi X ruota attorno al paziente [35].

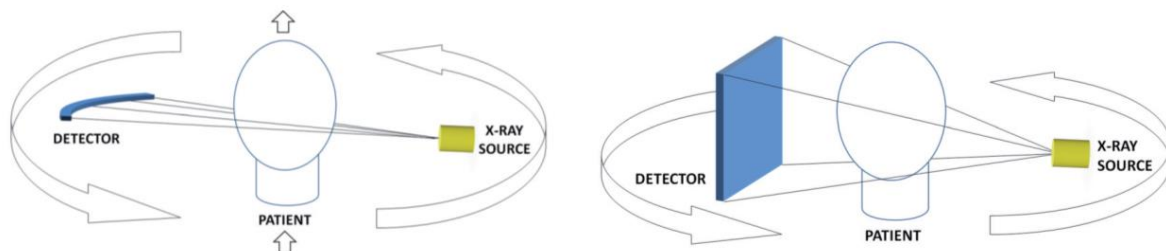


Figura 1. Differenza nel fascio di raggi X e del rilevatore. A sinistra: CT, a destra: CBCT (adattata da [37]).

I sistemi comprendono un acceleratore lineare (LINAC), che ospita una sorgente MV (MV-CBCT), a cui si aggiunge una sorgente kV (kV-CBCT) e il relativo rivelatore a pannello piatto, orientati ortogonalmente rispetto al fascio di trattamento MV [15,38,39]. La sorgente kV e il detector sono montati sul gantry del LINAC in modo che condividano un isocentro comune con l'unità di trattamento [15,38]. La configurazione kV-CBCT è quella principalmente utilizzata nella pratica clinica per evitare i problemi di contrasto dei tessuti molli che si riscontrano utilizzando la sorgente MV [15,38]: la miglior risoluzione di contrasto delle immagini kV-CBCT deriva dalla prevalenza dell'effetto fotoelettrico per energie dell'ordine del kilovolt [15], mentre ad energie dell'ordine del megavolt l'interazione dominante è l'effetto Compton [15,38–40] (fenomeno per cui, a seguito di un urto tra un fotone e un elettrone dei livelli più esterni di un atomo, il fotone devia la sua traiettoria cedendo energia cinetica all'elettrone, il quale viene espulso dall'atomo).

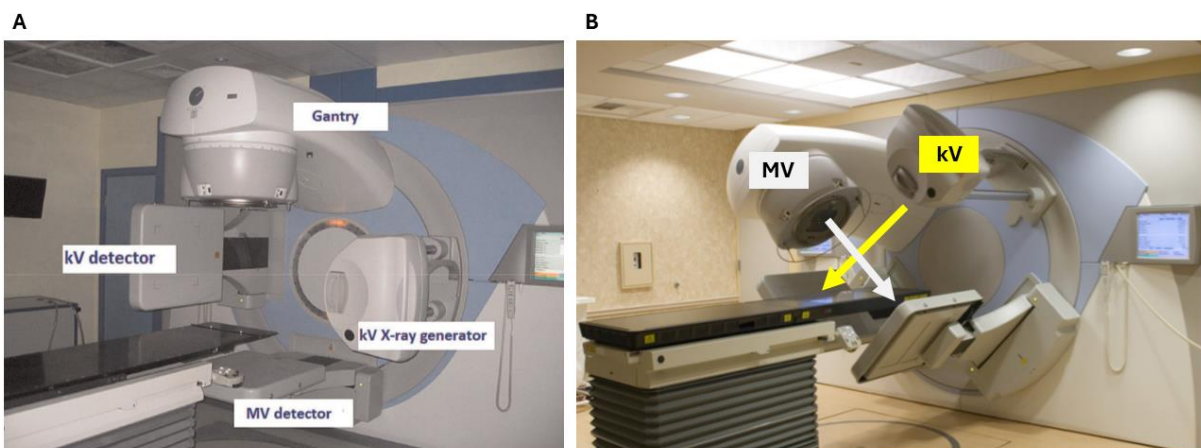


Figura 2. **A.** Esempio di sistema CBCT: Varian TrueBeam Linac [15]. **B.** Esempio di dispositivo CBCT (Elekta Synergy) con indicazione delle direzioni dei fasci MV e kV (adattata da [41])

1.1. Hardware della kV-CBCT

Gli elementi base di un sistema CBCT sono il tubo radiogeno, dove si ha la generazione dei raggi X, e il detector, dove i raggi X vengono convertiti in segnali elettrici.

Il tubo radiogeno è composto da un involucro in vetro contenente un catodo e un anodo, generalmente entrambi in tungsteno o in leghe di tungsteno, ai capi dei quali viene applicata una differenza di potenziale, in un range che va dalle poche decine fino al centinaio di kilovolt [17,36,39]. La generazione dei raggi X inizia nel catodo, un filamento nel quale viene iniettata

una corrente elettrica dell'ordine dei milliampere [42]; al passaggio della corrente, la temperatura nel filamento aumenta, inducendo il rilascio di elettroni per emissione termoionica [36]. Gli elettroni emessi dal catodo vengono accelerati verso l'anodo dalla differenza di potenziale, collidendo poi in corrispondenza della macchia focale, le cui dimensioni sono direttamente proporzionali con le dimensioni geometriche del fascio generato. La macchia focale è uno degli elementi che contribuisce alla nitidezza dell'immagine, nei sistemi CBCT ha dimensioni comprese tra i 0.5-0.8 mm [36,39]. I raggi X sono emessi in tutte le direzioni, ma l'assorbimento sull'anodo e nel tubo fa in modo che il fascio di elettroni esca dal tubo perpendicolarmente. Inoltre, l'anodo è leggermente inclinato (di un angolo che dipende dal dispositivo) per massimizzare l'uscita dalla finestra del tubo [36,39].

La dimensione del fascio, quindi del FOV, è determinata dal collimatore. Il collimatore circoscrive il fascio alla dimensione del sensore o alla dimensione della regione anatomica di interesse [17,39,42], limitando l'area di esposizione sul paziente, quindi la dose fornita [35,38]. La maggior parte dei sistemi CBCT ha varie dimensioni predefinite per il collimatore, conseguentemente ha varie dimensioni del FOV [17,36,39]. Il collimatore è in lega di piombo e presenta un'apertura (che può essere circolare o rettangolare) che consenta il passaggio dei raggi X solo sull'area di esposizione del paziente [36,42].

Rispetto ai tubi radiogeni utilizzati in radiografia 2D e in imaging CT, i tubi CBCT hanno differenti dimensioni dei collimatori, nei range dei fattori di esposizione e nella quantità di filtraggio del raggio [36].

Il raggio emesso dal tubo radiogeno è polienergetico [36,43]. La curva descritta dalle energie dei fotoni rilevati presenta una forma continua con dei picchi caratteristici di intensità relativa. L'andamento continuo nello spettro è determinato dalla radiazione di frenamento, per cui l'interazione degli elettroni accelerati con il materiale dell'atomo causa una perdita di energia cinetica che viene compensata con il rilascio di fotoni nella banda dei raggi X [36]. Lo spettro di frenamento si attenua attraverso un filtraggio intrinseco al tubo e un filtraggio aggiuntivo, utili per ridurre l'artefatto di beam hardening e la dose fornita al paziente: il primo è relativo all'interazione dei fotoni a bassa energia con il tubo stesso, il secondo consiste nell'uso di filtri in rame o in alluminio che impediscano l'uscita dal tubo dei fotoni a bassa energia [36]. L'energia media dello spettro è direttamente proporzionale allo spessore del filtro: i sistemi CBCT tipicamente utilizzano filtri di spessore 2.5 e i 10 mm [36].

I picchi discreti sono caratteristici del materiale componente l'anodo e si ottengono

dall'interazione del fascio di elettroni con gli atomi dell'anodo [36]: gli elettroni negli orbitali più vicini al nucleo vengono rimossi dai raggi X e le vacanze vengono colmate da elettroni dei livelli più esterni, più energetici, per cui l'energia in eccesso viene rilasciata in forma di fotoni.

Il gantry della maggior parte dei sistemi CBCT è realizzato come un braccio a C (C-arm) rotabile che alloggia tubo radiogeno e rivelatore [36]. La scansione si può effettuare in posizione supina, seduta o in piedi, in base allo scanner e alle necessità del paziente [16,37,43]: nei sistemi usati in ambito odontoiatrico il braccio ruota nel piano orizzontale, mentre negli scanner che sfruttano un posizionamento supino, tubo e detector ruotano sul piano verticale [36].

I detector attualmente in uso sono a pannello piatto, perché sono privi di distorsioni, hanno un'efficienza di dose più alta, un range dinamico più ampio e si possono produrre sia per FOV larghi che per FOV più stretti [36,37,43,44]. La maggior parte dei sistemi CBCT in commercio usano rilevatori a pannello piatto con uno strato di un materiale scintillatore, come ossalofuro di gadolinio e ioduro di cesio [36,43]. Più recentemente si utilizzano rilevatori CMOS con un FOV più largo, alta risoluzione, rapida lettura e basso rumore elettronico [36,43].

1.2. Acquisizione dell'immagine

Il principio di acquisizione delle immagini è comune tra i dispositivi CBCT: durante la scansione, tubo radiogeno e detector percorrono una traiettoria circolare con tempi di rotazione che possono variare tra i 10-40 s (ma esistano protocolli più veloci o più lenti [17,36,42]); dalla rotazione si producono diverse centinaia di proiezioni 2D [36,42].

Le differenze tra i sistemi CBCT riguardano i metodi di acquisizione e i parametri [36].

Una variabile è l'arco di rotazione: alcuni dispositivi consentono solo una rotazione parziale a causa dell'ingombro del C-arm, altri dispositivi permettono di effettuare sia una rotazione completa, sia una parziale [36]. La maggior parte dei dispositivi eseguono l'acquisizione compiendo una rotazione di 360° [16,35,38,41,42], sebbene la ricostruzione dell'intero FOV sia ottenibile con una rotazione di 180° aggiunta all'angolo del fascio [36]. L'ampiezza dell'arco di rotazione influenza la qualità dell'immagine e la dose fornita al paziente: in alcuni sistemi, un arco ridotto comporta tempi di scansione più brevi, risultando in una dose più bassa; tuttavia, una dose più bassa corrisponde ad un aumento del rumore [36].

Un'altra variabile riguarda la durata della scansione (tempo che intercorre tra la prima e l'ultima proiezione) e il tempo d'esposizione (tempo cumulativo durante la quale si effettua un'esposizione), che impattano sulla dose fornita al paziente, proporzionale al prodotto del

tempo di esposizione e corrente nel tubo (mA·s) [36]. Alcuni tubi radiogeni consentono di avere un'esposizione continua o pulsata: se l'esposizione è continua, durata della scansione e tempo d'esposizione coincidono [36]. L'esposizione pulsata viene adoperata per non avere esposizione tra le proiezioni, quindi il tempo di esposizione risulta molto minore rispetto alla durata della scansione, impattando sulla dose fornita al paziente [16,36,42,43]. Un'ulteriore vantaggio dell'esposizione pulsata risiede in una migliore risoluzione spaziale legata ad un minore effetto di movimento del gantry [36,44].

1.3. Ricostruzione dell'immagine

Durante il processo di ricostruzione dell'immagine, i dati grezzi provenienti dalle numerose proiezioni scansionate vengono assemblate insieme in un'immagine che rappresenta l'attenuazione relativa dei raggi X dei diversi materiali componenti l'oggetto scansionato [16,35,36,43]. Per un'accurata ricostruzione è necessario calibrare il sistema di imaging tramite oggetti di prova specifici del produttore [36].

Il volume digitale viene ricostruito come una matrice tridimensionale composta da voxel, solitamente isotropici, a cui viene assegnato un valore di grigio sulla base dell'attenuazione dei materiali in esso rappresentati [17,35,36].

In generale, i valori di grigio sono numeri interi dove i numeri più bassi corrispondono ai voxel più scuri [36]; il numero di valori di grigio assegnabili è determinato dalla profondità di bit dell'immagine, che nei sistemi CBCT è solitamente compresa tra 12 e 16 bit [16,37,42,43].

Nell'ambito dell'imaging CT, i toni di grigio vengono normalizzati in unità Hounsfield (HU), che rappresentano il coefficiente di attenuazione lineare dei raggi X in un voxel relativo al coefficiente di attenuazione dell'acqua. Nella scala Hounsfield, 0 HU corrisponde al coefficiente d'attenuazione dell'acqua, -1000 HU a quello dell'aria, mentre i materiali con una maggiore assorbimento dei raggi X presentano valori HU più alti. Tuttavia, l'utilizzo delle unità Hounsfield nella CBCT è limitato da vari fattori e l'incertezza risultante legata all'accuratezza e alla consistenza delle HU è spesso troppo grande per un'applicazione clinica di routine [36,45].

Da un punto di vista degli algoritmi di ricostruzione dell'immagine, si hanno tre tipologie di modelli: retroproiezione filtrata (FBP), tecniche di ricostruzione algebrica e metodi statistici [36]. In questa trattazione verrà brevemente approfondito il processo FBP, essendo il metodo

più comunemente adottato per i vantaggi legati alla breve durata di ricostruzione [36,39,46]. Prima di applicare l'algoritmo di ricostruzione, si eseguono degli step di pre-processing delle proiezioni 2D grezze per correggere le distorsioni legate alle variazioni nella corrente del rivelatore, al guadagno e a difetti nei pixel, e per rimuovere l'immagine latente delle proiezioni precedenti [36,43]. Dopodiché, si applica l'algoritmo FBP, il quale retroproietta le immagini pre-elaborate lungo i percorsi originali, ottenendo una serie di immagini raffiguranti la distribuzione di attenuazione dei raggi X nel volume [36]. Infine, si effettuano due step di filtraggio, uno con un filtro a rampa per correggere la sfocatura intrinseca al processo di retroproiezione, uno con un filtro di smoothing per ridurre il rumore ad alta frequenza amplificato dal filtro a rampa (tenendo però presente che questo compromette la risoluzione spaziale) [36].

1.4. Qualità dell'immagine

Nell'imaging medico la qualità di un'immagine viene valutata in base al compito specifico che si deve effettuare. In generale, è possibile effettuare quest'analisi attraverso quattro aspetti fondamentali: risoluzione spaziale, risoluzione di contrasto, rumore e artefatti.

La risoluzione spaziale è l'abilità di discriminare strutture separate nello spazio. Nell'imaging CBCT, i fattori che determinano la risoluzione spaziale sono la dimensione della macchia focale, la dimensione degli elementi del rivelatore, l'uso di uno filtro di smoothing e la dimensione dei voxel [36]. È possibile caratterizzare la risoluzione spaziale in termini della funzione di trasferimento della modulazione (MTF) che descrive l'abilità di un sistema di trasferire il segnale data una frequenza spaziale [36,47].

La risoluzione di contrasto è la capacità di distinguere tra le differenze in intensità di un'immagine. Anche in questo caso si ha una dipendenza da numerosi fattori, quali range dinamico del detector, profondità del bit dell'immagine, fattori di esposizione [36]. La percezione del contrasto, inoltre, dipende dalle impostazioni del display che possono essere gestite per ottimizzare la visualizzazione delle strutture anatomiche o delle lesioni [36]. Tra queste si trova la trasformazione window/level, la quale permette di visualizzare solo una porzione del range completo dei livelli di grigio, definendo l'ampiezza della finestra (il numero di livelli considerati nella visualizzazione) e il livello della finestra (il valore centrale dell'ampiezza della finestra).

Per rumore, tipicamente, si intende la variabilità casuale dei valori dei voxel nell'immagine. Le principali sorgenti di rumore nelle immagini radiografiche sono il rumore quantico e quello dovuto all'elettronica: il primo è causato dalla natura intrinsecamente casuale delle interazioni durante la produzione e attenuazione dei raggi X, il secondo è dovuto alla conversione e trasmissione dei segnali del detector [36].

È essenziale valutare questi parametri nel complesso, o perché sono interconnessi (ad es.: solitamente si effettua un compromesso tra rumore e risoluzione spaziale, poiché i fattori che migliorano un parametro, degradano l'altro), o perché presi singolarmente non consentono di effettuare una valutazione completa (ad es.: il contrasto, essendo una caratteristica generale dell'immagine, necessita della MTF per valutare la capacità del dispositivo di distinguere le caratteristiche più fini) [36].

Le sorgenti e le tipologie di artefatti nella CBCT sono molteplici. Una delle principali sorgenti di artefatti è lo scatter dei raggi X che provoca striature e ombreggiature sull'immagine. Questo tipo di artefatto nasce dalle inconsistenze nei dati associate ai fotoni di raggi X che subiscono interazioni di Compton all'interno del paziente [16,36]. È possibile attenuare i disturbi utilizzando griglie antiscatter, comportando, però, un aumento di dose [19,24,29,36,39]. Esistono, inoltre, algoritmi di correzione dello scattering [36].

Un artefatto molto comune ai dispositivi CT è il beam hardening, legato alla natura policromatica del fascio di raggi X [16,21,43]: man mano che il fascio attraversa un oggetto, la sua energia media aumenta perché i fotoni a energia inferiore vengono assorbiti più rapidamente rispetto a quelli ad energia più alta, causando un "indurimento" del fascio [21,22,36]. Il beam hardening, dunque, causa una distorsione nella rappresentazione di eventuali strutture metalliche presenti a causa dell'assorbimento differenziale (cupping artifact) e striature e bande scure tra due oggetti densi [16,22].

Un altro artefatto comune nella tomografia è l'effetto di volume parziale, che si verifica quando la risoluzione dei voxel è maggiore della risoluzione spaziale o di contrasto dell'oggetto, causando una rappresentazione non accurata del tessuto o dei confini [16,21,22].

Alcuni artefatti sono correlati allo scanner, come la presenza di anelli concentrici sull'immagine (ring artefact), causato da imperfezioni nella rilevazione dello scanner o una calibrazione non adeguata [16,22,37].

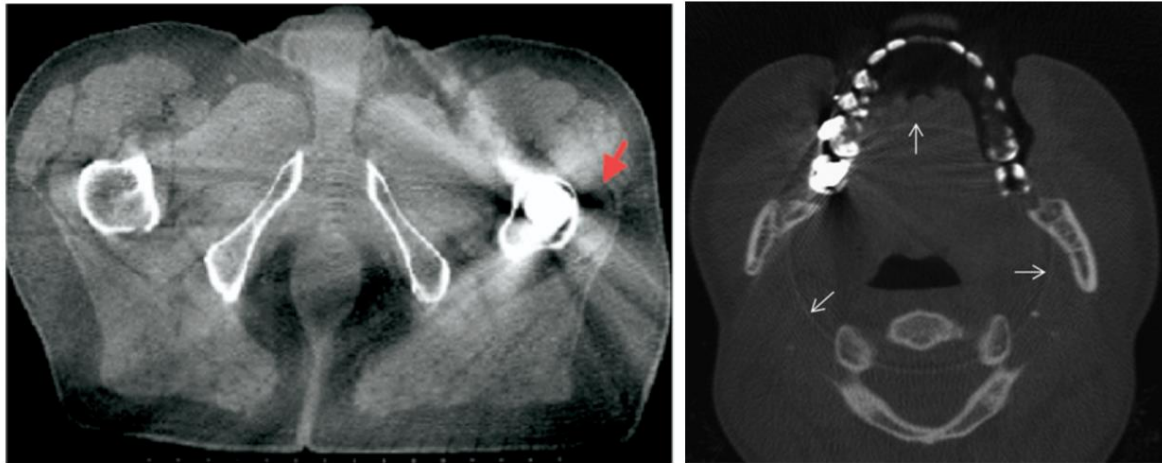


Figura 3. Esempi di artefatto di beam hardening: a sinistra, distretto pelvi (adattata da [48]); a destra: artefatto di beam hardening in ambito odontoiatrico [24]

Un artefatto specifico della CBCT è il cone-beam effect, per cui la divergenza del fascio durante la rotazione della sorgente attorno al paziente fa sì che l'attenuazione rilevata dal rilevatore nelle zone più periferiche irradiate sia ridotta rispetto a quella delle zone più centrali, risultando in striature e maggior rumore periferico (**Figura 4**) [16].

Si menzionano, infine, gli artefatti causati dal movimento del paziente [16,22].

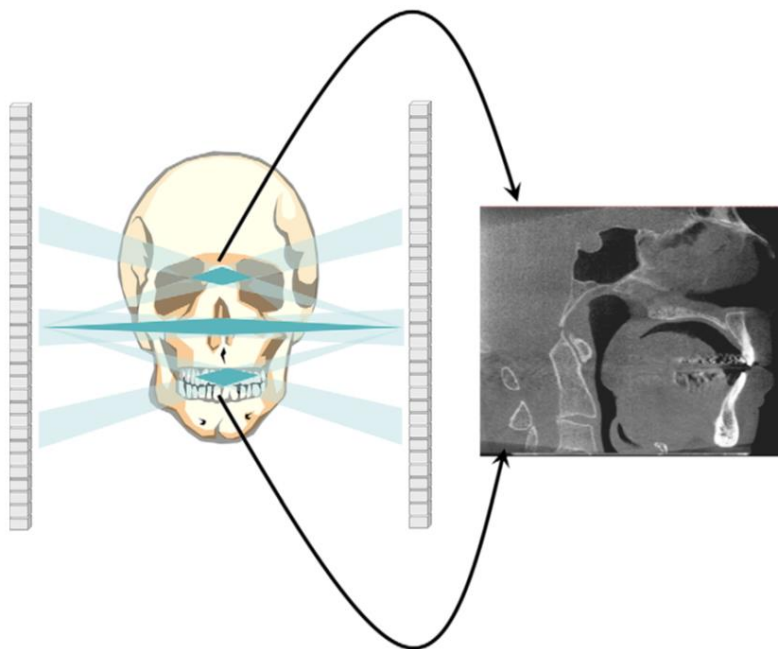


Figura 4. Artefatto da cone-beam in ambito odontoiatrico [16].

1.5. Ottimizzazione dell'esposizione

La natura volumetrica dell'imaging CBCT comporta una quantità importante di dose fornita ai tessuti immediatamente adiacenti al volume di trattamento [38]. Ci sono variazioni distinte nella dose efficace tra le unità CBCT, che possono essere attribuite a fattori come il FOV, l'impostazione di mA, la tensione di picco (anche se in molti sistemi è fissa), il tempo di scansione (inclusa la dose a impulsi rispetto a quella continua), la sensibilità del sensore e il numero di acquisizioni d'immagine [35,36].

In termini di ottimizzazione, il parametro di imaging più semplice da controllare è la dimensione del FOV: FOV più grandi aumentano la dose di radiazioni per il paziente [35,36], e aumentano la quantità della radiazione dispersa che raggiunge il detector, aumentando, in ultimo, rumore e artefatti; d'altra parte, FOV più piccoli aumentano i troncamenti e la formazione di ombreggiature che solitamente non si riesce a compensare tramite gli algoritmi di ricostruzione [36]. In ogni caso, il FOV dovrebbe essere mantenuto il più piccolo possibile, in modo che comprenda la sola ROI.

Capitolo 2

Machine Learning e GAN

Il Machine Learning (ML) è una branca della computer science nata dall'approfondimento del pattern recognition e dalla teoria computazionale dell'apprendimento nel campo dell'intelligenza artificiale, della quale è un sottogruppo [49]. Il ML comprende l'insieme di tecniche in grado di migliorare l'esecuzione di un task per mezzo di un apprendimento automatico dai dati forniti, senza la necessità di una programmazione esplicita [50].

È possibile classificare le tecniche di ML in base al tipo di apprendimento, che può essere supervisionato, non supervisionato, semi-supervisionato o rinforzato.

Nell'apprendimento supervisionato, il modello riceve i dati in input e le relative etichette di output. Durante l'allenamento, il modello impara a mappare gli input agli output, regolando i suoi parametri per minimizzare l'errore tra le previsioni e le etichette di output [49]. I dati in input vengono suddivisi in un Training Set (TRS) e in un Test Set (TSS). Il modello impara a partire dai dati del TRS e testa il suo apprendimento sui dati del TSS non etichettati, cercando di ottenere l'accuratezza più alta possibile [49]. Solitamente si utilizza questo tipo di apprendimento per eseguire task di classificazione e di regressione. Alcuni esempi di algoritmi con allenamento supervisionato sono Decision Tree e Support Vector Machine [49].

Nell'apprendimento non supervisionato non vengono fornite etichette di output, per cui il

modello determinerà autonomamente somiglianze tra i dati in input o ne identificherà pattern ricorrenti. I task generalmente svolti sono il clustering, la riduzione della dimensionalità e l'analisi dell'insieme di dati [49]. Gli algoritmi principali dell'apprendimento non supervisionato sono il clustering con il k-Means e la Principal Component Analysis (PCA) [49]. L'apprendimento semi-supervisionato combina elementi dall'allenamento supervisionato e dall'allenamento non supervisionato. Non tutti i dati di input avranno la corrispondente etichetta di output [49], dunque si sfruttano le etichette a disposizione per il processo di allenamento e si estrae l'informazione implicita nei dati non etichettati per migliorarne le prestazioni. Un esempio di algoritmo con questo apprendimento sono i modelli generativi [49].

Nell'allenamento rinforzato il modello apprende interagendo con un ambiente in cui si vuole massimizzare le *rewards* o minimizzare le *penalties* [49]: esso prende delle decisioni sequenziali e riceve dei feedback dopo il completamento di ogni azione, che influenzerà le decisioni successive.

Un limite delle tecniche di Machine Learning risiede nella gestione di grandi quantità di dati che causa una saturazione nelle performance [51]. In questo contesto emerge il Deep Learning (DL), sottocategoria delle tecniche di Machine Learning [50–52] in grado di svolgere task complessi, utilizzando numerosi livelli di unità d'elaborazione non lineari in cui le caratteristiche significative dei dati forniti vengono individuate in modo automatico [51,52]. Le reti di DL sono progettate per emulare il funzionamento del cervello umano [51]; i blocchi fondamentali che le costituiscono sono neuroni di input, layer nascosto e neuroni di output. A partire da questi elementi le differenti architetture sviluppate si distinguono per dimensione dei layer, profondità della rete, tipologia di collegamenti tra i vari livelli [52].

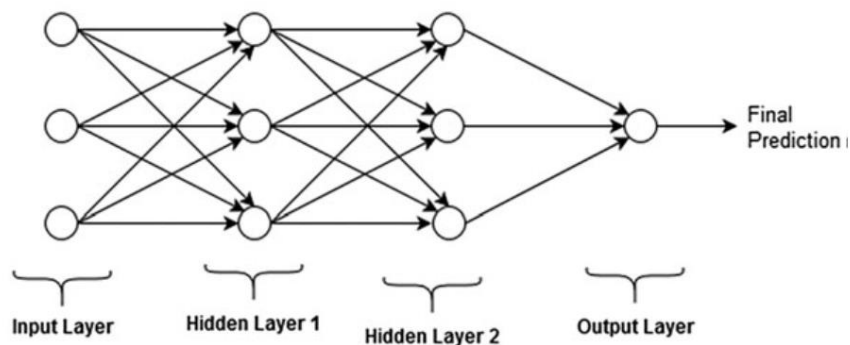


Figura 5. Esempio di struttura di una Deep Neural Network

Anche nelle tecniche di DL si può distinguere un addestramento supervisionato, non supervisionato, semi-supervisionato o rinforzato [51,52]. Nel medical imaging l'allenamento supervisionato è il più utilizzato e l'architettura più diffusa che si basa su questa modalità di allenamento è la Convolutional Neural Network (CNN) [52]. I quattro componenti fondamentali di una CNN sono i layer convoluzionali, nei quali vengono applicati dei kernel per generare la mappa delle features di output, i layer di *pooling*, che si occupano del sottocampionamento delle mappe delle features, la funzione di attivazione, che introduce non linearità nel modello, e il livello *Fully Connected*, dove la mappa delle features viene vettorializzata e l'output che si ottiene è quello finale della CNN [52].

Nell'ambito medico, le CNN assolvono a task quali classificazione, detection e segmentazione, ma possono essere implementate all'interno di architetture più complesse per eseguire altri compiti. Un esempio è quanto avviene nelle GAN (Generative Adversarial Network), reti di DL che si occupano della generazione di immagini sintetiche, o come supporto all'allenamento di un classificatore qualora una classe sia poco rappresentata, o come strumento per convertire (o tradurre) le immagini, come accade nella generazione di immagini CT sintetiche a partire da scansioni CT e CBCT.

Di seguito verranno approfondite le tecniche utilizzate in questo lavoro di tesi e si concluderà con una descrizione delle Generative Adversarial Network, il cui uso è l'obiettivo ultimo del progetto in cui si inserisce questa trattazione.

2.1. Random Forest

Il Random Forest (RF) è un metodo di ensemble nato per migliorare le performance del Decision Tree (DT), un algoritmo ML che può effettuare task di classificazione e di regressione. Per meglio comprendere il RF, è necessario esaminare il principio di funzionamento del Decision Tree, che costituisce gli elementi base del Random Forest.

Il Decision Tree è un metodo basato su una struttura ad albero, in cui si distinguono dei nodi e degli archi [53]. I nodi costituiscono i punti di decisione nei quali l'informazione viene suddivisa in base a determinate caratteristiche o criteri. Il processo decisionale inizia nel primo nodo, chiamato *nodo radice*, dove si ha la prima suddivisione dell'intero dataset. I nodi intermedi che originano dalla radice prendono il nome di *nodi di test*; ogni nodo di test ha un arco in entrata da un nodo genitore e uno o più archi che collegano i nodi figlio. I nodi terminali

dell'albero decisionale sono chiamati *nodi foglia* e contengono l'output predetto.

Il processo di decisione e suddivisione del dataset che si verifica nei nodi si realizza attraverso il calcolo dell'entropia e del Gini Index [53] (una misura di casualità o disomogeneità di una distribuzione di classi, dove 0 indica una maggiore omogeneità, 1 indica una distribuzione più casuale delle classi). L'obiettivo di ogni nodo è trovare la suddivisione che minimizzi questi due criteri per ottenere un insieme di dati il più omogeneo possibile.

L'ampio uso di questa tecnica è legato ai vantaggi che essa offre anzitutto in termini di chiarezza e semplicità nella comprensione del funzionamento, se confrontata con altri classificatori [53]. È un algoritmo che non richiede selezione di parametri, né ipotesi a priori, sufficientemente capace di gestire grandi quantità di dati, in grado di elaborare dati di formati differenti; inoltre, può fornire risultati rapidi rispetto ad altre tecniche, con alte performance [53]. Di contro, il DT ha una forte propensione all'overfitting, incline ad una scarsa generalizzazione; in aggiunta, gli alberi decisionali sono particolarmente suscettibili a variazioni nei dati di addestramento, rendendoli instabili.

Per superare i limiti di un singolo albero decisionale, è possibile utilizzare un approccio ensemble, il Random Forest, che consiste nell'addestramento di più alberi decisionali su un sottoinsieme dei dati appartenenti allo stesso dataset. Nel task di classificazione, gli output dei DT che compongono la "foresta" vengono aggregati tramite *majority voting* o *averaging* [54]. I singoli DT vengono generati dall'unione di due approcci, il bagging e la random selection of features [55], che consentono una migliorata stabilità e precisione nel modello predittivo. Il *bagging* (bootstrap aggregating) consiste nella generazione di diversi TRS attraverso un campionamento casuale (*resampling*) che può prevedere anche la presenza di duplicati degli elementi originali (*replacement*). La *random selection of features* introduce casualità nell'addestramento tramite una selezione casuale delle variabili utilizzate per la previsione, così da evitare la produzione di alberi correlati tra di loro [53].

Contrariamente al DT, il RF richiede la selezione di parametri, quali il numero di alberi, il numero di elementi all'interno dei nuovi TRS creati tramite bagging e il numero massimo di variabili che ogni DT può utilizzare.

Il RF è un metodo che può fornire previsioni molto accurate, molto robusto rispetto all'overfitting (sebbene si debba fare attenzione al numero di DT e alla profondità massima degli alberi), in grado di gestire dataset numerosi e con classi sbilanciate.

Gli svantaggi sono principalmente correlati alla numerosità dei DT che lo compongono.

Rispetto ad altri metodi, si ha una complessità interpretativa più alta, sia in termini di modello, che in termini di interpretazione delle previsioni; inoltre, può risultare computazionalmente esigente quando l'input è un dataset molto ampio o il modello prevede un gran numero di alberi.

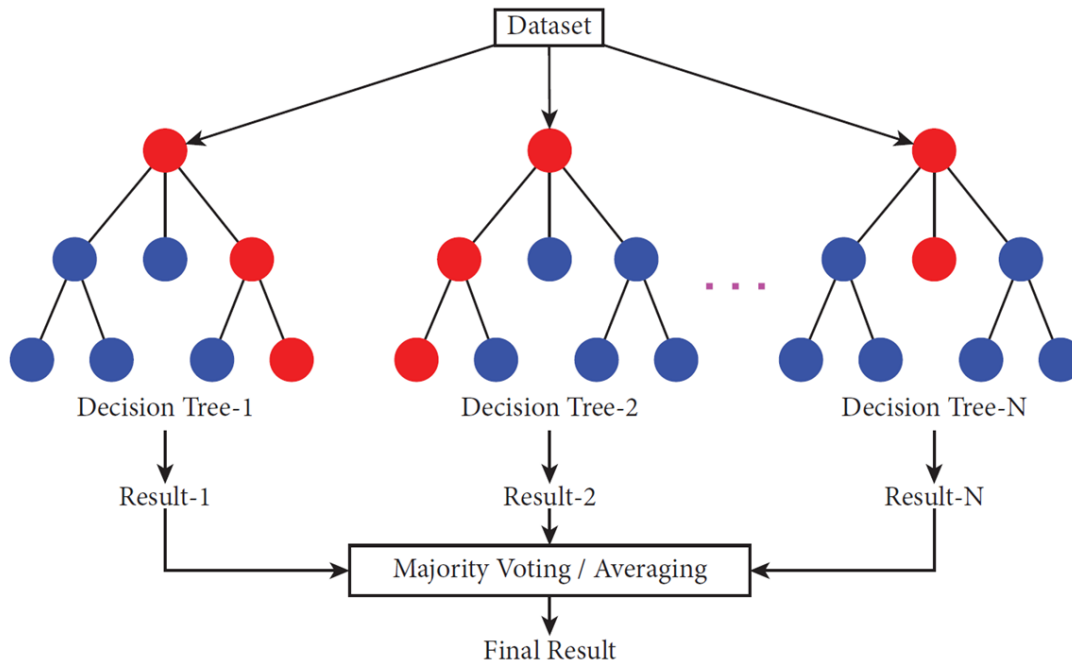


Figura 6. Schematizzazione del funzionamento del Random Forest: il dataset viene suddiviso in N subset negli N alberi decisionali componenti la foresta; in ogni nodo viene presa una decisione che si conclude in un output nei nodi foglia. L'output finale si ottiene per *majority voting* o *averaging* tra i risultati di ogni DT [54].

2.2. k-Nearest Neighbors

Il k-Nearest Neighbors (kNN) è un algoritmo di apprendimento supervisionato principalmente usato nei problemi di classificazione. Esso basa l'appartenenza alla classe sulla valutazione dell'esempio di allenamento più vicino (*neighbor*) o più vicino in una determinata regione. Per un nuovo dato in ingresso, vengono individuati i k vicini più prossimi e la classe più frequentemente rappresentata nei vicini determinerà la classificazione del nuovo dato [56]. La metrica maggiormente adoperata per calcolare la distanza tra i punti di dati è la distanza Euclidea, ovvero la distanza geometrica tra due punti. È possibile selezionare altre metriche, di cui si citano la distanza Manhattan (somma delle differenze assolute tra le coordinate dei punti) e la distanza di Minkowski (generalizzazione della distanza Euclidea e Manhattan).

Il parametro k determina il numero di punti vicini da tenere in considerazione; solitamente valori più grandi di k conducono a risultati più precisi perché riducono il rumore complessivo,

sebbene questo non sia sempre garantito [56]. Il valore di k si può determinare dall'osservazione del dataset o utilizzando la cross-validation [56].

L'algoritmo kNN si può riassumere in un primo step in cui si memorizza il TRS, seguito da una serie di operazioni, eseguite per ogni dato privo di etichetta, in cui si calcola la distanza con tutti i punti di allenamento, si identificano i k punti più prossimi, si assegna al nuovo dato la classe prevalente tra i suoi k vicini.

I fattori che influenzano le prestazioni del kNN sono i valori di k , la metrica scelta per il calcolo della distanza e la normalizzazione dei parametri [56].

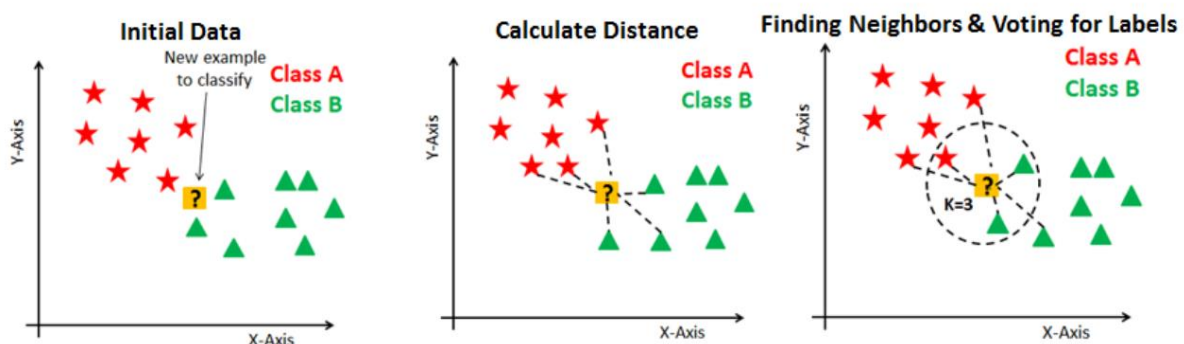


Figura 7. Rappresentazione grafica degli step del kNN: all'inserimento di un nuovo dato, si calcolano le distanze del dato con gli N punti più vicini; al nuovo dato verrà assegnata la classe predominante tra gli N esempi più vicini (adattata da [57]).

Il kNN è noto per la sua semplicità, sia da un punto di vista della comprensione dell'algoritmo in sé, sia da un punto di vista dell'interpretazione dei risultati. Trattandosi di un algoritmo non parametrico [56], esso non fa nessuna assunzione sulla distribuzione dei dati, rendendolo un metodo flessibile per vari problemi di classificazione. In ultimo, rispetto ad altri algoritmi, è meno influenzato dalla presenza di dati rumorosi o outlier, esattamente in virtù del fatto che la predizione dipende dai punti più prossimi.

Gli svantaggi legati a questa tecnica riguardano la dispendiosità computazionale necessaria per il calcolo delle distanze di ogni esempio di addestramento; questo si riflette anche sulla velocità di esecuzione, che può risultare bassa nel caso di dataset ampi. Il kNN è definito un algoritmo

di apprendimento “pigro”, poiché non apprende dai dati d’addestramento, ma li memorizza per effettuare la classificazione dei nuovi input [56].

2.3. Support Vector Machine

Il Support Vector Machine (SVM) è un algoritmo di apprendimento supervisionato utilizzato principalmente per effettuare task di classificazione binaria. Nei problemi lineari, l’algoritmo SVM si usa per effettuare predizioni attraverso la ricerca di un iperpiano, che separi il dataset o le classi, riducendo al minimo l’overfitting dei dati [58]. Gli esempi di training vengono proiettati in un nuovo spazio in cui gli oggetti che appartengono a classi differenti sono linearmente separabili. Tra le varie soluzioni possibili, l’iperpiano che si cerca è quello che massimizza il margine tra le classi [58], quindi la distanza tra l’iperpiano e i punti più vicini che appartengono ad ogni classe, che prendono il nome di vettori di supporto.

Nei problemi non lineari è possibile consentire degli errori nella classificazione, quindi non rendendo necessaria la divisione esatta della frontiera decisionale tra le due classi. In questi casi si parla di *soft margin classification* [58], per cui si usa una trasformazione non lineare, il *kernel*, che effettua una mappatura del problema in uno spazio delle variabili a più alta dimensione, dove il TRS è linearmente separabile. Se il TRS non è linearmente separabile, è possibile inserire delle variabili di scarto che consentano la classificazione errata di esempi difficili o rumorosi. Il parametro che gestisce questo aspetto è il termine di penalità C, che definisce la quantità di violazioni dei margini ammesse; valori elevati di C corrispondono a margini più stretti, con un numero di errori basso, mentre valori bassi corrispondono a margini più larghi, con un numero di errori concesso più elevato. È possibile scegliere tra un kernel lineare, gaussiano, RBF, polinomiale o sigmoide [58,59].

Il Support Vector Machine offre flessibilità nei dati, grazie alla possibilità di selezionare un kernel che possa gestire problemi non lineari [61], e buone capacità di generalizzazione. Tuttavia, l’interpretazione dei modelli può risultare complessa [61] e potrebbe non essere adeguato per dataset di grandi dimensioni, che potrebbero comportare costi computazionali molto elevati [59].

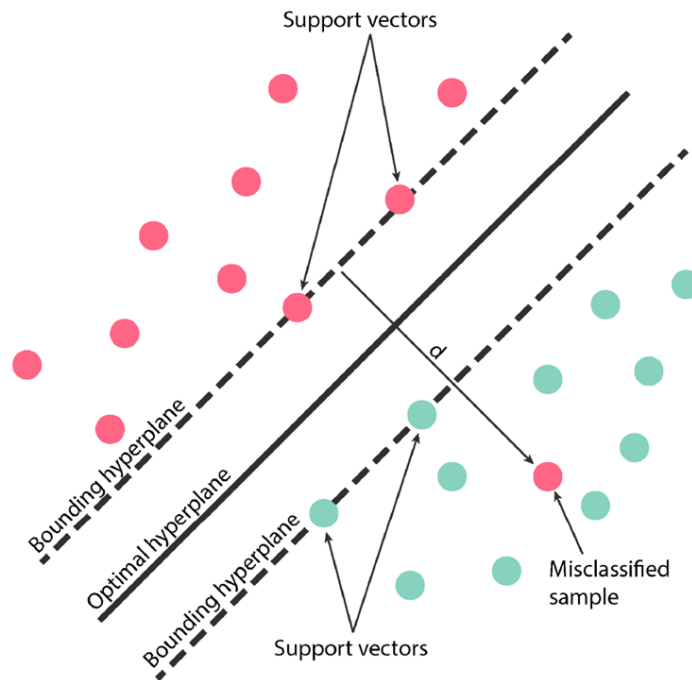


Figura 8. SVM: l'iperpiano ottimo è equidistante dagli iperpiani limite descritti dai vettori di supporto; la distanza tra i due iperpiani limite è detta margine [60].

2.4. Generative Adversarial Network

La GAN è un'architettura di DL costituita da due network, uno generativo e uno discriminativo, che interagiscono attraverso un processo avversario, dove il generatore converte l'immagine di partenza nell'immagine target, mentre il discriminatore valuta se l'immagine sintetica è reale o falsa [31]; la rete raggiunge la convergenza quando generatore e discriminatore raggiungono uno stato di equilibrio dinamico. La generazione di immagini sintetiche può essere descritta come un problema di conversione di immagini, in cui si crea un modello che mappi un'immagine di origine su un'immagine di destinazione [31].

Il modello GAN *pix2pix* è il primo modello *general-purpose* progettato per svolgere questa traduzione tra immagini; è una tipologia di GAN condizionali, dove la generazione dell'output è condizionata dall'input, ovvero l'immagine sorgente [62,63].

L'allenamento di generatore e discriminatore avviene simultaneamente: il discriminatore riceve un'immagine sorgente e un'immagine target, il suo compito sarà determinare se l'immagine target costituisce una trasformazione plausibile dell'immagine sorgente; il generatore è allenato tramite una funzione di perdita chiamata *adversarial loss*, calcolata in base alla capacità del discriminatore di distinguere correttamente le immagini reali da quelle sintetiche [62].

Tipicamente una GAN pix2pix si compone di un generatore implementato con un modello encoder-decoder simmetrico usando un'architettura U-Net e di un discriminatore realizzato con una CNN che effettui un task di classificazione [24,62]. La U-Net del generatore realizza anzitutto una fase di *encoding* (*downsampling*), durante la quale l'immagine viene ridotta di dimensione per estrarre le features a scale differenti, e successivamente una fase di *decoding* (o *upsampling*), dove l'immagine viene riportata alle dimensioni originali; una caratteristica importante di quest'architettura risiede nella presenza delle *skip-connections* che collegano i layer di encoding e decoding preservando le informazioni nelle diverse scale di risoluzione [62]. La **Figura 9** ritrae un esempio di GAN pix2pix.

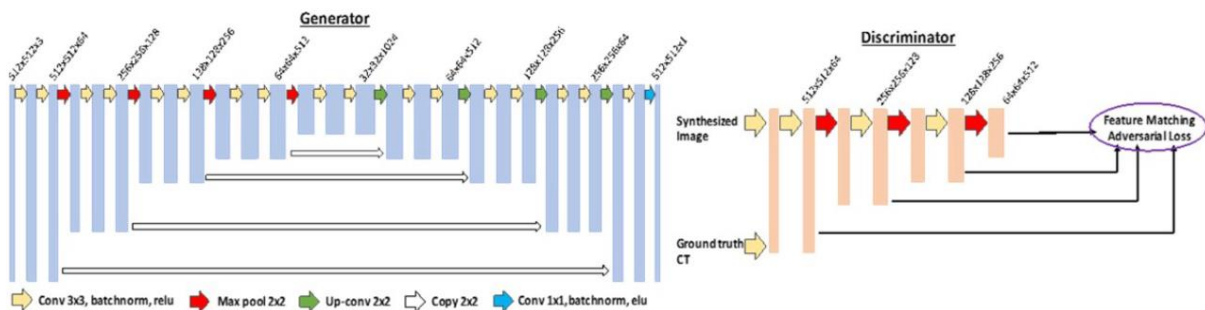


Figura 9. GAN pix2pix. Il generatore è realizzato con una U-Net: gli step di downsampling sono collegati con gli step di upsampling tramite le skip-connection. Il discriminatore è un classificatore CNN [24].

Un'applicazione delle GAN pix2pix nell'imaging medico è la generazione di CT sintetiche per migliorare la qualità delle immagini CBCT [24]. Le immagini in input al generatore saranno coppie di immagini CT e CBCT corrispondenti, di cui l'immagine CT costituisce il ground truth: il generatore crea un'immagine CT sintetica a partire dall'input CBCT e il discriminatore cercherà di distinguere le immagini sintetiche e quelle reali.

Capitolo 3

Materiali e procedura sperimentale

La procedura adottata è finalizzata alla classificazione di immagini attraverso tre metodi di Machine Learning. L'allenamento è effettuato sulla base di 20 metriche valutate in 3 regioni di interesse circolari concentriche con lo scopo di confrontare le immagini CBCT deformate rispetto alle corrispondenti immagini CT.

3.1. Preparazione dataset

Il dataset di immagini utilizzato in questo lavoro di tesi proviene dalle scansioni CT e CBCT del distretto pelvico di 124 pazienti anonimizzati provenienti da quattro differenti centri ospedalieri. Di questi, 15 pazienti non sono stati valutati in quanto destinati al successivo testing della GAN.

3.2. Valutazioni preparatorie

I dati iniziali a disposizione includono i file DICOM relativi alle scansioni CT e i file MetaImage e Raw relativi alle scansioni CBCT deformate ricavate da RayStation[®] di RaySearch Laboratories (software di pianificazione del trattamento radioterapico). In particolare, la generazione delle CBCT deformate avviene attraverso un processo chiamato *Deformable Image Registration* (DIR), eseguito dal software con lo scopo di allineare le

immagini CBCT con le immagini CT. La DIR sfrutta un algoritmo proprietario sviluppato da RaySearch Laboratories, ANACONDA, che risolve un problema di ottimizzazione non lineare: l'output è una funzione target composta da una somma pesata di un termine di similarità dell'immagine, un termine di regolarizzazione della griglia dell'immagine da deformare e un termine di penalità [64]. Trattandosi di un problema di ottimizzazione, l'obiettivo è minimizzare l'errore complessivo calcolabile attraverso la funzione target.

Per ciascun paziente è stata effettuata una suddivisione in base al produttore del sistema utilizzato per le acquisizioni CBCT: due centri ospedalieri hanno impiegato sistemi Varian Medical Systems, mentre gli altri due centri hanno utilizzato i sistemi di Elekta (da ora indicati rispettivamente come Varian ed Elekta). Tale suddivisione è stata effettuata per avere un maggior controllo dei processi, dal momento che visivamente la qualità delle immagini ottenute dai sistemi dei due produttori è notevolmente diversa, come mostrato nella **Figura 10**.

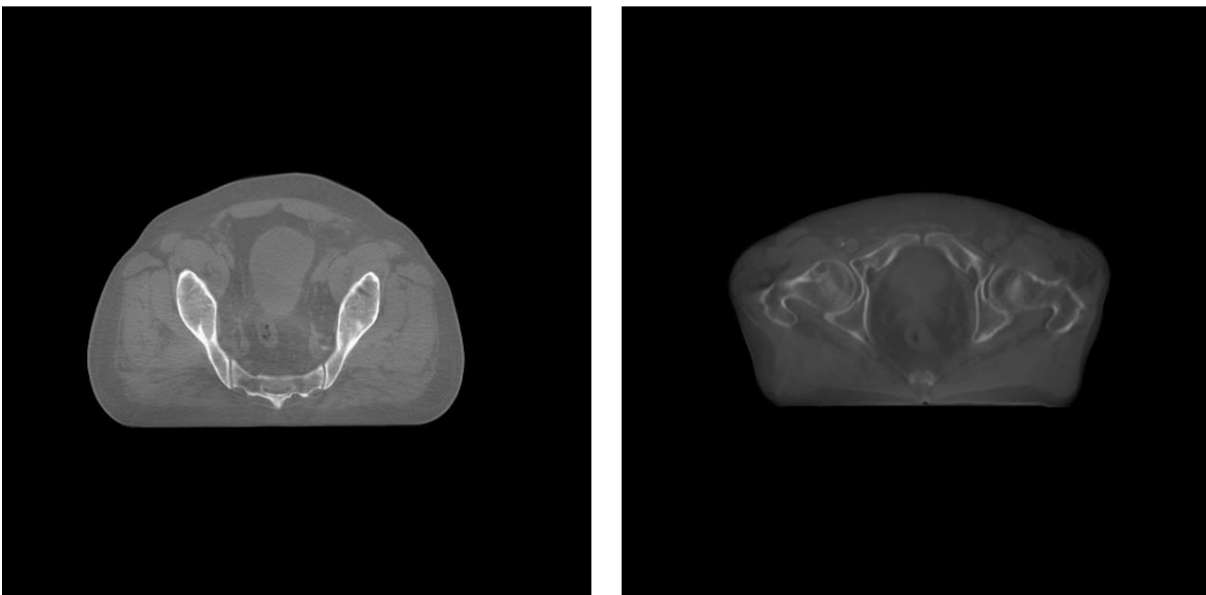


Figura 10. Scansioni CBCT dei produttori Varian Medical Systems (a sinistra) ed Elekta (a destra). Rispetto alle immagini Elekta, le immagini Varian risultano visivamente più definite e meno soggette ad artefatti.

Nei 109 pazienti presi in considerazione è stata condotta una raccolta preliminare di dati e un controllo della corrispondenza dei file a disposizione. Durante questa fase è stata verificata la simultanea presenza dei file DICOM e dei file MetaImage e Raw, necessari per estrarre le immagini per il calcolo delle metriche; inoltre, sono state identificate informazioni come il numero di frazioni CBCT presenti per paziente, il numero delle slice delle scansioni CT e

CBCT. A seguito di questa analisi, ulteriori sette pazienti sono stati esclusi a causa di incongruenze riscontrate tra le scansioni CT e i file contenenti le informazioni sulla CBCT deformata.

Tra i pazienti Varian esaminati, in 20 casi sono disponibili scansioni multiple di CBCT, con un numero di frazionamenti compreso tra 2 e 5. Inoltre, sono state individuate 20 acquisizioni a FOV ridotto: 11 di queste appartenevano a pazienti in cui è stata effettuata solo la scansione a FOV ridotto, mentre le restanti 9 provenivano dai casi di scansioni multiple.

3.3. Estrazione delle immagini

Per ciascuno dei 102 pazienti rimanenti è stato eseguito un processo di estrazione delle immagini nel formato e nelle caratteristiche necessarie per essere messe in input nell'addestramento della GAN. Di seguito sono indicati i passaggi principali del processo di estrazione delle immagini.

1. Lettura dei dati: i dati contenuti nei file DICOM delle scansioni CT e nei file MetaImage e Raw delle scansioni CBCT deformate vengono letti e caricati come volumi completi in due matrici distinte.
2. Conversione in unità Hounsfield: i valori dei pixel delle immagini CT vengono portati in unità Hounsfield attraverso le informazioni sulla retta di conversione contenute nel file DICOM.
3. Ricampionamento e rescaling: le matrici CT e CBCT vengono sottoposte a un'operazione di ricampionamento seguita da un *rescaling slice-by-slice* tramite *padding*, producendo in uscita delle immagini di dimensioni 512x512 pixel.
4. Definizione del range dinamico: si imposta un range dinamico compreso tra -1000 HU e 2000 HU.
5. Creazione della maschera CBCT: viene creata una maschera basata sulla CBCT che viene applicata sia alle immagini CBCT che alle immagini CT; ai pixel al di fuori della maschera viene assegnato un valore pari al limite inferiore del range dinamico stabilito nel punto precedente.
6. Conversione in uint8: le matrici CT e CBCT vengono convertite da unità Hounsfield in uint8.
7. Concatenazione delle immagini: le immagini CBCT e CT corrispondenti vengono accoppiate

- Salvataggio delle immagini accoppiate: le immagini accoppiate non nere vengono salvate in due cartelle differenti per le acquisizioni CBCT effettuate con un sistema Varian o Elekta

Al termine del processo le immagini CT e CBCT sono contenute in un'unica immagine di dimensione 512x1024 pixel, come mostrato nella **Figura 11**. È possibile osservare come le immagini salvate comprendano anche ricostruzioni parziali, che si manifestano tipicamente ad inizio (Figura 11C) e fine volume (Figura 11D).

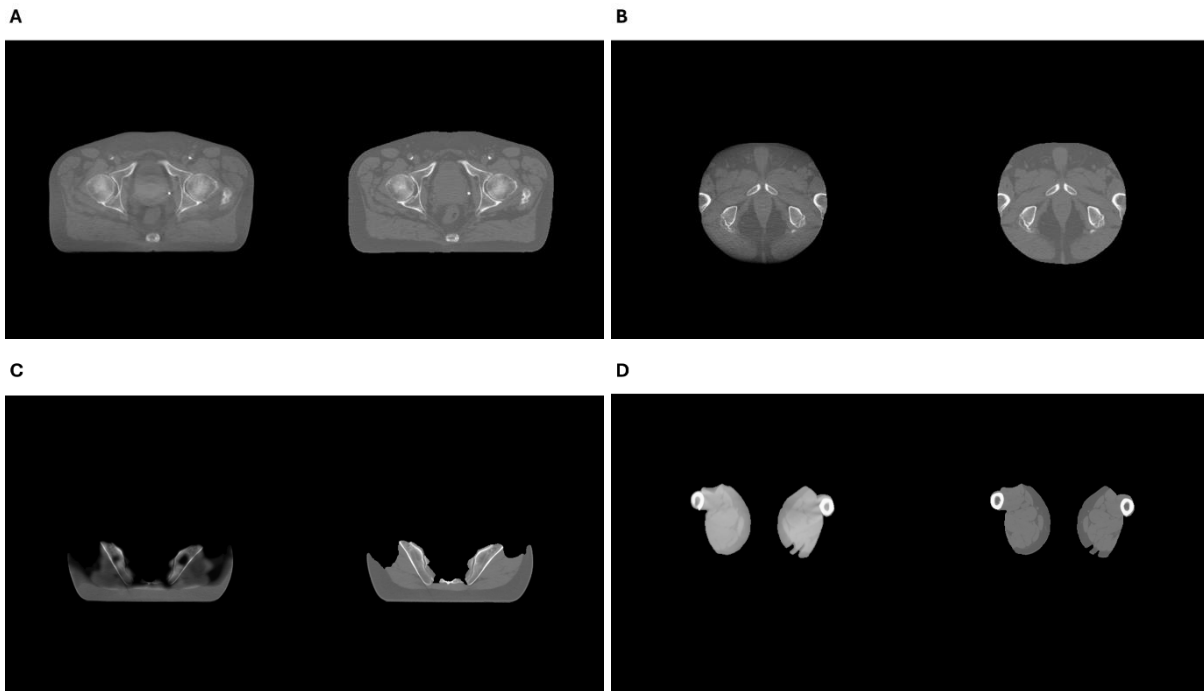


Figura 11. Esempio di immagini in output dal processo di estrazione delle immagini: l'oggetto a sinistra è la scansione CBCT, l'oggetto a destra la scansione CT. **A.** FOV completo. **B.** FOV ridotto. **C.** Immagine parzialmente ricostruita ad inizio volume. **D.** Immagine parzialmente ricostruita a fine volume

3.4. Suddivisione del dataset

Il dataset completo così ottenuto è stato successivamente suddiviso per individuare le due classi per l'allenamento dei metodi di Machine Learning, dove per classe 0 si intende l'insieme delle immagini da scartare e per classe 1 si intende l'insieme delle immagini da selezionare per l'addestramento della GAN.

Una precedente operazione di scrematura manuale delle immagini aveva individuato un sottoinsieme di immagini utile per il training della GAN: a questo sottogruppo è stata assegnata

l'etichetta 1; di conseguenza, la classe 0 è stata assegnata all'insieme differenza del dataset completo estratto e le immagini del sottoinsieme di training.

	VARIAN	ELEKTA
CLASSE 0	5537	3031
CLASSE 1	1697	898

Tabella 1. Suddivisione in classi e numerosità del dataset.

3.5. Determinazione delle ROI

Le considerazioni che hanno portato alla scelta di utilizzare delle ROI per il calcolo delle metriche sono due. Anzitutto è stato considerato il fatto che nei pazienti Varian 20 acquisizioni su 83 sono state realizzate con un FOV ridotto: sapendo che le immagini estratte per acquisizione sono poco meno di un centinaio, le immagini a FOV ridotto sono approssimativamente 2000, costituendo circa il 25% dei dati Varian a disposizione e il 18% delle immagini totali.

La seconda considerazione riflette sulla necessità di escludere lo sfondo per favorire considerazioni relative esclusivamente all'oggetto, dal momento che si ha interesse nel valutare una corretta ricostruzione delle strutture anatomiche, ma anche l'eventuale presenza di artefatti o errori nella ricostruzione (**Figura 12B**). Da un punto di vista delle caratteristiche delle immagini estratte, sebbene durante l'estrazione si applichi una maschera per la segmentazione dell'oggetto, non sempre tale segmentazione individua il solo oggetto, com'è visibile nella **Figura 12A**. L'applicazione della ROI consente di uniformare a 0 i valori dei pixel della regione esterna alla ROI stessa, estromettendo la quasi totalità dello sfondo dalle valutazioni sulle metriche.

Per riprendere la forma del FOV ridotto, si è scelta una ROI circolare, il cui raggio è stato selezionato a 110 pixel, in modo da corrispondere con le dimensioni del FOV ridotto delle scansioni dei pazienti Varian. Poiché con la ROI di 110 pixel venivano comunque comprese regioni non appartenenti all'oggetto, si è scelto di valutare le metriche anche su altre due ROI di dimensione inferiore, nell'ipotesi che si potesse beneficiare in termini di riduzione del rumore

e per ottenere una valutazione più accurata delle caratteristiche dell'immagine in una regione tendenzialmente critica nel processo di creazione della sCT.

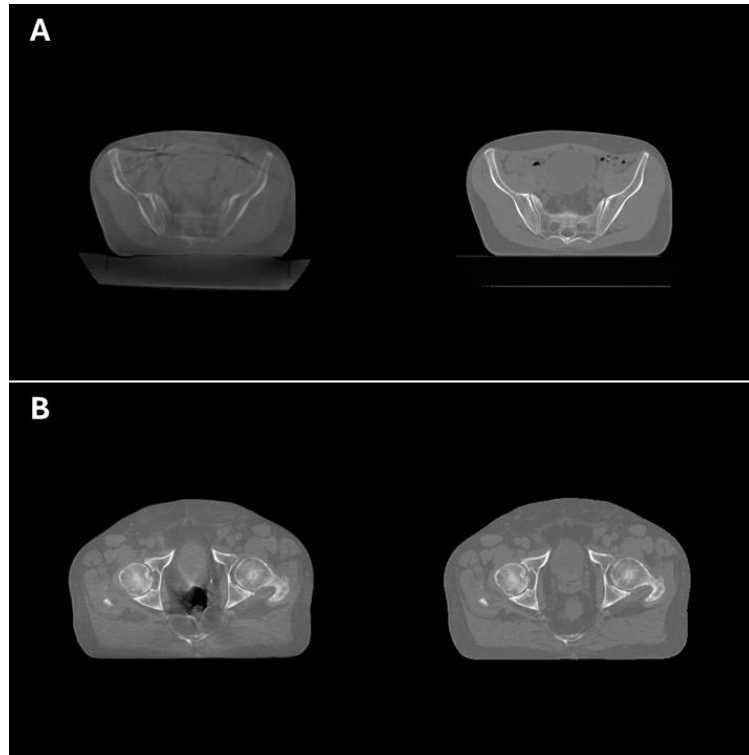


Figura 12. Due esempi di problematicità nelle immagini; in entrambi a sinistra si ha la CBCT; a destra CT. **A.** Oggetto non isolato: l'alone visibile nell'immagine CBCT è dovuto ad una rimozione del lettino non efficace. **B.** Errore ricostruzione delle strutture anatomiche.

I raggi di dimensioni minori sono stati determinati considerando sia la rimozione di eventuali elementi di sfondo ancora presenti, sia la necessità di avere una rappresentazione adeguata dell'oggetto, mantenendo comunque delle differenze apprezzabili tra le varie ROI: i raggi scelti per le due ROI minori sono 110 pixel, 85 pixel, 60 pixel (**Figura 13**). Il centro della circonferenza della ROI è stato fatto allineato con il centroide dell'oggetto, poiché è raro che l'oggetto stesso sia perfettamente centrato nell'immagine.

Per realizzare la ROI, innanzitutto è stata effettuata una segmentazione dell'oggetto nell'immagine creando una maschera binaria tramite Global Thresholding; per la selezione della soglia è stata ritenuta sufficiente una valutazione empirica, date le caratteristiche e le modalità di costruzione dell'immagine accoppiata. Sulla maschera di segmentazione ottenuta è

stata applicata una chiusura, quindi un'operazione di dilatazione a cui segue un'erosione, per realizzare *hole filling*.

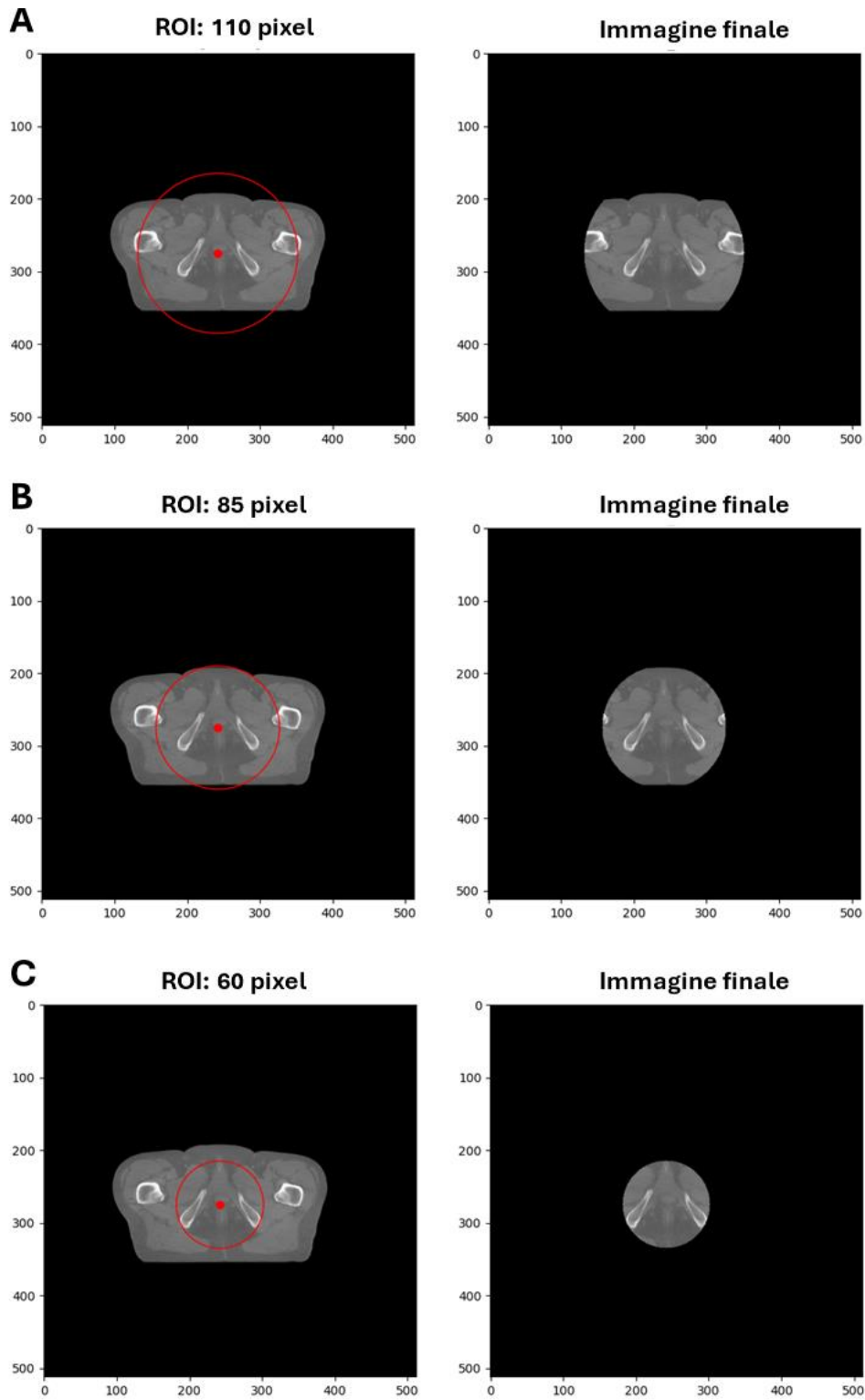


Figura 13. Visualizzazione delle tre ROI su un'immagine d'esempio (Varian).

A questo punto si è proceduto con il calcolo delle coordinate del centroide sull'area della maschera binaria dell'oggetto. Noto il raggio della circonferenza e il suo centro, è stata realizzata la maschera della ROI, assegnando alla maschera di segmentazione un valore di 0 a tutti i pixel al di fuori della circonferenza.

Una volta ottenuta la maschera binaria della ROI, questa è stata sovrapposta all'immagine in ingresso. Il risultato di questa operazione è l'immagine di output sulla quale effettuare il calcolo delle metriche. Nella **Figura 14** è possibile osservare come l'immagine sia stata elaborata durante questa fase.

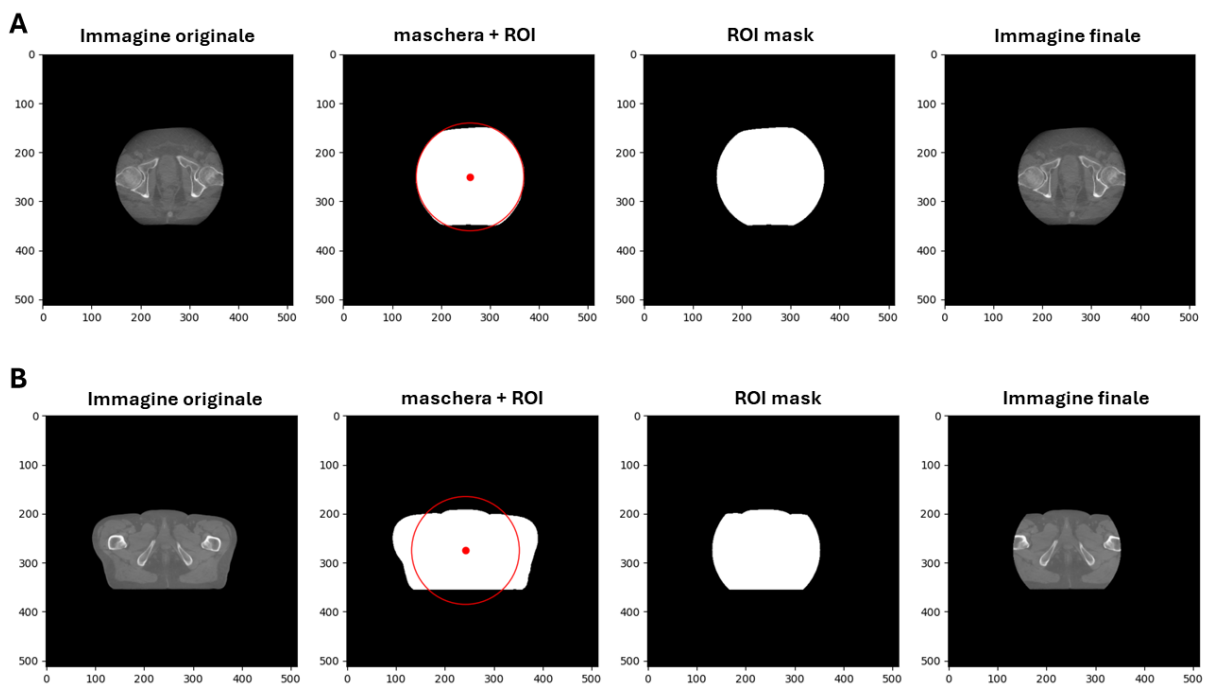


Figura 14. Elaborazione delle immagini durante gli step di preparazione al calcolo delle metriche. Dall'immagine di partenza viene ricavata la maschera di segmentazione che viene troncata dalla circonferenza delimitante la ROI; la maschera troncata viene poi applicata all'immagine di partenza per ottenere l'immagine sulla quale avverrà il calcolo delle metriche. **A.** Immagine a FOV ridotto. **B.** Immagine a FOV completo.

3.6. Calcolo delle metriche

Le osservazioni sono state effettuate a partire dal calcolo di 20 metriche (riportate nella **Tabella 2** e più in dettaglio in **Appendice A**) per osservare, in declinazioni differenti, quattro aspetti dell'immagine, ovvero fedeltà dell'informazione, correlazione o similarità, informazione o incertezza, errore o discrepanza. Esaminare questi aspetti ha lo scopo di individuare e

selezionare le immagini idonee per essere utilizzate come input nella GAN per la generazione di sCT: si vuole garantire che le immagini selezionate siano caratterizzate da una ricostruzione precisa e siano prive di artefatti e/o errori di acquisizione che potrebbero influenzare negativamente l'addestramento della rete e la qualità delle immagini sintetiche prodotte.

Le metriche sono state valutate tra le immagini CBCT e le immagini CT, mantenendo sempre la distinzione tra immagini Varian e immagini Elekta, sulle tre regioni di interesse (ROI) circolari concentriche.

Aspetto valutato	Metriche	
Fedeltà dell'informazione	PSNR – Peak Signal-to Noise Ratio	VIF – Visual Information Fidelity
Correlazione/ similarità	NCC – Normalized Cross-Correlation	PCC – Pearson Correlation Coefficient
	SCC – Spatial Correlation Coefficient	SAM – Spectral Angle Mapper
	MS-SSIM – Multi-Scale Structural Similarity	FSIM – Feature Similarity Index
		SSIM – Structural Similarity Index
Informazione	MI – Mutual Information	ED – Entropy Difference
	JE – Joint Entropy	RUI – Relative Uncertainty Index
Errore/ Discrepanza	ME – Mean Error	MSGE – Mean Squared Gradient Error
	MAE – Mean Absolute Error	SRE – Signal Reconstruction Error
	MSE – Mean Squared Error	DSC – Dice Similarity Coefficient
	RMSE – Root Mean Squared Error	

Tabella 2. Elenco delle metriche calcolate raggruppate per aspetto valutato.

3.7. Step di calcolo

Il calcolo delle metriche è stato effettuato valutando ogni slice salvata al seguito del processo di estrazione, distinguendo tra classe 0 e classe 1. La procedura ha inizio con il caricamento di tutte le immagini presenti nella cartella (Varian o Elekta) e prevede tre blocchi di operazioni effettuate sulla singola immagine per tutte le ROI considerate:

1. Separazione della regione CT e della regione CBCT dall'immagine accoppiata
2. Pipeline di definizione della ROI, sia per l'immagine CT, sia per l'immagine CBCT
3. Calcolo delle metriche, considerando l'immagine CT come ground truth.

L'output di questo processo è un file Excel che raccoglie in tre fogli di calcolo i valori delle 20 metriche per ogni ROI.

3.8. Classificazione delle immagini

Per effettuare il task di classificazione sono stati testati tre differenti metodi di ML, il Random Forest, il k-Nearest Neighbor e il Support Vector Machine, valutando per ognuno tre modelli a dimensionalità progressivamente ridotta:

- Modello 1: allenamento con le 60 features calcolate sulle 3 ROI;
- Modello 2: allenamento con le 20 features della ROI di raggio 110 pixel;
- Modello 3: allenamento con le feature selezionate a seguito di un'analisi di feature importance della ROI di raggio 110 pixel.

In generale, l'algoritmo per la classificazione è organizzato nel seguente modo. Si ha il caricamento e la lettura dei dati classe 0 e classe 1. Come si può osservare nella

Tabella 1, le due classi sono fortemente sbilanciate, per cui il numero di esempi della classe più numerosa viene pareggiato con quello della classe meno rappresentata: la selezione degli elementi è stata resa casuale attraverso uno *shuffle*, per cui gli elementi della classe più numerosa sono stati mischiati.

Le due classi sono state riunite in un unico *dataframe* e i dati in esso contenuti sono stati standardizzati. La divisione tra dati di training e dati di test segue una proporzione 70:30.

Per l'addestramento di ogni classificatore è stata adoperata la tecnica della *Grid Search*, un metodo che consente di ricercare la configurazione degli iperparametri in modo automatico. I parametri da determinare per effettuare la ricerca sono: stimatore, spazio dei parametri, metodo per la ricerca o il campionamento dei candidati, schema di cross-validazione e funzione di score.

Lo stimatore è il metodo per la quale si effettua la ricerca: in questo lavoro i metodi utilizzati sono RF, kNN e SVM. Lo spazio dei parametri consiste in un dizionario nel quale vengono indicati gli iperparametri che si vuole testare e i relativi valori che si vuole esplorare durante la ricerca. Il metodo di ricerca può essere esaustivo o randomizzato: nel caso della ricerca randomizzata i valori della griglia dei parametri vengono scelti casualmente; nel caso della ricerca esaustiva vengono valutate tutte le possibili combinazioni degli iperparametri. La cross-validation è un metodo automatico per la valutazione delle prestazioni di un modello predittivo

attraverso la partizione dei dati in un numero predeterminato di *fold* di dimensioni simili. Durante ogni iterazione del processo di cross-validation, il modello viene addestrato su tutti i fold ad eccezione di uno che viene utilizzato come set di test: il processo si ripete fino a quando tutti i fold non sono stati usati come test. Dunque, nella Grid Search, la cross-validation viene adoperata per valutare le prestazioni dei modelli esplorati, con la possibilità di selezionare il numero di fold da utilizzare. La funzione di score è il metodo con cui viene restituita la valutazione sul modello.

In questa applicazione è stata utilizzata una Grid Search esaustiva, una cross-validation su 5 fold e valutazione dei modelli data tramite accuracy score (precisione nella classificazione). I parametri esplorati per le tre tecniche analizzate sono indicati nella **Tabella 3**.

METODO	IPERPARAMETRO	VALORI	DESCRIZIONE
RANDOM FOREST	n_estimators	5, 10, 20, 50, 100	Numero di alberi decisionali (DT)
	criterion	gini, entropy	Criterio per la suddivisione degli alberi
	max_depth	None, 5,10, 20, 30	Profondità massima di ogni albero decisionale “None” indica l’assenza di un limite alla profondità
	min_samples_split	2, 4, 5 ,10	Numero minimo di campioni per suddividere un nodo di test
	min_samples_leaf	1, 2, 4, 5	Numero minimo di campioni per avere un nodo foglia valido
KNN	n_neighbors	valori da 2 a 15	Numero di vicini da considerare
	weights	uniform, distance	Peso da assegnare ad ogni vicino “uniform” corrisponde ad un peso uguale per ogni vicino, “distance” corrisponde ad un peso inversamente proporzionale alla distanza
	metric	minkowski, euclidean, manhattan	Metrica per il calcolo della similarità tra i punti
SVM	C	0.1, 1, 10, 100, 1000	Valore del termine di penalità sulle variabili di scarto
	kernel	linear, rbf, sigmoid	Funzione per la trasformazione dello spazio delle features

Tabella 3. Spazio dei parametri valutato per i tre metodi ML.

3.9. Descrizione dei modelli

3.9.1. Modello 1. 60 features: ROI 110, ROI 85, ROI 60

- Random Forest.

Per la valutazione delle immagini Varian lo stimatore selezionato ha 5 DT, suddivide i DT tramite Gini Index, la profondità degli alberi è non limitata, un numero minimo di campioni per foglia pari ad 1 e numero minimo di campioni per nodo di test pari a 5. Per le immagini Elekta si hanno 20 DT, suddivisione tramite Gini Index, profondità non limitata, numero minimo di campioni per foglia pari ad 1 e numero minimo di campioni nei nodi di test pari a 2.

- kNN

Sia per le immagini Varian che per le immagini Elekta la metrica adottata è Minkowski, con un numero di vicini pari a 2 e pesi uniformi.

- SVM

Nel modello allenato sulle immagini Varian lo stimatore migliore usa un termine di penalità pari a 1 e un kernel basato sulla funzione radiale di base (RDF). Per Elekta, il termine di penalità è pari a 10 e si usa un kernel lineare.

3.9.2. Modello 2. 20 features: ROI 110

- Random Forest

Per il gruppo Varian il miglior subset di parametri prevede un numero di stimatori pari a 5 DT, criterio di suddivisione dato dal Gini Index, una profondità non limitata, numero minimo di campioni per foglia pari ad 1 e un numero minimo di campioni per i nodi intermedi pari a 2.

Per le immagini Elekta il numero di DT è 5, il criterio di suddivisione è il Gini Index, profondità massima degli alberi pari a 30, un numero di campioni per foglia pari ad 1 e un numero minimo di campioni nei nodi di test pari a 4.

- kNN

I migliori parametri selezionati sono uguali sia per il caso Varian che per il caso Elekta, con un modello che prevede un numero di vicini pari a 2, pesi valutati sulla distanza e Minkowski come metrica di similarità.

- SVM

Sia per le immagini Varian che per quelle Elekta è stato selezionato un termine di penalità pari a 10, ma si differiscono per il kernel, rispettivamente RBF e lineare.

3.9.3. Modello 3. Features selezionate: ROI 110

Il terzo modello è stato allenato valutando le features selezionate a seguito di una valutazione di feature importance effettuata tramite Random Forest.

- Random Forest

Per le immagini Varian il modello ha un numero di DT pari a 50, la suddivisione basata sul Gini Index, profondità dei DT non limitata, numero minimo di campioni per foglia pari ad 1 e un numero minimo di campioni per i nodi intermedi pari a 2.

Rispetto a Varian, per Elekta il numero di DT è 20 e la profondità massima selezionata è 30.

- kNN

Per i gruppi di immagini è stato selezionato un numero di vicini pari a 2 e pesi valutati sulla distanza, ma si distinguono per la metrica selezionata, Minkowski il set Varian e Manhattan per il set Elekta.

- SVM

Sia per il gruppo Varian che per quello Elekta è stato scelto un kernel RBF, differenziandosi per il termine di penalità pari a 10, rispettivamente pari a 100 e 1000.

Capitolo 4

Analisi dei dati e dei risultati

4.1. Valutazione preliminare dei dati

Nelle fasi iniziali di sviluppo, sono state osservate le distribuzioni di quattro metriche generalmente adoperate nel confronto tra due immagini, Mean Error (ME), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) e DSC (Dice Similarity Coefficient) [8], valutate nella ROI di raggio 110.

Dataset Varian

La **Figura 15A** riporta la rappresentazione statistica e in frequenza delle distribuzioni di ME, PSNR, SSIM e DSC, dove è evidente una presenza significativa di outlier notevolmente dispersi nei campioni di classe 0; in generale, si è osservato un certo grado di sovrapposizione delle distribuzioni delle due classi. A seguito della rimozione degli outlier (**Figura 15B**) una parte della distribuzione della classe 1 presenta una separazione netta rispetto alla classe 0, tuttavia permane una porzione di dati con una distribuzione sovrapposta a quella della classe 1.

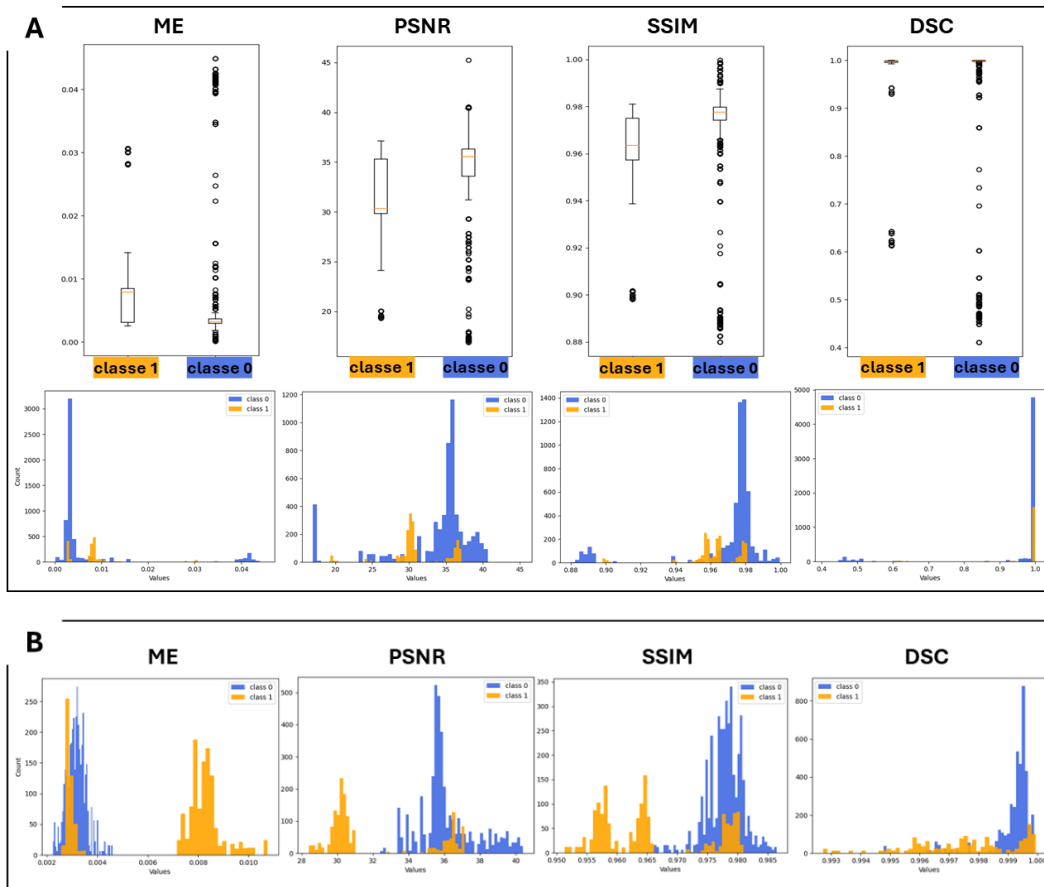


Figura 15. Distribuzione dei dati e boxplot per ME, PSNR, SSIM e DSC tra le due classi, caso Varian **A.** prima della rimozione degli outlier, **B.** dopo la rimozione degli outlier.

Dataset Elekta

Nelle metriche osservate, le distribuzioni dei valori dei dati grezzi (**Figura 16A**) della classe 0 inglobano quasi totalmente le distribuzioni relative alla classe 1; la rimozione degli outlier (**Figura 16B**) non ha prodotto differenze significative nella sovrapposizione delle distribuzioni, che è rimasta sostanzialmente marcata. Si nota, inoltre, che il DSC assume valori praticamente costanti.

Considerazioni

Data la significativa sovrapposizione delle distribuzioni dei valori, è stato ritenuto necessario ampliare il set di metriche a disposizione per poter beneficiare di più osservazioni, quindi per avere una quantità di informazioni sul campione in esame più elevata. Con la stessa finalità è stato scelto di valutare le metriche anche su ROI di dimensioni minori, che permettessero sia di estromettere eventuali elementi di sfondo compresi nella ROI da 110 pixel, sia di avere a

disposizione delle stime della regione più centrale all'oggetto, che solitamente presenta delle differenze nella rappresentazione dei tessuti molli tra l'immagine CT e l'immagine CBCT. La gran quantità di dati così ottenuta è stata elaborata con delle tecniche di Machine Learning.

Per le nuove metriche introdotte è stata condotta la stessa analisi delle distribuzioni (grafici riportati in **Appendice B**). Le osservazioni sono analoghe a quanto riportato per ME, PSNR e SSIM. Nessun'altra metrica studiata ha riportato una rappresentazione omogenea come per il DSC.

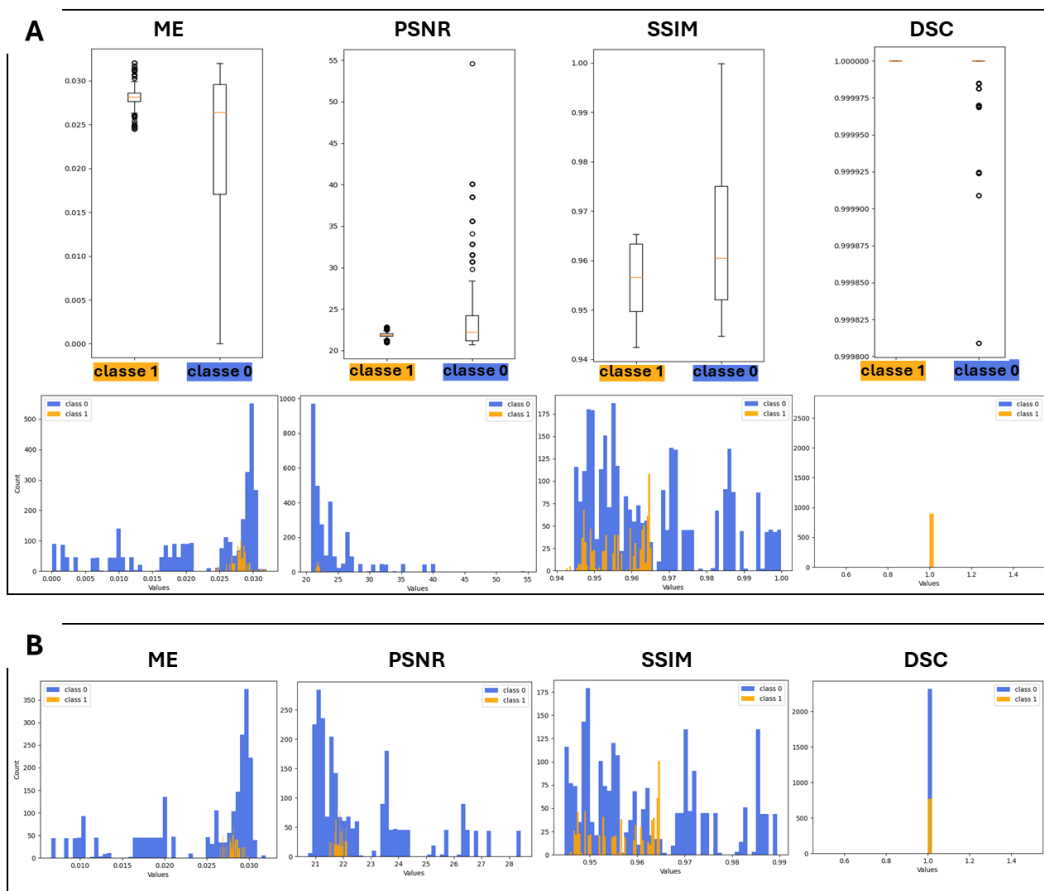


Figura 16. Distribuzione dei dati e boxplot per ME, PSNR, SSIM e DSC tra le due classi, caso Varian: **A.** prima della rimozione degli outlier, **B.** dopo la rimozione degli outlier.

4.2. Valutazioni sulle metriche

Per analizzare le coppie di immagini in esame, sono state selezionate 20 metriche (**Tabella 2**) per osservare quattro caratteristiche nel confronto tra immagini: fedeltà dell'informazione, correlazione o similarità, informazione, errore o discrepanza. Era previsto che gran parte delle metriche avrebbe condiviso un alto grado di correlazione, poiché valutano aspetti simili ma con

approcci leggermente diversi, tuttavia si è scelto comunque di esplorare il potenziale di ognuna nel processo di apprendimento automatico.

Per approfondire ulteriormente le relazioni tra le metriche nelle tre ROI considerate, è stata condotta un'analisi comparativa della loro correlazione attraverso delle mappe di colore in cui il nero/viola indica una correlazione inversa, l'arancione rappresenta un'assenza di correlazione, mentre il beige denota una correlazione diretta.

Dataset Varian

Nella **Figura 17A** e **17B** sono riportate le mappe di colore della classe 1 e classe 0 rispettivamente.

Nella classe 1, si osserva un alto grado di correlazione, sia positiva che inversa, per le metriche valutate nella ROI 110; al diminuire della dimensione della ROI il grado di correlazione generale si riduce, soprattutto confrontando la ROI 85 con la ROI 110. Il colore bianco delle caselle del DSC nella ROI 60 è dovuto all'impossibilità di valutare la correlazione per via dell'alta omogeneità di valori che la metrica assume.

Nella classe 0 si osserva una tendenza opposta alla classe 1: nella ROI 110 la correlazione tra le metriche è moderata, principalmente orientata verso valori negativi; rispetto alla ROI 85, la ROI 60 presenta correlazioni più evidenti, sia positivamente che negativamente.

Dataset Elekta

Le mappe colore della classe 1 (**Figura 17C**) mostrano un'elevatissima correlazione, sia positiva che negativa, tra le metriche della ROI 110. Il livello di correlazione si riduce con le ROI di dimensioni inferiori, fino alle metriche della ROI 60 che risultano scarsamente correlate oppure con una correlazione moderatamente inversa.

Per la classe 0 (**Figura 17D**), la ROI 110 mostra tendenzialmente una polarizzazione verso gli estremi di correlazione, ma si osservano comunque metriche che si collocano nel mezzo dello spettro di colori, suggerendo una scarsa correlazione. Rispetto alla ROI 110, nella ROI 85 la polarizzazione verso i due estremi di correlazione è globalmente più evidente; valutando le metriche nella ROI 60, si torna ad una situazione tendenzialmente analoga a quella osservata per la ROI 110.

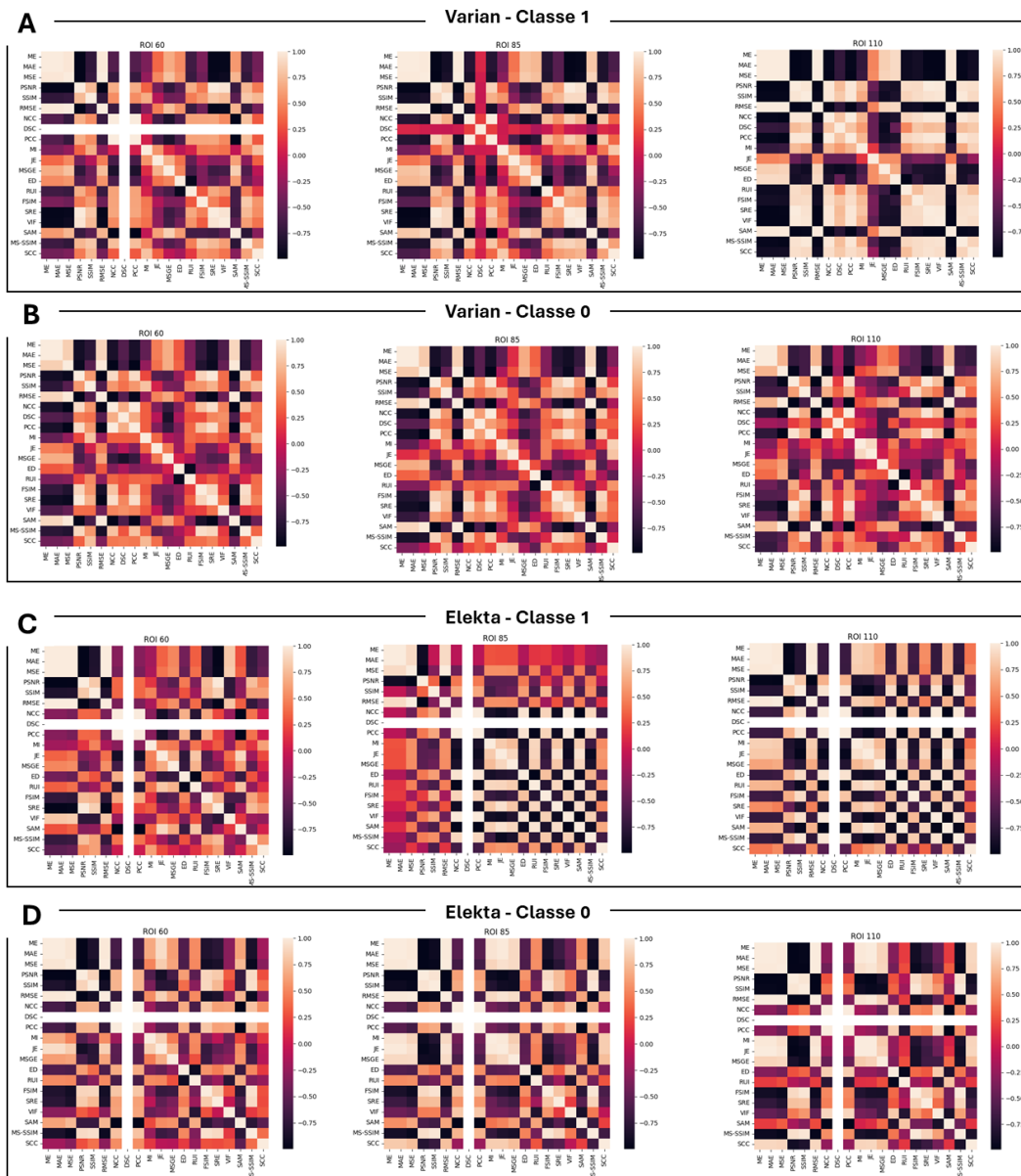


Figura 17. Heatmap di correlazione. Si osserva il DSC di colore bianco nelle mappe di Elekta (C e D) e la ROI 60 di Varian (A) perché non è stato possibile valutare la correlazione, dal momento che la metrica assume valori costanti o comunque poco variabili.

Considerazioni

In generale, si osserva la presenza di una certa variabilità nelle relazioni tra le metriche relativamente alla classe e alle dimensioni della ROI. Da un lato, l'alta correlazione suggerirebbe una ridondanza nelle informazioni, il che potrebbe condurre a problemi di generalizzazione dei modelli di ML; d'altra parte, la variabilità tra le ROI e inter-classe potrebbe consentire l'individuazione di pattern, o in generale, di informazioni rilevanti per l'esecuzione del task di classificazione.

Sulla base di queste considerazioni, si è scelto di iniziare l'analisi dei metodi ML utilizzando per l'allenamento la totalità delle metriche valutate nelle tre ROI, quindi valutando 60 features, per poi proseguire con una progressiva diminuzione della dimensionalità delle features coinvolte.

4.3. Performance dei modelli di Machine Learning

Sono stati comparati tre algoritmi di ML con differenti livelli di complessità: Random Forest (RF), k-Nearest Neighbors (kNN) e Support Vector Machine (SVM). Il RF è stato scelto per i benefici in termini di versatilità, stabilità e capacità di generalizzazione, che lo rendono uno dei metodi più usati in ambito biomedico. Si è deciso di valutare anche il kNN, un algoritmo molto semplice ma efficace con dati di alta dimensionalità. L'SVM è stato valutato perché rappresenta una via intermedia tra RF e kNN, in termini di complessità.

4.3.1. Modello a 60 features: ROI 110, ROI 85, ROI 60

Il primo modello analizzato prevede l'uso di tutte le 60 metriche valutate nelle tre ROI. Nella **Tabella 4** sono riportati i risultati per i tre metodi nei dataset Varian ed Elekta.

Dataset Varian

La rimozione degli outlier rilevati in ogni metrica ha comportato l'esclusione di 2603 su 7234 campioni, di cui 2011 dalla classe 0 e 592 dalla classe 1. A seguito del bilanciamento delle classi il Training Set (TRS) e il Test Set (TSS) sono costituiti rispettivamente da 1547 e 663 campioni. Le performance nella classificazione sono del 100% sia nel TRS che nel TSS. Analizzando le curve di apprendimento (**Figura 18A**), per tutti gli algoritmi si osserva overfitting sui dati di training sin dalle prime istanze di addestramento. La convergenza viene raggiunta per punteggi elevati, ma si nota come siano sufficienti un centinaio di campioni per avere degli score superiori al 90% sia nel training score, che nel cross-validation score.

Dataset Elekta

La rimozione degli outlier in ogni metrica ha portato all'esclusione di 895 campioni di classe 0 e 248 campioni di classe 1, per un totale di 1143 immagini scartate su 3929. A seguito del bilanciamento delle classi, il numero di campioni nel TRS è 910, mentre nel TSS è 390. Le performance nella classificazione sono prossime al 100% in tutti i metodi valutati. Le curve di apprendimento (**Figura 18B**) mostrano una situazione analoga a quanto descritto

nel paragrafo precedente per il dataset Varian: il punteggio di training è pari ad 1 sin dalle prime istanze di apprendimento e le curve di cross-validation raggiungono la convergenza rapidamente.

Considerazioni

Le curve di apprendimento hanno mostrato un elevato adattamento ai dati di addestramento, per cui è stato ritenuto opportuno ridurre la complessità dei modelli diminuendo il numero di features fornite.

	<i>Varian</i>			<i>Elekta</i>				
	TRS (%)	TSS (%)	Confusion Matrix (TSS)		TRS (%)	TSS (%)	Confusion Matrix (TSS)	
<i>Random Forest</i>	100	100	327	0	100	99,74	195	1
			0	336			0	194
<i>kNN</i>	98,35	100	327	0	100	99,74	195	1
			0	336			0	194
<i>SVM</i>	100	100	327	0	100	100	196	0
			0	336			0	194

Tabella 4. Performance per il modello a 60 features. Nelle Confusion Matrix: $C_{0,0}$ =TN, $C_{0,1}$ =FP, $C_{1,0}$ =FN, $C_{1,1}$ =TP.

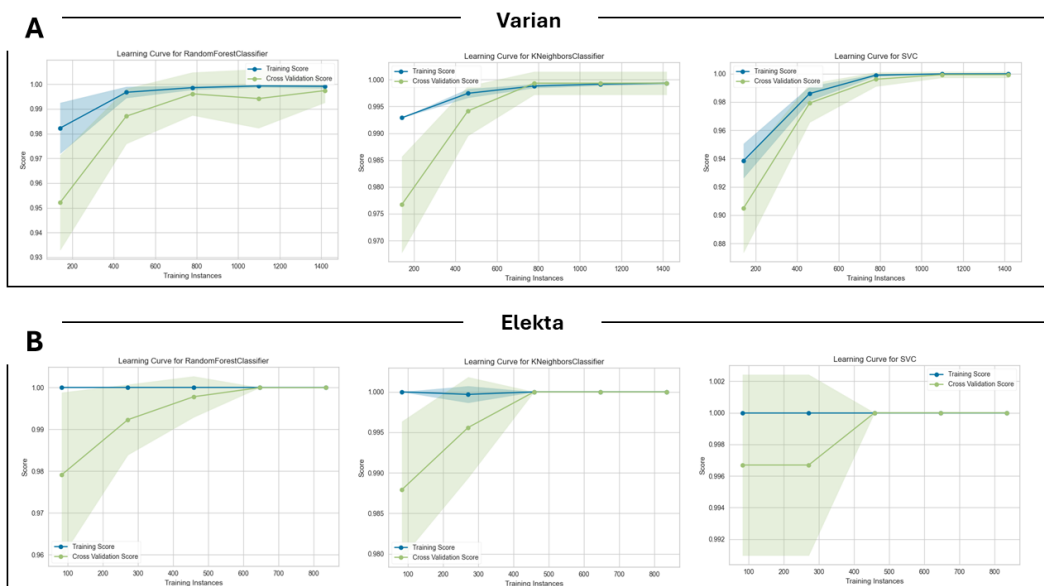


Figura 18. Curve di apprendimento: modello a 60 features. Lo score è l'accuracy nella classificazione, le curve riportano dei punteggi medi e le sfumature rappresentano la variabilità in termini di deviazione standard. In blu: training score; in verde: cross-validation score. A sinistra: RF; al centro: kNN; a destra: SVM.

4.3.1. Modello a 20 features, ROI 110

Per effettuare la riduzione di dimensionalità sono state eseguite delle prove che esplorassero i risultati attraverso differenti combinazioni di numero di features e ROI, con esiti sostanzialmente simili. Per brevità di trattazione, verranno illustrati i test relativi alla ROI 110. La **Tabella 5** riporta i risultati dei tre algoritmi di ML per il caso Varian e per il caso Elekta per l'addestramento sulle 20 features della ROI 110.

Dataset Varian

In questa analisi, nell'eliminazione degli outlier sono state rimosse 1693 immagini, di cui 1566 dalla classe 0 e 127 dalla classe 1. Dopo il bilanciamento delle classi la divisione in TRS e TSS è risultata rispettivamente in 2198 campioni e 942 campioni.

Non sono state osservate differenze nei risultati rispetto a quanto ottenuto con il modello allenato sulle 60 features di tutte le ROI: le performance per i tre algoritmi ML sono pari o prossime al 100% sia per TRS che per TSS.

Le curve di apprendimento (**Figura 19A**) riportano andamenti sostanzialmente analoghi a quelli riscontrati per il modello a 60 features (**Figura 18A**).

Dataset Elekta

Il processo di pulizia del dataset ha portato all'esclusione di 713 immagini dalla classe 0 e 124 dalla classe 1, per un totale di 837 immagini rimosse. Il set di immagini pulito e bilanciato è stato diviso in un TRS composto di 1083 immagini e di un TSS di 465 immagini.

Le performance di kNN e SVM sono rimaste invariate rispetto al modello allenato con le 60 features di tutte le ROI, con prestazioni pari o prossime al 100% sia nel TRS che nel TSS; nel RF si nota un lieve aumento di misclassificazioni bilanciate tra le due classi, ma delle performance comunque pari al 99,54% sul TRS e 98,92% sul TSS.

L'analisi delle curve di apprendimento (**Figura 19B**) non ha messo in evidenza variazioni significative rispetto a quanto osservato nel modello a 60 features (**Figura 18B**). Si nota un lieve aumento della variabilità nelle curve di cross-validation.

Considerazioni

Il comportamento dei classificatori è sostanzialmente rimasto invariato: per validare i risultati ottenuti i modelli sono stati semplificati ulteriormente eseguendo una nuova riduzione della dimensionalità.

	Varian			Elekta				
	TRS (%)	TSS (%)	Confusion Matrix (TSS)		TRS (%)	TSS (%)	Confusion Matrix (TSS)	
Random Forest	99,95	100	463	0	99,54	98,92	235	2
			0	479			3	225
kNN	100	100	463	0	100	99,78	236	1
			0	479			0	228
SVM	100	99,89	462	1	100	99,78	236	1
			0	479			0	228

Tabella 5. Performance per il modello a 20 features. Nelle Confusion Matrix: $C_{0,0}$ =TN, $C_{0,1}$ =FP, $C_{1,0}$ =FN, $C_{1,1}$ =TP.

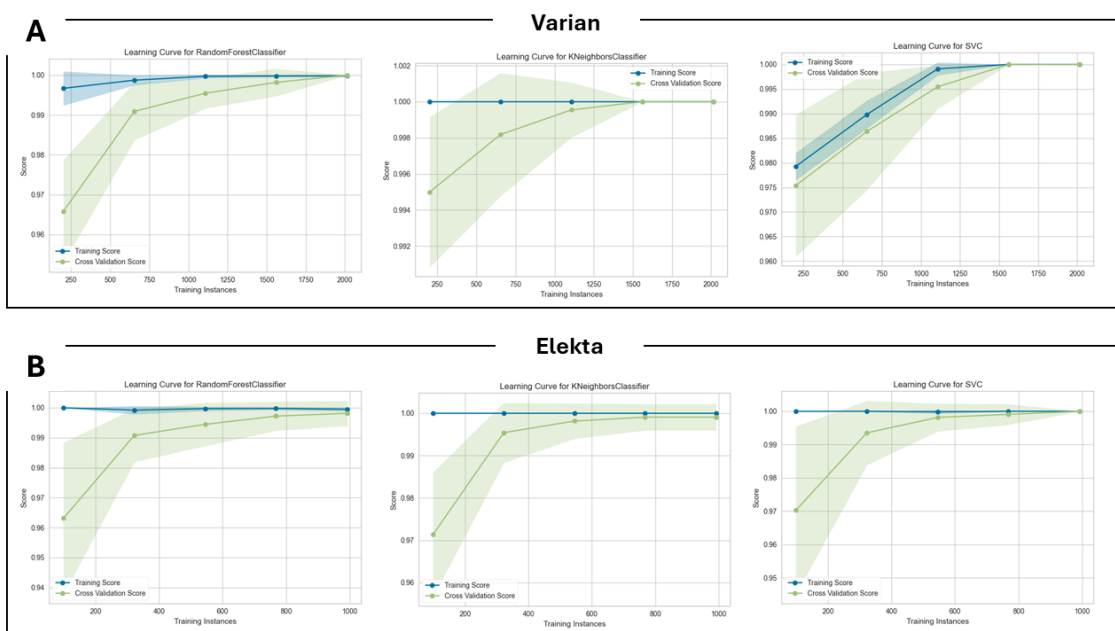


Figura 19. Curve di apprendimento: modello a 20 features (ROI 110). Lo score è l'accuracy nella classificazione, le curve riportano dei punteggi medi e le sfumature rappresentano la variabilità in termini di deviazione standard. In blu: training score; in verde: cross-validation score. A sinistra: RF; al centro: kNN; a destra: SVM.

4.3.2. Modello con features selezionate

La selezione delle feature è stata realizzata analizzando il contributo della capacità predittiva del modello tramite feature importance implementata con l'algoritmo Random Forest. I risultati per il dataset Varian e per il dataset Elekta sono riportati nella **Figura 20**, che visualizza le metriche ordinate per importanza decrescente. La soglia di selezione delle features è stata determinata per limitare ragionevolmente la dimensionalità: considerando i profili ottenuti, per entrambi i set di dati sono state considerate le features con uno score di importanza almeno pari all'8%. La **Tabella 6** i risultati dei tre algoritmi di ML per il caso Varian e per il caso Elekta.

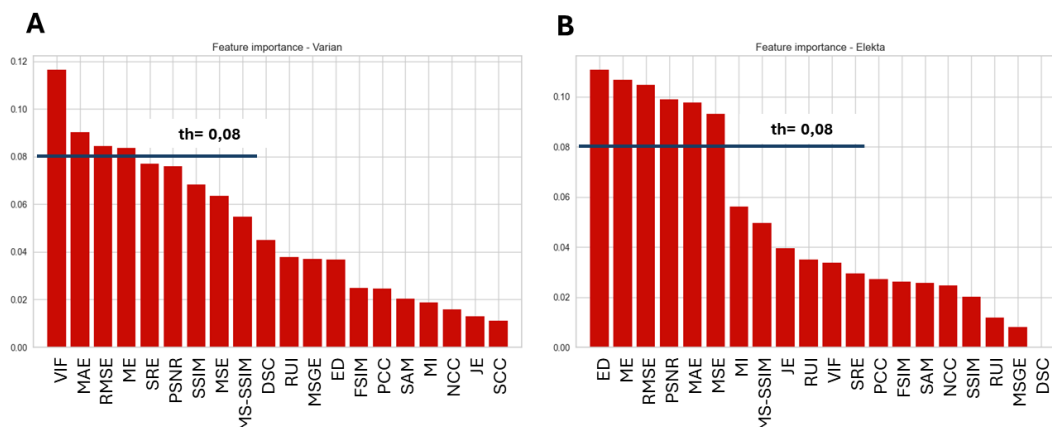


Figura 20. Feature Importance. A sinistra: Varian; a destra: Elekta

Dataset Varian

Nel profilo d'importanza (**Figura 20A**), la feature Visual Information Fidelity (VIF) emerge particolarmente sulle altre metriche, che invece riportano differenze spesso inferiori al 1% nel punteggio di importanza, denotando un andamento gradualmente decrescente. Le metriche selezionate sono VIF (11.64%), MAE (9.02%), RMSE (8.44%) e ME (8.38%).

In questa analisi, l'operazione di pulizia degli outlier ha condotto alla rimozione di 1693 immagini, di cui 1566 dalla classe 0 e 127 dalla classe 1. Dopo il bilanciamento delle classi la divisione in TRS e TSS è risultata rispettivamente in 2198 campioni e 942 campioni.

Le performance ottenute con il Random Forest e con il kNN sono invariate rispetto a quelle osservate nei modelli a 20 features e a 60 features, con risultati prossimi o pari al 100% per TRS e TSS. L'algoritmo SVM ha mostrato un calo nelle performance, riportando sul TRS il 96,72% e sul TSS il 96,18% di corretti classificati, con una sensibilità sul TSS pari al 98,74 % e una specificità del 94,13%.

L'analisi delle curve di allenamento (**Figura 21A**) riporta per Random Forest e kNN un punteggio di training è pari ad 1 sin dalle prime istanze di apprendimento, come già osservato nei modelli a 20 features e a 60 features. Per SVM, training score e del cross-validation score convergono su un punteggio pari al 97%.

Dataset Elekta

Dal grafico della feature importance (**Figura 20B**) si individuano sei metriche con uno score significativamente più elevato rispetto alle altre metriche valutate, con una differenza tra la

prima per importanza e la prima tra le escluse circa pari al 6%. Le feature selezionate sono ED (11.08%), ME (10.68%), RMSE (10.49%), PSNR (9.90%), MAE (9.78%) e MSE (9.31%).

La pulizia del dataset dagli outlier ha portato all'esclusione di 713 immagini dalla classe 0 e 124 dalla classe 1, per un totale di 837 immagini rimosse. Il dataset pulito e bilanciato è stato diviso in un TRS e un TSS 1083 e 465 campioni rispettivamente.

Le performance ottenute sono lievemente inferiori rispetto a quelle osservate nei modelli a più alta dimensionalità. Nel Random Forest si ottiene una sensibilità del 100% e una specificità del 98,73%; il kNN riporta una sensibilità del 98,68% e una specificità del 99,57%; infine, il SVM restituisce una sensibilità del 100% e una specificità del 97,46%, dimostrandosi il metodo più abile nell'identificare le immagini di classe 1, ma il meno performante nel riconoscimento delle immagini di classe 0.

L'osservazione delle curve di learning (**Figura 21B**) ha fatto notare un aumento della variabilità nel cross-validation score in tutti i metodi. Si evidenzia la curva relativa al SVM, in cui si raggiunge lentamente la convergenza ad uno score di poco superiore al 98%.

Considerazioni

La selezione delle features ha mantenuto fondamentalmente invariate le performance per gli algoritmi RF e kNN, mentre il SVM è risultato essere il metodo più sensibile alla riduzione di dimensionalità. Si nota, comunque, che l'andamento delle curve di apprendimento è rimasto sostanzialmente inalterato rispetto ai modelli a dimensionalità superiore.

	<i>Varian</i>			<i>Elekta</i>		
	TRS (%)	TSS (%)	Confusion Matrix (TSS)	TRS (%)	TSS (%)	Confusion Matrix (TSS)
<i>Random Forest</i>	100	99,79	462 1 1 478	99,63	99,35	234 3 0 228
<i>kNN</i>	100	99,89	462 1 0 479	100	99,14	236 1 3 225
<i>SVM</i>	96,72	96,18	433 30 6 473	98,61	98,71	231 6 0 228

Tabella 6. Performance per il modello con le features selezionate. Nelle Confusion Matrix: $C_{0,0}$ =TN, $C_{0,1}$ =FP, $C_{1,0}$ =FN, $C_{1,1}$ =TP.

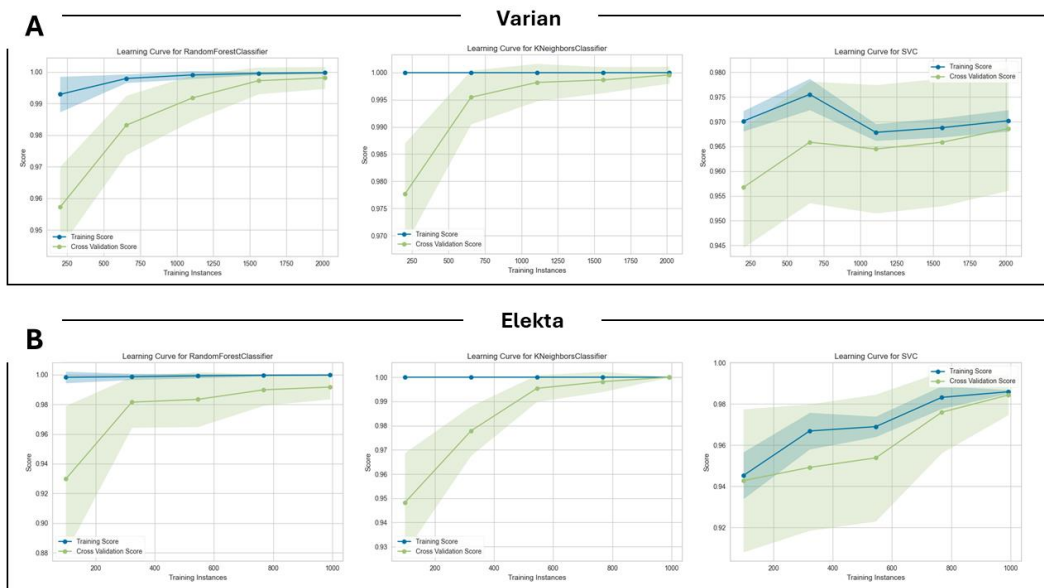


Figura 21. Curve di apprendimento: modello con feature selezionate (ROI 110). Lo score è l’accuracy nella classificazione, le curve riportano dei punteggi medi e le sfumature rappresentano la variabilità in termini di deviazione standard. In blu: training score; in verde: cross-validation score. A sinistra: RF; al centro: kNN; a destra: SVM.

4.4. Metodo delle soglie

I risultati ottenuti con i metodi di ML hanno portato a considerare la possibilità che il problema potesse essere risolto mediante un approccio più semplice. Si è quindi intrapreso uno studio più dettagliato delle distribuzioni delle metriche a partire dai dati grezzi, al fine di determinare se fosse fattibile effettuare la classificazione attraverso una soglia delle metriche.

Sono state selezionate le metriche caratterizzate da una distribuzione dei valori che mostrava una netta separazione tra le due classi, con il minimo margine d’errore. Successivamente, è stata esaminata la distribuzione dei valori delle due classi per ciascuna metrica al fine di stabilire le soglie per la classificazione. Sulla base dei range così definiti, ciascuna metrica ha contribuito a determinare la classe di appartenenza dell’immagine. Per determinare la classe finale di ciascun campione, è stato adottato un metodo di majority voting, dove la classe più frequente tra le diverse metriche determina l’assegnazione finale. Al fine di evitare situazioni di parità, il numero di metriche scelte deve essere dispari. Il metodo è stato valutato sulle ROI 110 e 85 sia di Varian che di Elekta.

Per le immagini Varian le features selezionate sono ME, PSNR, SSIM, RMSE, SRE, VIF e MS-SSIM: la prova effettuata sulla ROI 110 ha restituito una corretta classificazione di 2870 esempi su 3394, con delle prestazioni del 84,56%, mentre sulla ROI 85 le performance sono di 71,66%,

con 1242 classificazioni corrette.

Per le immagini Elekta non è stato possibile applicare questo metodo poiché la sovrapposizione nelle distribuzioni dei valori delle metriche nelle due classi è tale da avere margini d'errore troppo ampi, per cui non è stato possibile definire alcun range.

Questo studio ha contribuito a spiegare la ragione per cui i risultati ottenuti tramite l'applicazione di tecniche di Machine Learning risultassero così elevati. Tuttavia, nonostante l'adozione di soglie possa sembrare una soluzione più intuitiva per risolvere il problema di classificazione, tale approccio comporta una complessità maggiore nell'analisi e una minore capacità di generalizzazione dell'algoritmo a soglia implementato. Questo aspetto invece potrebbe essere meglio gestito dall'implementazione di solidi metodi di ML che permettano di analizzare il campione con un più alto grado di generalizzazione.

4.5. Valutazioni su altro dataset

Come ulteriore analisi si è deciso di valutare lo stesso approccio adottato per il distretto pelvi con un dataset nuovo, relativo al distretto testa-collo. L'analisi di questo nuovo campione è stata inserita nell'ultima fase di testing per validare la procedura utilizzata per il distretto pelvico.

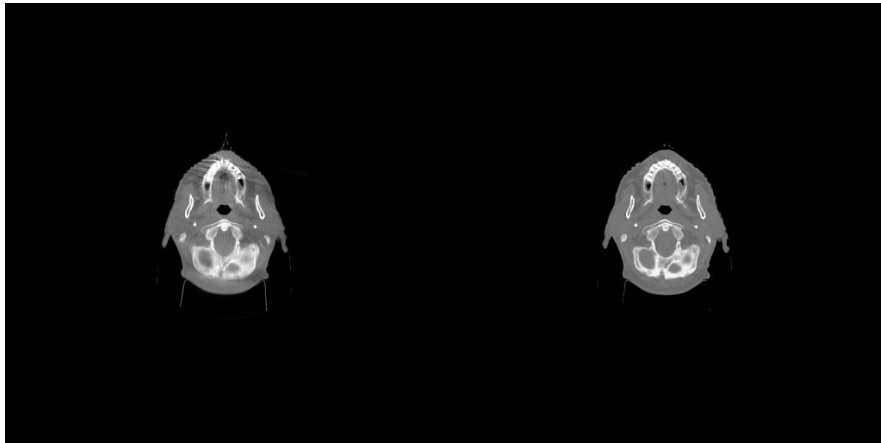


Figura 22. Esempio di immagine dal dataset testa-collo (classe 1).

Il nuovo dataset analizzato comprende immagini CBCT ottenute con sistemi Varian e con sistemi Elekta, non distinte per dispositivo, ma già suddivise in classe 0 (2014 campioni) e classe 1 (5676 campioni). In una prima fase, per eseguire il task di classificazione, sono stati utilizzati i modelli parametrizzati sul distretto pelvi e le performance sono state confrontate con

quelle del distretto pelvico (modello a 60 features: **Tabella 4**; modello a 20 features sulla ROI 110: **Tabella 5**; modello con features selezionate: **Tabella 6**).

4.5.1. Valutazione con i modelli Varian

L'uso degli iperparametri selezionati sul distretto pelvi ha comportato un abbassamento delle performance rispetto a quanto ottenuto nel distretto pelvi in tutti gli algoritmi valutati. Nella **Tabella 7** sono riportati i risultati.

Nel modello a 60 features l'algoritmo più performante sul TSS è il kNN, con un'accuratezza del 92,51%, una specificità del 97,80% e una sensibilità del 87,27%. Le prestazioni più basse le ha riportate il SVM, con una percentuale di corrette classificazioni dell'87% sul TRS e dell'84% sul TSS, riportando una specificità del 74,61% e una sensibilità del 94,72%.

Nel modello allenato con le 20 features sulla ROI 110 le prestazioni generali sono inferiori rispetto al modello a 60 features: il kNN si conferma essere il metodo più accurato nella classificazione (99,94% su TRS, 93,18% su TSS), mentre il SVM il metodo con performance più basse tra i tre algoritmi valutati (88,46% su TRS, 85,83% su TSS).

Nell'analisi del modello allenato con le feature selezionate per le immagini Varian le prestazioni dei tre algoritmi sul TSS si sono ridotte ulteriormente. Il kNN è il metodo che ha riportato il calo più significativo, con una percentuale di corrette classificazioni pari all'85,65%, quasi il 10% in meno rispetto al modello allenato sulle 20 features. Il RF è il metodo con le prestazioni migliori (85,3% su TSS), mentre il SVM è l'algoritmo con le performance peggiori, con l'80,58% di corrette classificazioni sul TSS.

	<i>60 features</i>			<i>20 features</i>			<i>Feature selezionate</i>		
	TRS (%)	TSS (%)	CM	TRS (%)	TSS (%)	CM	TRS (%)	TSS (%)	CM
<i>Random Forest</i>	98,19	89,54	287 32 35 287	98,42	87,80	354 39 54 315	99,83	85,30	334 59 53 316
<i>kNN</i>	96,32	92,51	312 7 41 281	99,94	93,18	379 14 38 331	99,94	84,65	337 56 61 308
<i>SVM</i>	87,14	84,71	238 81 17 305	88,46	85,83	327 66 42 327	82,38	80,58	322 71 77 292

Tabella 7. Riassunto delle performance dei modelli Varian nel distretto testa-collo. La confusion matrix (CM) è relativa al TSS. Confusion Matrix: $C_{0,0}$ =TN, $C_{0,1}$ =FP, $C_{1,0}$ =FN, $C_{1,1}$ =TP

4.5.2. Valutazione con i modelli Elekta

La classificazione con i modelli parametrizzati sul distretto pelvico ha riportato performance inferiori a quanto ottenuto nel distretto pelvi in tutti gli algoritmi valutati. La **Tabella 8** riporta i risultati.

Il modello che presenta le performance più elevate è quello a 60 features, dove il RF è l'algoritmo con le prestazioni migliori sul TSS (93,14%), il kNN è il metodo con la sensibilità più bassa (87,26%), mentre il SVM è il metodo con performance peggiori, ma sensibilità più elevata (92,86%).

L'analisi del modello a 20 features ha riportato un lieve aumento delle prestazioni sul TSS per kNN e un calo delle performance per RF e SVM. Il kNN è l'algoritmo con la più alta percentuale di corretti classificati sul TSS (93,18%) e con la sensibilità più elevata (89,70%).

Valutando il modello allenato con le feature selezionate per le immagini Elekta, si osserva un'ulteriore diminuzione delle performance per tutti gli algoritmi, con il kNN che registra la diminuzione più significativa (84,65% su TSS). Il Random Forest è il metodo con le performance migliori sul TSS (85,30%) e con la sensibilità più alta (85,64%), mentre il SVM è l'algoritmo con le prestazioni peggiori (80,58%).

	60 features			20 features			Feature selezionate		
	TRS (%)	TSS (%)	CM	TRS (%)	TSS (%)	CM	TRS (%)	TSS (%)	CM
Random Forest	99,87	93,14	300 19	98,09	88,06	359 34	99,83	85,30	334 59
			25 297			57 312			53 316
kNN	96,32	92,51	312 7	99,94	93,18	379 14	99,94	84,65	337 56
			41 281			38 331			61 308
SVM	92,50	88,14	266 53	82,66	83,33	306 87	82,38	80,58	322 71
			23 299			40 329			77 292

Tabella 8. Riassunto delle performance dei modelli Elekta nel distretto testa-collo. La confusion matrix (CM) è relativa al TSS. Confusion Matrix: $C_{0,0}$ =TN, $C_{0,1}$ =FP, $C_{1,0}$ =FN, $C_{1,1}$ =TP

4.5.3. Nuovo allenamento

L'uso dei modelli parametrizzati sul distretto pelvi per la valutazione del distretto testa-collo ha dato come risultato performance più basse ma comunque promettenti, Si è comunque deciso di

effettuare un allenamento specifico per il nuovo distretto in esame per valutare il comportamento dei modelli di ML. Dunque, sulla base dei modelli precedentemente definiti per il distretto pelvi, sono stati rivalutati gli iperparametri ed è stato ripetuto l'addestramento con i campioni del dataset testa-collo.

In linea generale i tre modelli hanno restituito performance ottimali, ma inferiori rispetto a quanto ottenuto nel distretto pelvi. La **Tabella 9** riporta un riassunto delle performance.

	60 features			20 features			Feature selezionate		
	TRS (%)	TSS (%)	CM	TRS (%)	TSS (%)	CM	TRS (%)	TSS (%)	CM
<i>Random Forest</i>	100	92,67	296 23 24 298	99,49	90,68	365 28 43 326	99,27	82,15	336 57 79 290
<i>kNN</i>	100	94,54	304 15 20 302	99,94	93,18	379 14 38 331	99,94	80,71	334 59 88 281
<i>SVM</i>	99,40	92.67	296 26 21 301	95,95	90,68	363 30 41 328	81,25	76,51	324 69 110 259

Tabella 9. Riassunto delle performance dei modelli riallenati nel distretto testa-collo. La confusion matrix (CM) è relativa al TSS. Confusion Matrix: $C_{0,0}$ =TN, $C_{0,1}$ =FP, $C_{1,0}$ =FN, $C_{1,1}$ =TP

Il modello a 60 features è il più performante, con risultati superiori al 99% sul TRS e una percentuale media di 93,29% sul TSS per tutti e tre i metodi valutati. Per il Random Forest risulta una sensibilità del 93,42% e una specificità pari al 92,79%, mentre il kNN ha riportato una miglior capacità di riconoscimento delle immagini di classe 0 (specificità pari a 95,30%).

L'addestramento con le 20 metriche della ROI da 110 pixel ha portato a un lieve calo delle performance complessive. Per Random Forest e kNN le performance sul TRS sono rimaste sostanzialmente stabili, mentre sul TSS hanno registrato un calo del 2%; per quanto riguarda SVM, si è verificata una diminuzione media delle performance sul TRS e sul TSS pari a circa il 3%. Sensibilità e specificità media dei tre metodi si attesta rispettivamente all'86,36% e al 93,89%.

L'analisi sull'importanza delle features per l'addestramento del modello basato sulla feature selection ha visto la selezione di 4 metriche applicando una soglia del 6,5% di importance: VIF (10,15%), ED (8,61%), DSC (8,08%) e RUI (7%). Rispetto al modello a 20 features, in questo

caso si è verificato un peggioramento delle performance sul TSS, con un calo del 10% per Random Forest e kNN e di circa il 15% per SVM. Come nel modello a 20 features, RF e kNN non hanno subito variazioni significative nelle prestazioni sul TRS, mentre SVM si è confermato il meno performante con una diminuzione di circa il 15%. In generale, le confusion matrix dei tre metodi (**Tabella 9**) riportano una capacità di riconoscere le immagini di classe 1 inferiore.

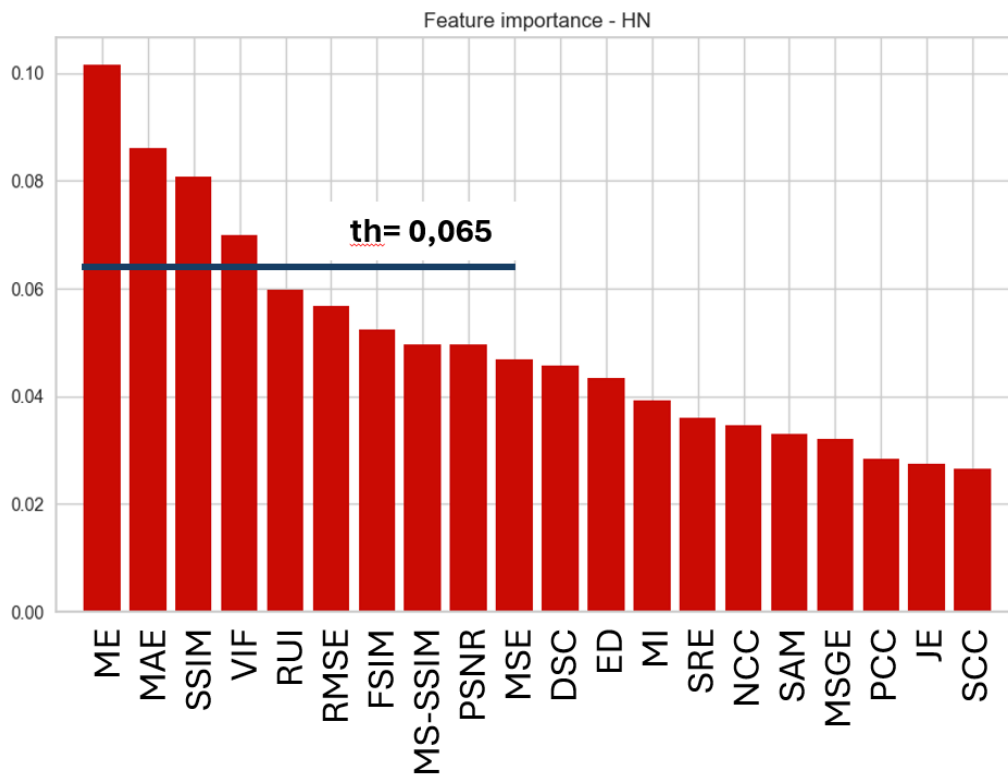


Figura 23. Feature importance per il distretto testa-collo.

Dalle curve di apprendimento (**Figura 24**) è emerso che nessun metodo raggiunge una convergenza completa, con l’eccezione di SVM nel modello dove è stata valutata la feature importance. Questa situazione suggerirebbe la presenza di underfitting, indicando che l’aggiunta di nuovi dati potrebbe migliorare le prestazioni per questo distretto.

Considerazioni

Rispetto al distretto pelvico, l’analisi di questo nuovo set di dati ha rivelato che la progressiva riduzione della dimensionalità ha impattato sui risultati di classificazione. La diminuzione delle prestazioni tra il modello che considera tutte le ROI e quello che valuta le sole metriche della

ROI 110 suggerisce un contributo, seppur modesto, delle informazioni contenute nelle ROI di dimensioni minore, differenza non osservata nel distretto pelvico.

Più sorprendente è stato il significativo declino delle prestazioni a seguito della feature selection tramite feature importance: mentre nel distretto pelvico un numero di features pari a 4-6 è risultato sufficiente per eseguire il task di classificazione, in questo contesto sembrerebbe necessario valutare un numero di metriche maggiore per avere prestazioni più elevate.

Le ragioni del calo così evidente riscontrato nelle performance di questo distretto potrebbero riflettersi nella simultanea presenza di campioni ottenuti con dispositivi Varian e con dispositivi Elekta, i quali producono immagini qualitativamente molto differenti.

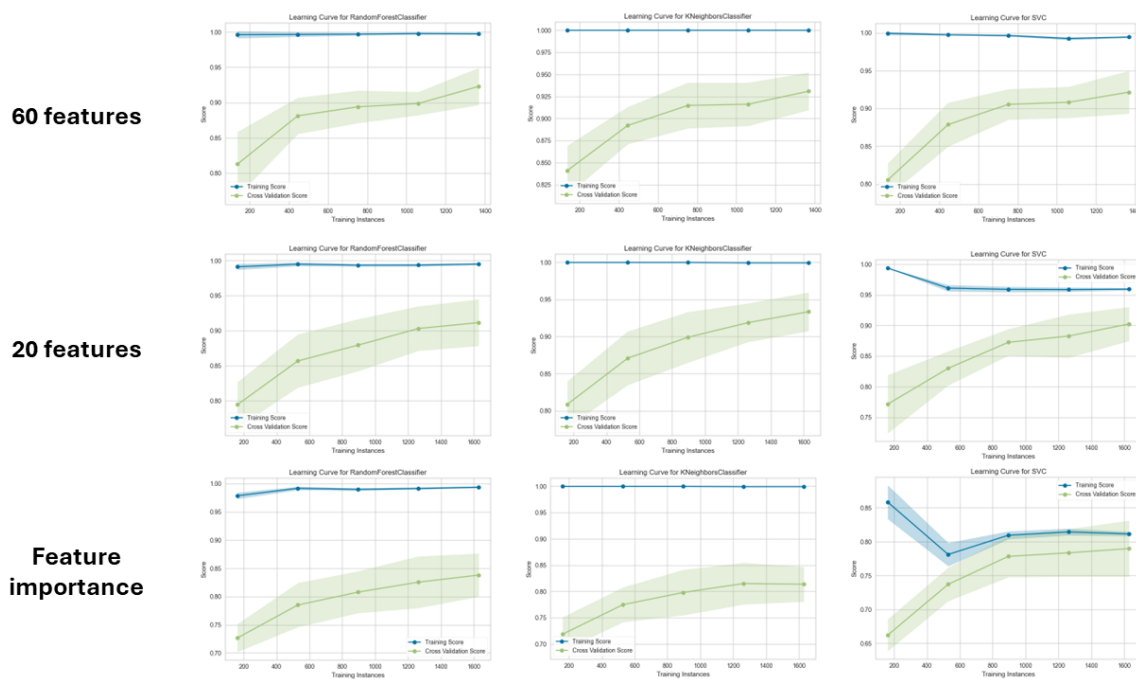


Figura 24. Curve di apprendimento per i tre modelli valutati sul distretto testa-collo.

Capitolo 5

Conclusioni

In questo lavoro di tesi è stato effettuato uno studio per individuare metodi e algoritmi per la selezione automatica di dataset destinati all'allenamento di una GAN che generi immagini di CT sintetiche a partire da coppie di immagini CT e di immagini CBCT deformate. Per raggiungere l'obiettivo sono state selezionate 20 metriche di valutazione dell'immagini su tre ROI circolari il cui centro è localizzato sul centroide della segmentazione dell'oggetto nell'immagine; i dati così ottenuti sono stati elaborati da tre metodi di Machine Learning, Random Forest, k-Nearest Neighbor e Support Vector Machine, per effettuare un task di classificazione binaria, che permettesse di distinguere in modo automatico le immagini da fornire in input alla GAN e quelle da scartare. L'allenamento è stato effettuato su tre modelli, uno che considerasse tutte le metriche valutate sulle ROI valutate (60 features), uno addestrato solo sulle features della ROI di dimensioni maggiori, uno con le feature selezionate tramite feature importance.

I metodi di Machine Learning si sono dimostrati efficaci nell'esecuzione del compito di classificazione e i modelli sviluppati si sono rivelati molto performanti, anche quando applicati a nuovi dataset. Inoltre, la strategia adottata consente una rapida selezione del dataset per il training della GAN, consentendo di ridurre significativamente i tempi di sviluppo.

Nonostante le performance siano nettamente superiori alle aspettative, è importante evidenziare le criticità osservate nel metodo adottato.

In primo luogo, il calcolo del centroide su cui viene centrata la ROI è fortemente influenzato dalla qualità della segmentazione dell'oggetto: l'ottimizzazione della soglia per il Global Thresholding risulta fondamentale per il corretto posizionamento della ROI e conseguentemente per un corretto confronto tra le immagini CT e CBCT.

Un altro limite significativo è dato dalla forte correlazione dei modelli con il dataset utilizzato: le performance nella classificazione sono influenzate dal dispositivo utilizzato per l'acquisizione.

Alla luce di queste considerazioni, si propongono alcune aree di sviluppo per il miglioramento della procedura adottata.

Effettuare una valutazione selettiva sulla numerosità dei dati potrebbe essere utile per esaminare l'impatto delle dimensioni del dataset sulle performance dei modelli sviluppati. Un ulteriore approfondimento si potrebbe attuare in uno studio sulla separazione dei dataset Varian ed Elekta nel distretto testa-collo, che in questa trattazione sono stati valutati insieme, sia nell'uso dei modelli parametrizzati sul distretto pelvi, sia durante un nuovo allenamento, riportando performance più basse nelle analisi menzionate rispetto a quanto ottenuto con il distretto pelvico. L'implementazione del procedimento adottato su nuovi distretti consentirebbe di proseguire il processo di validazione e di analisi su contesti differenti. Infine, si potrebbe valutare un unico modello per i vari distretti, separando i dati solo in base al dispositivo d'acquisizione.

Appendice A

Metriche

È riportata la tabella con le informazioni salienti delle metriche utilizzate come features nel testing dei modelli di Machine Learning.

Per X si intende l'immagine CBCT, per Y l'immagine CT.

<i>Metrica</i>	<i>Definizione</i>	<i>Misura</i>
ME: <i>Mean Error</i>	$ME = \frac{1}{N} \sum_{i=1}^N X_i - Y_i$	Errore medio sull'intensità dei pixel
MAE: <i>Mean Absolute Error</i>	$MAE = \frac{1}{N} \sum_{i=1}^N X_i - Y_i $	Errore medio assoluto sull'intensità dei pixel
MSE: <i>Mean Square Error</i>	$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2$	Errore medio quadratico sull'intensità dei pixel
PSNR: <i>Peak Signal-to Noise Ratio</i>	$PSNR = 10 \log_{10} \left(\frac{MAX^2}{\sqrt{MSE}} \right) \quad (1)$	Rapporto segnale-rumore di picco
SSIM: <i>Structural Similarity Metric</i>	$SSIM = \frac{(2\mu_x\mu_y+c1)(2\sigma_{xy}+c2)}{(\mu_x^2+\mu_y^2+c1)(\sigma_x^2+\sigma_y^2+c2)} \quad (2)$	Similarità strutturale
RMSE: <i>Root Mean Square Error</i>	$RMSE = \sqrt{MSE}$	Radice dell'errore quadratico medio
NCC: <i>Normalized Cross-Correlation</i>	$NCC = \frac{\sum_{i,j} (Y_{i,j} - \mu_y)(X_{i,j} - \mu_x)}{\sqrt{\sum_{i,j} (Y_{i,j} - \mu_y)^2 \cdot \sum_{i,j} (X_{i,j} - \mu_x)^2}}$	Similarità tra le intensità dei pixel
DSC: <i>Dice Similarity Coefficient</i>	$DSC = \frac{2 X \cap Y }{ X + Y }$	Sovrapposizione geometrica
MSE: <i>Mean Squared Gradient Error</i>	$MSG_E = \frac{1}{N} \sum_{i=1}^N (\nabla X_i - \nabla Y_i)^2$	Errore quadratico medio sui gradienti dei pixel
MI: <i>Mutual Information</i>	$MI(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$	Dipendenza statistica
JE: <i>Joint Entropy</i>	$JE(X,Y) = - \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2(p(x,y))$	Incertezza congiunta tra le distribuzioni di intensità

¹ MAX: valore massimo del range dei pixel nell'immagine.

² c1 = (0.01·MAX)²; c2 = (0.03·MAX)²

PCC: <i>Pearson Correlation Coefficient</i>	$PCC = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	Correlazione lineare tra i valori dei pixel
ED: Entropy Difference	$ED = - \sum_{i=1}^N p_i \cdot \log_2(p_i) \Big _y + \sum_{i=1}^N p_i \cdot \log_2(p_i) \Big _x$	Differenza di casualità tra le distribuzioni dei livelli di grigio
RUI: Relative Uniformity Index	$RUI = \frac{\sigma_y / \mu_y - \sigma_x / \mu_x}{\sigma_y / \mu_y}$	Differenza relativa nell'uniformità delle intensità dei pixel
FSIM: Feature Similarity Index	$FSIM = [S_{PC}]^\alpha \cdot [S_G]^\beta \text{ } ^{(3)}$	Similarità strutturale e delle features
SRE: Signal to Reconstruction Error	$SRE = 10 \log_{10} \left(\frac{\sum (Y_{i,j})^2}{\sum (Y_{i,j} - X_{i,j})^2} \right)$	Errore di ricostruzione
VIF: Visual Information Fidelity	Implementato con libreria sewar [65]	Fedeltà delle informazioni visive tra due immagini
MSSSIM: Multi-Scale Structural Similarity Index	Implementato con libreria sewar [65]	Similarità a diverse scale spaziali
SAM: Spectral Angle Mapper	Implementato con libreria sewar [65]	Similarità tra le intensità dei pixel di due immagini

³ S_{PC} : similarità tra la congruenza di fase; S_G : similarità tra le intensità dei gradienti; α, β : parametri di regolarizzazione [66]

Appendice B

Grafici delle distribuzioni

B.1 Immagini Varian

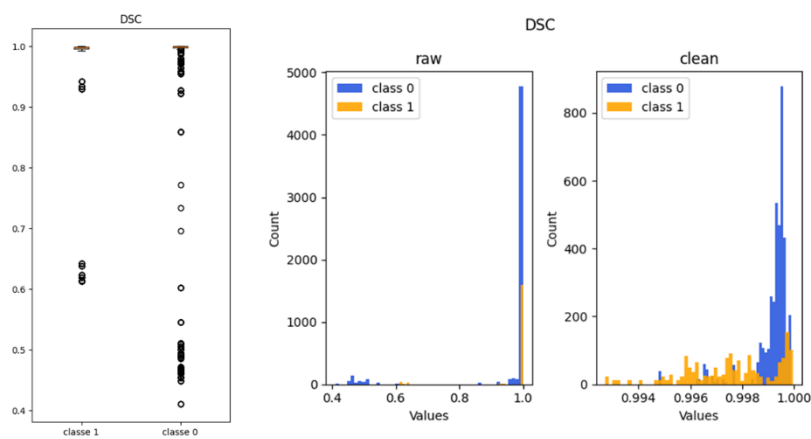


Figura 25. Distribuzioni dei valori: DSC.

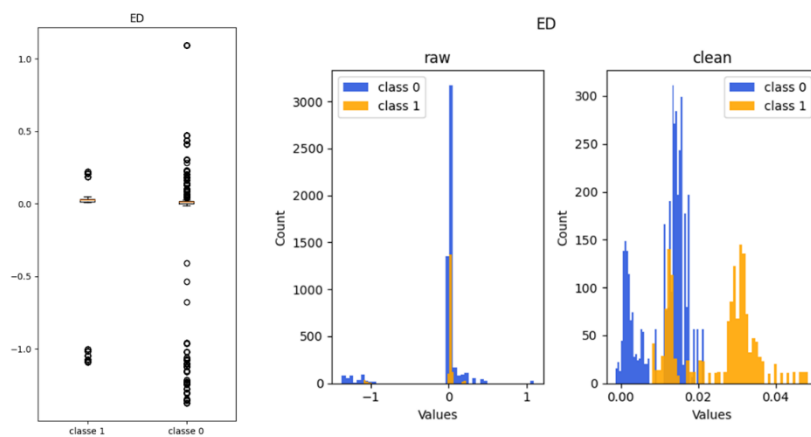


Figura 26. Distribuzioni distribuzione dei valori: ED.

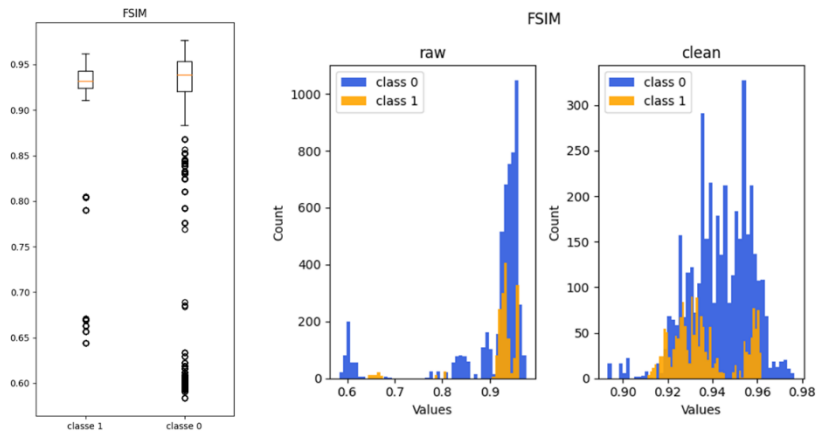


Figura 27. Distribuzioni dei valori: FSIM.

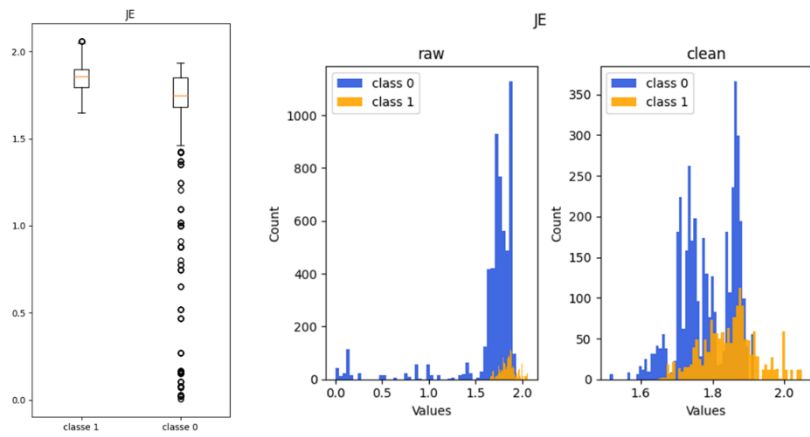


Figura 28. Distribuzioni dei valori: JE.

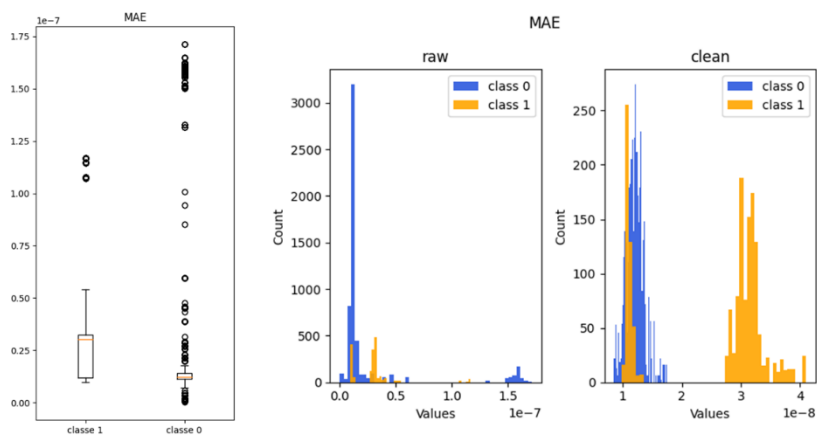


Figura 29. Distribuzioni dei valori: MAE.

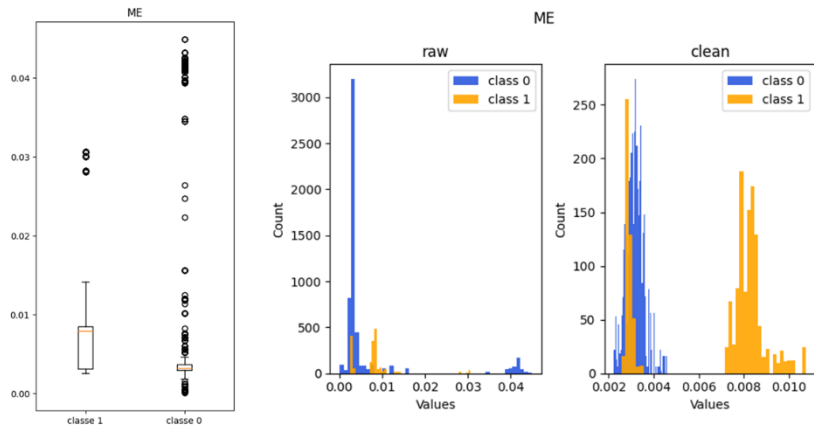


Figura 30. Distribuzioni dei valori: ME.

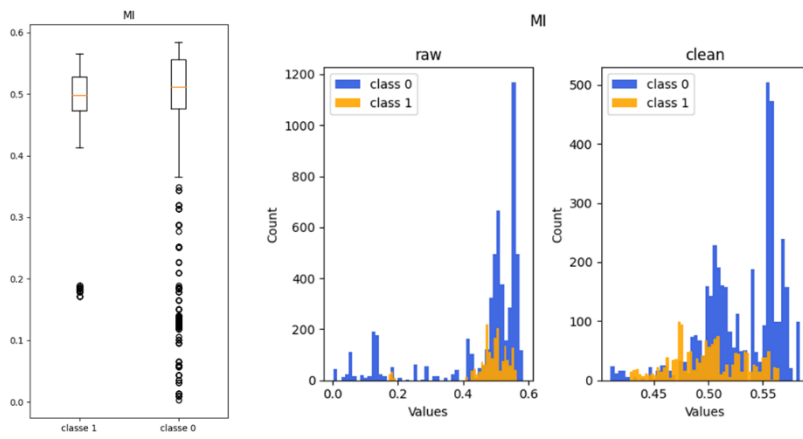


Figura 31. Distribuzioni dei valori: MI.

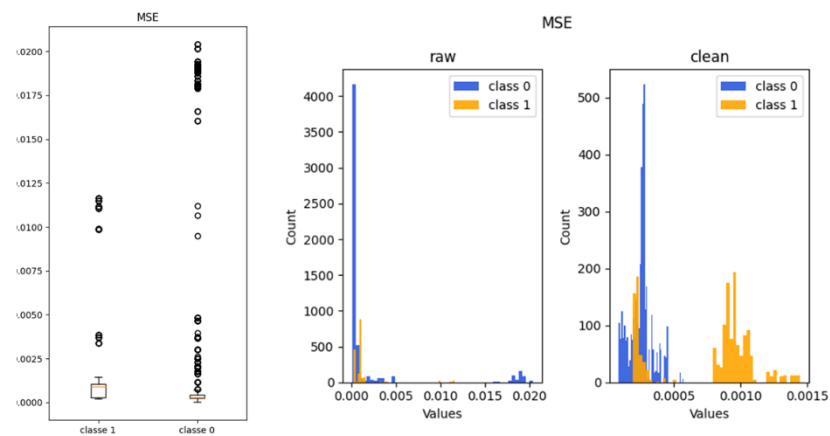


Figura 32. Distribuzioni dei valori: MSE.

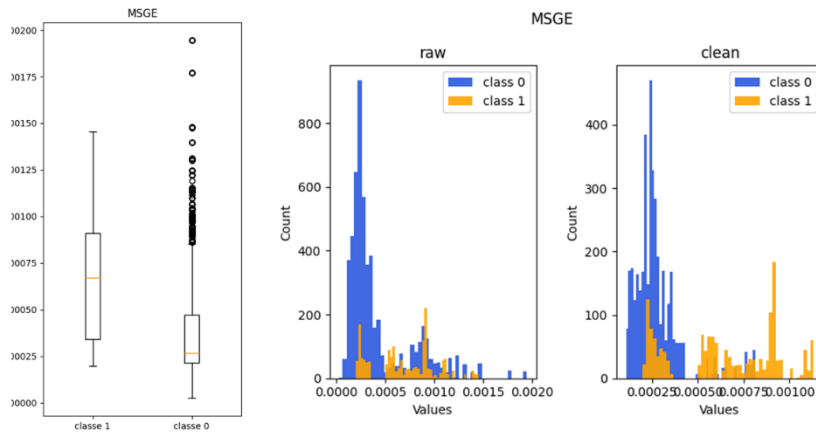


Figura 33. Distribuzioni dei valori: MSGE.

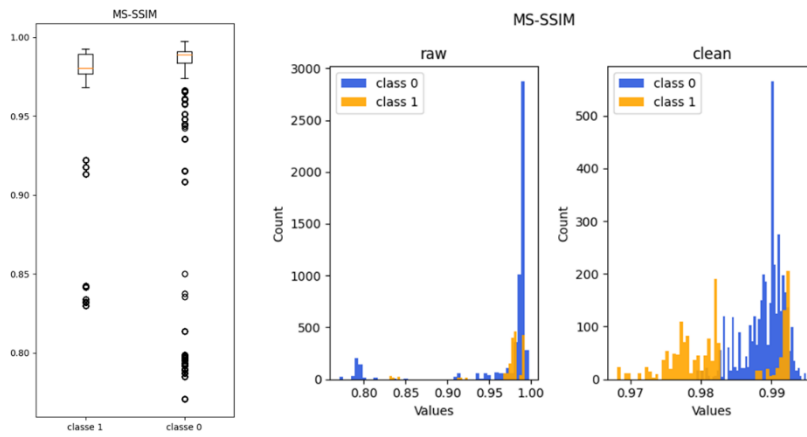


Figura 34. Distribuzioni dei valori: MS-SSIM

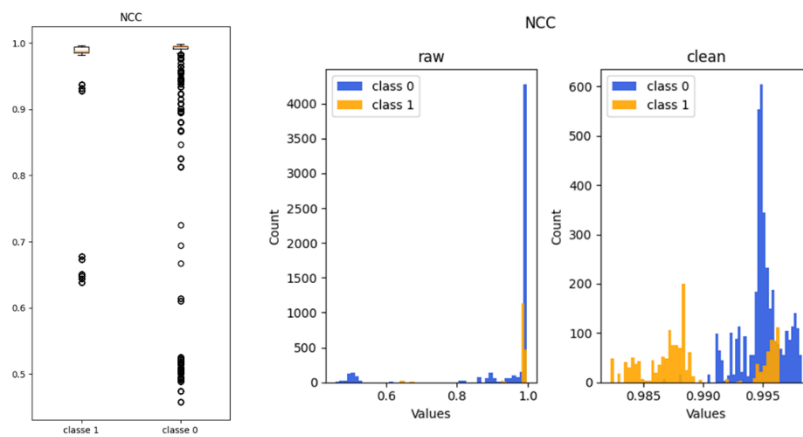


Figura 35. Distribuzioni dei valori: NCC.

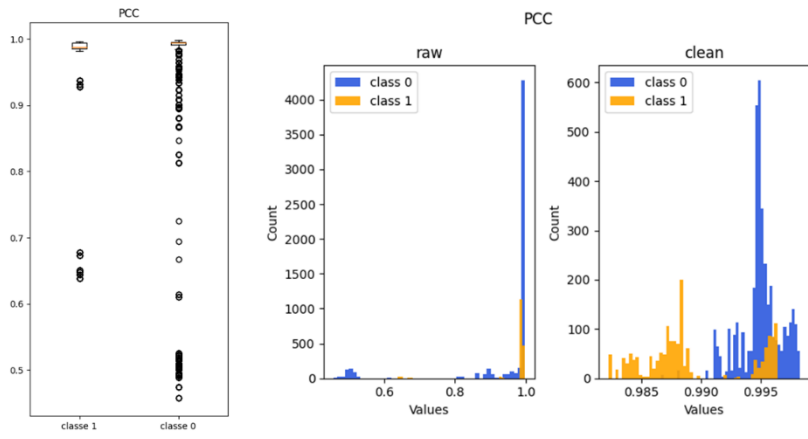


Figura 36. Distribuzioni dei valori: PCC.

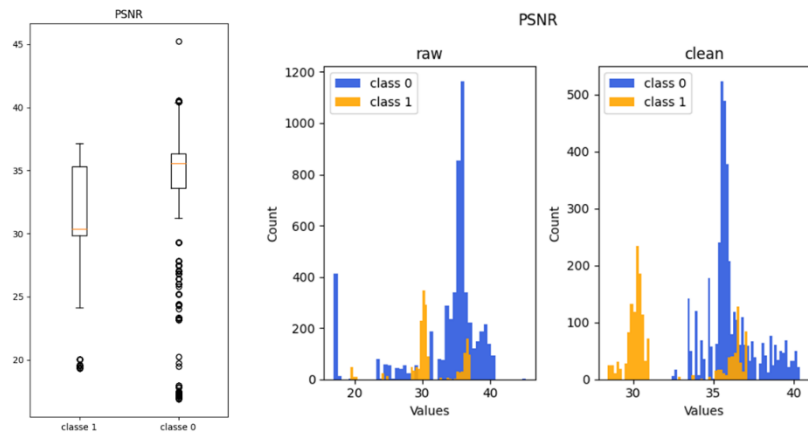


Figura 37. Distribuzioni dei valori: PSNR.

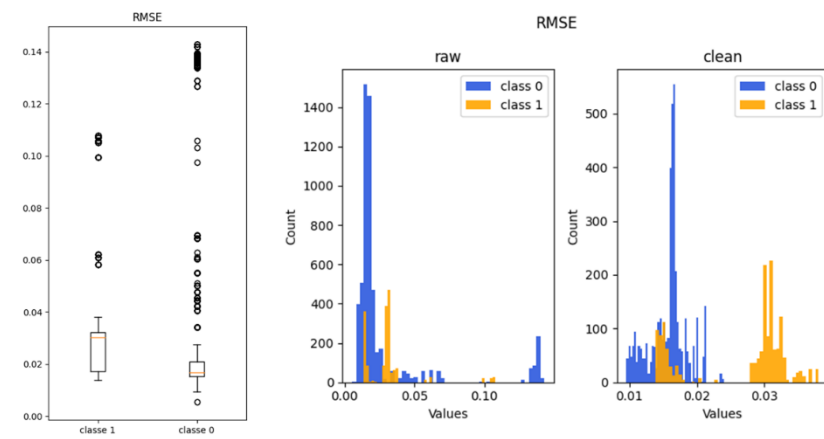


Figura 38. Distribuzioni dei valori: RMSE.

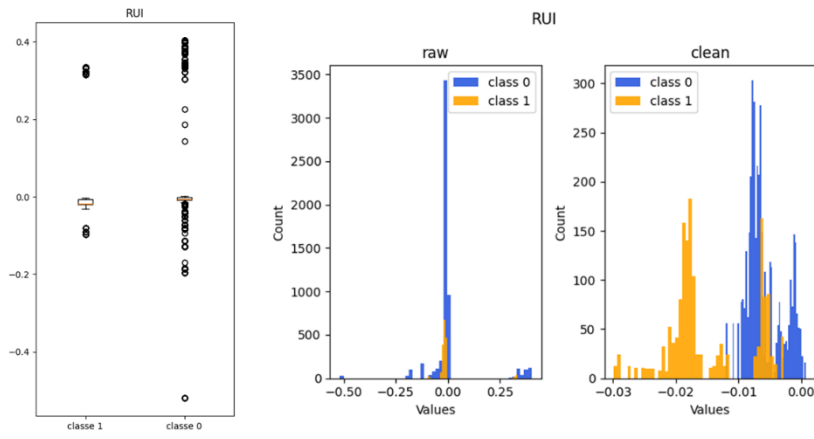


Figura 39. Distribuzioni dei valori: RUI.

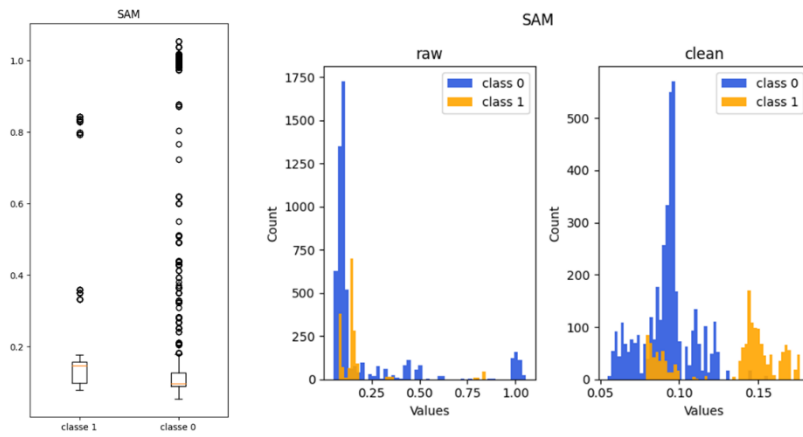


Figura 40. Distribuzioni dei valori: SAM.

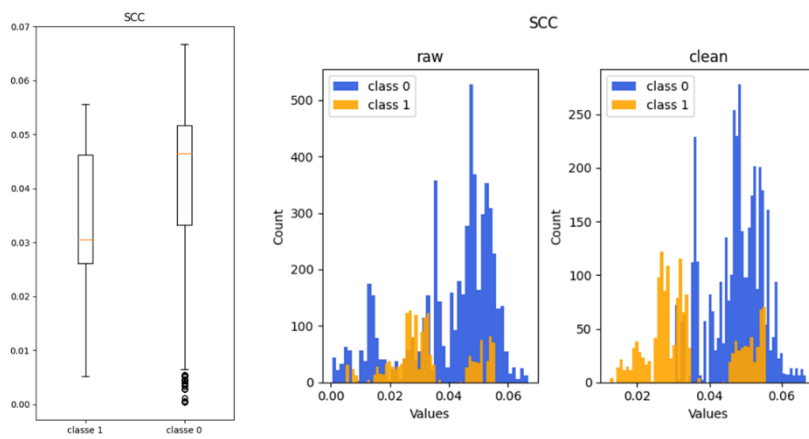


Figura 41. Distribuzioni dei valori: SCC.

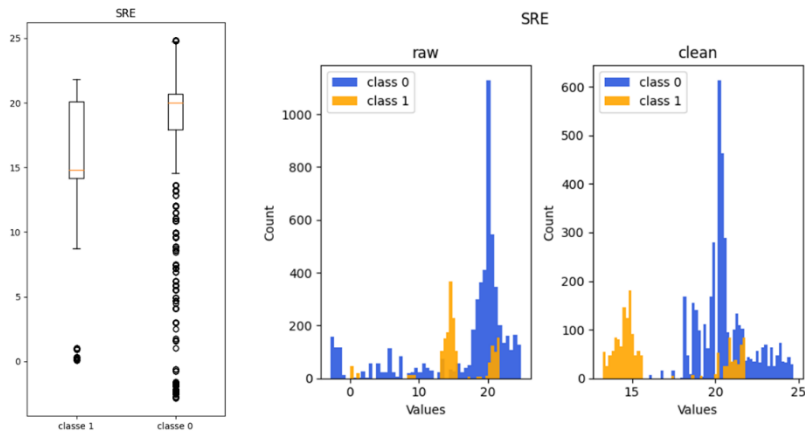


Figura 42. Distribuzioni dei valori: SRE.

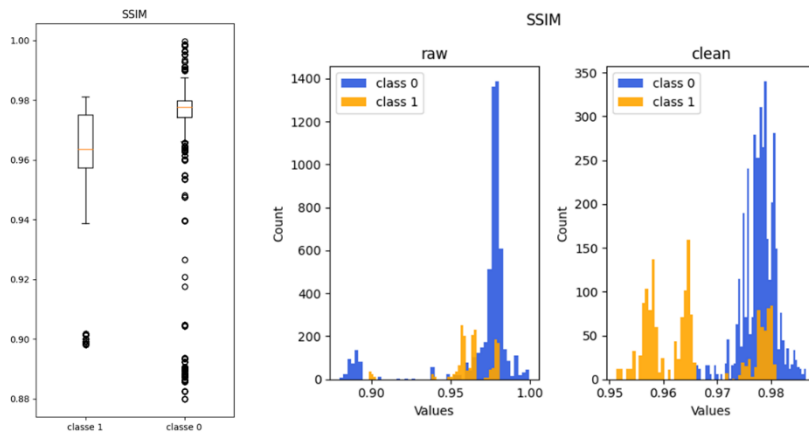


Figura 43. Distribuzioni dei valori: SSIM.

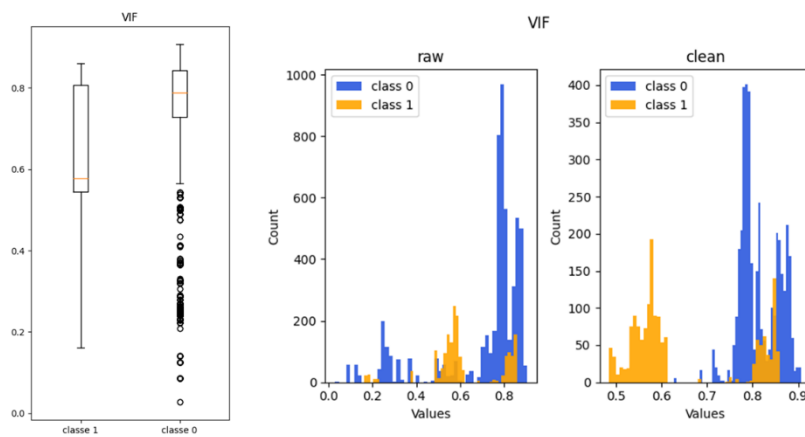


Figura 44. Distribuzioni dei valori: VIF.

B.2 Immagini Elekta

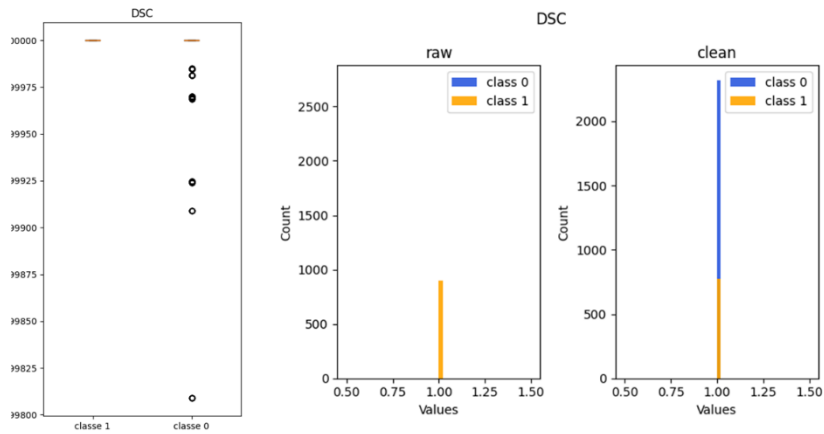


Figura 45. Distribuzioni dei valori: DSC.

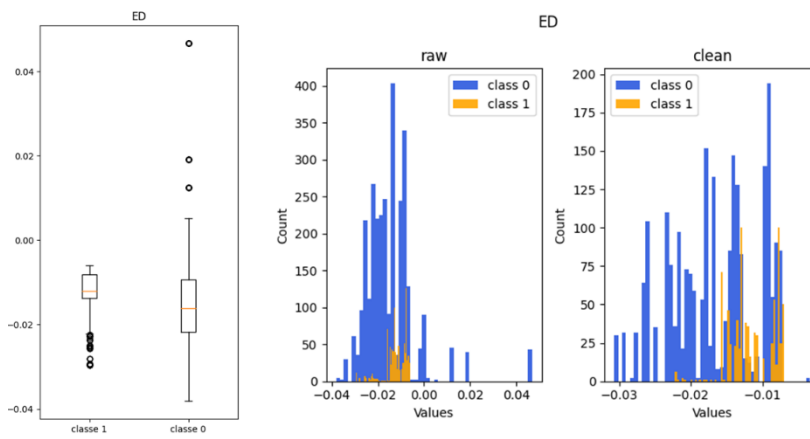


Figura 46. Distribuzioni dei valori: ED

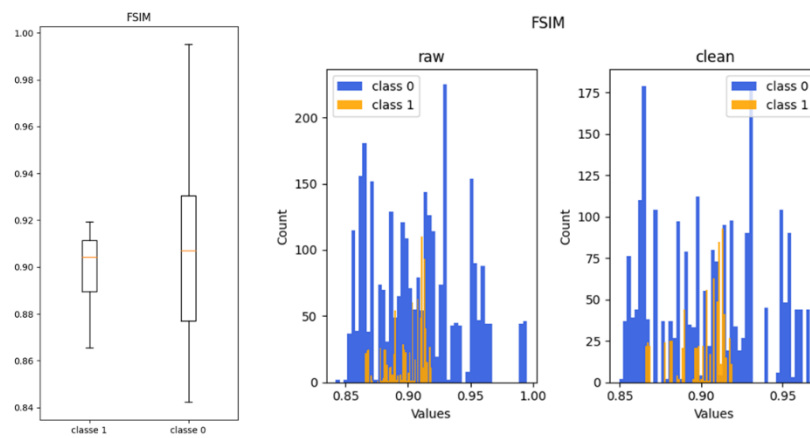


Figura 47. Distribuzioni dei valori: FSIM.

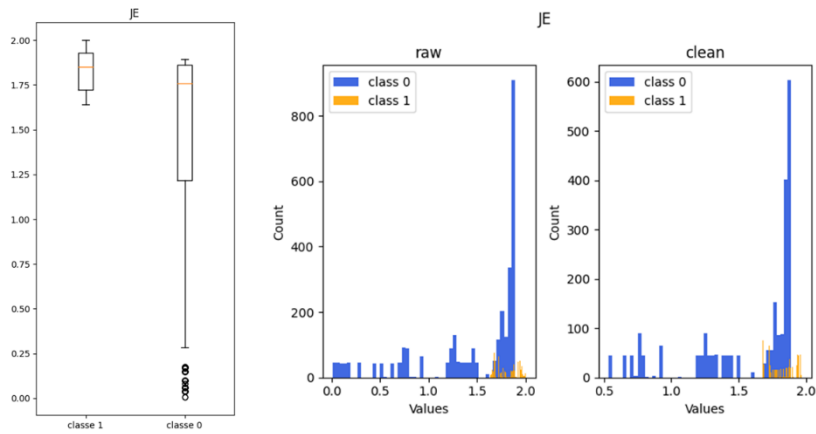


Figura 48. Distribuzioni dei valori: JE.

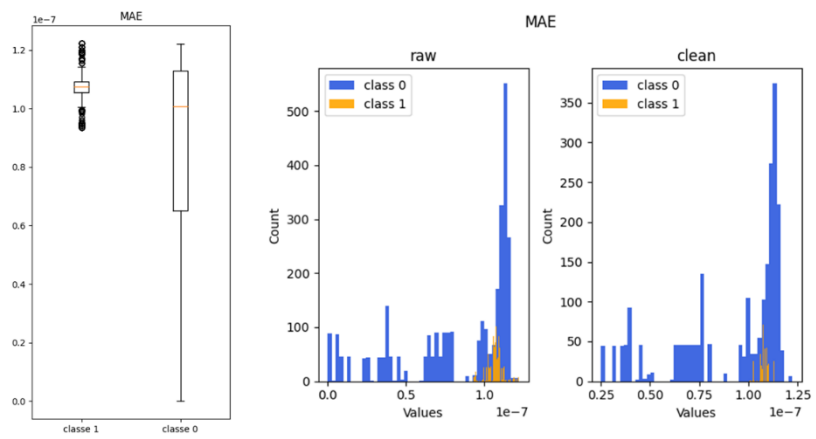


Figura 49. Distribuzioni dei valori: MAE.

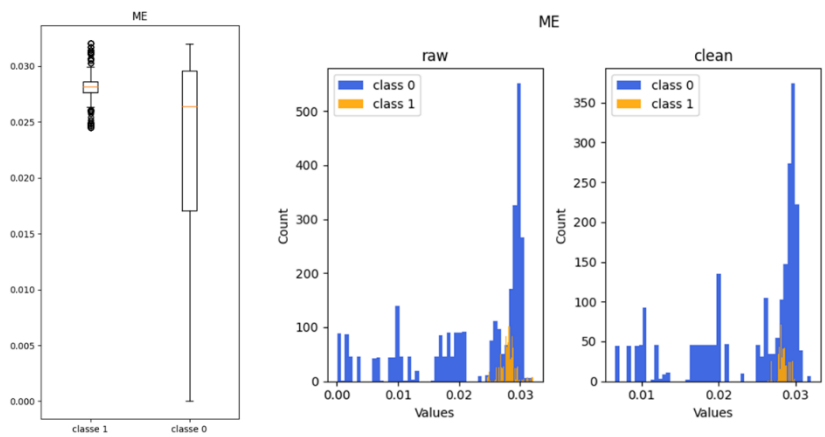


Figura 50. Distribuzioni dei valori: ME.

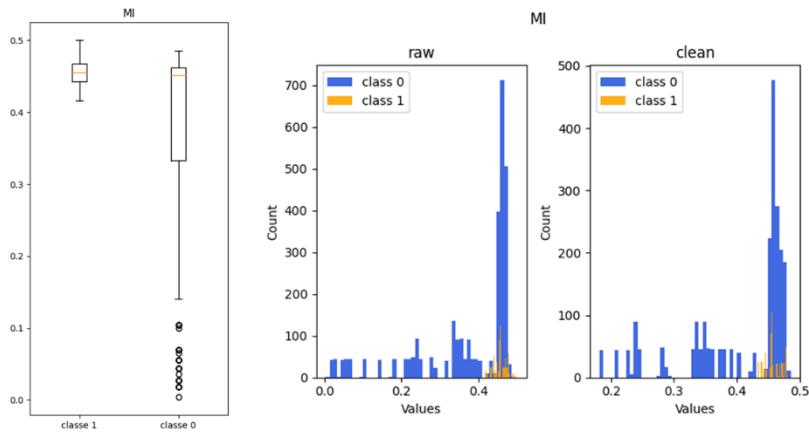


Figura 51. Distribuzioni dei valori: MI.

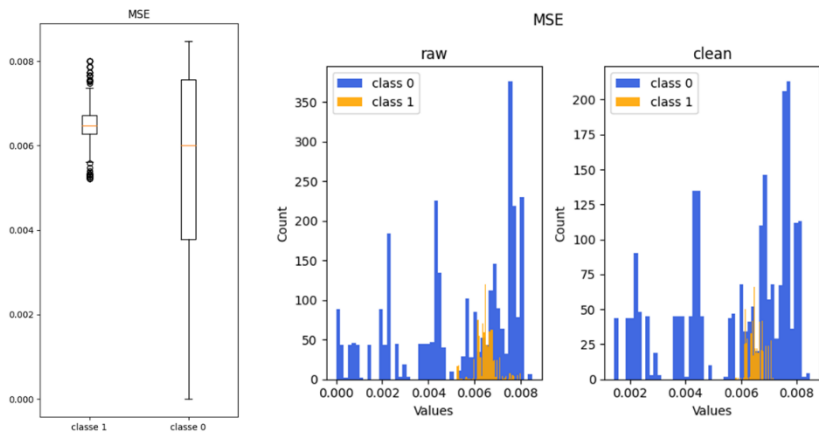


Figura 52. Distribuzioni dei valori: MSE.

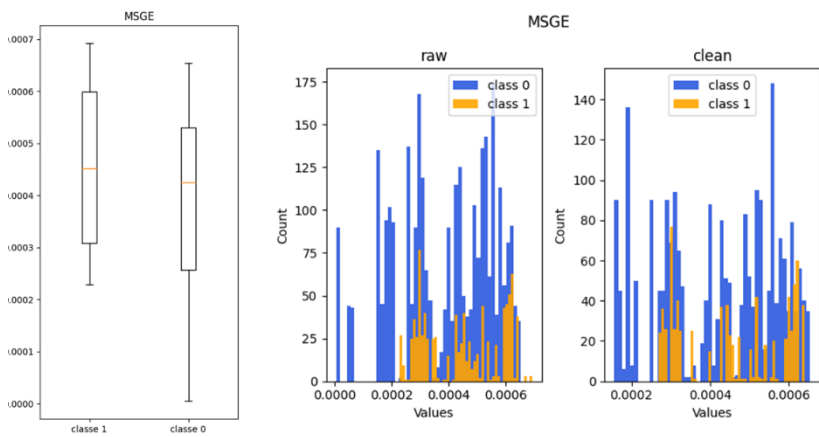


Figura 53. Distribuzioni dei valori: MSGE.

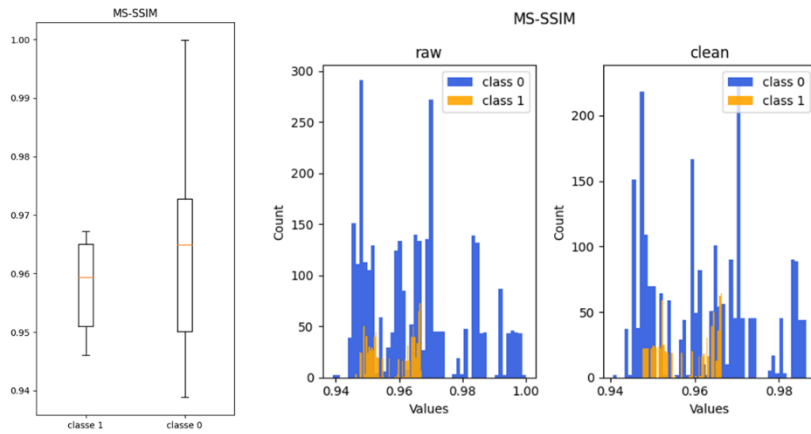


Figura 54. Distribuzioni dei valori: MS-SSIM.

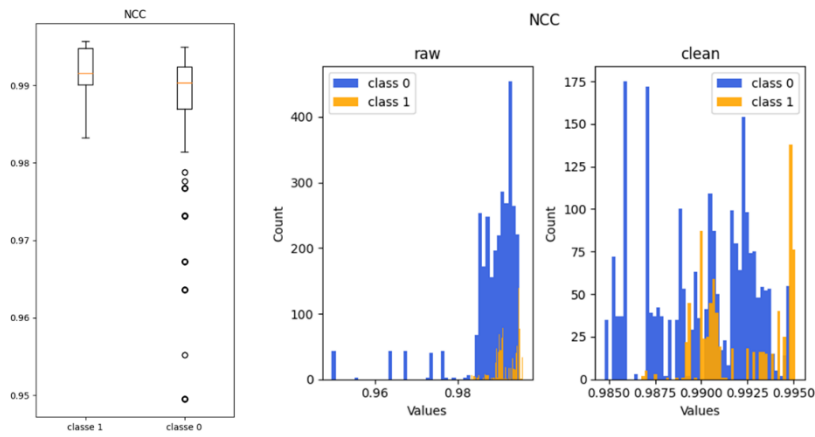


Figura 55. Distribuzioni dei valori: NCC.

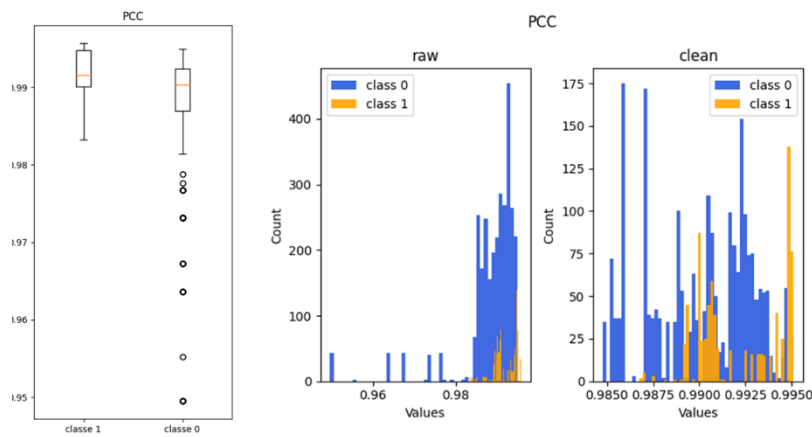


Figura 56. Distribuzioni dei valori: PCC.

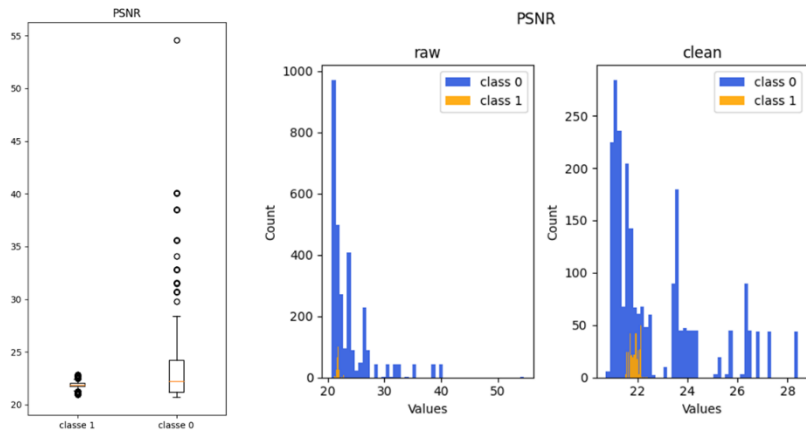


Figura 57. Distribuzioni dei valori: PSNR.

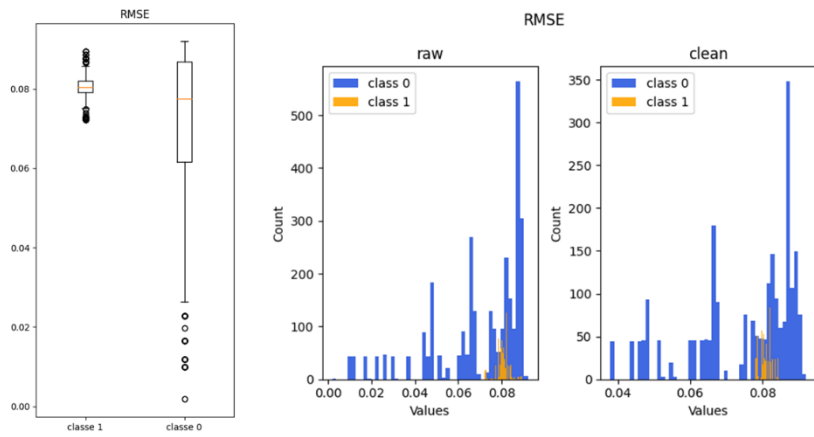


Figura 58. Distribuzioni dei valori: RMSE.

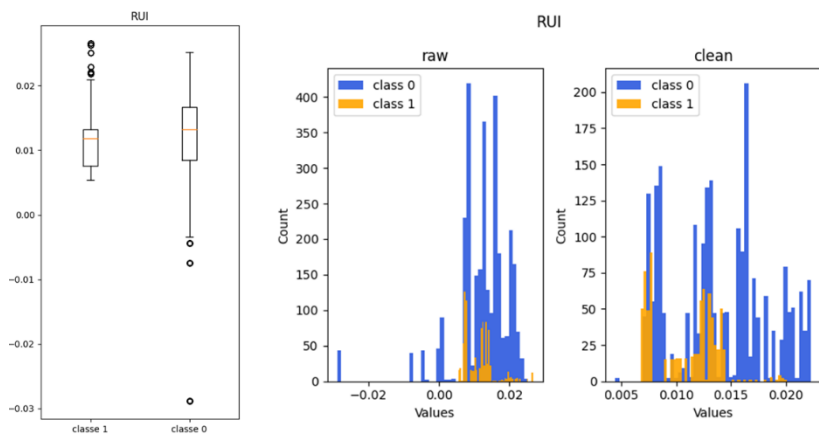


Figura 59. Distribuzioni dei valori: RUI.

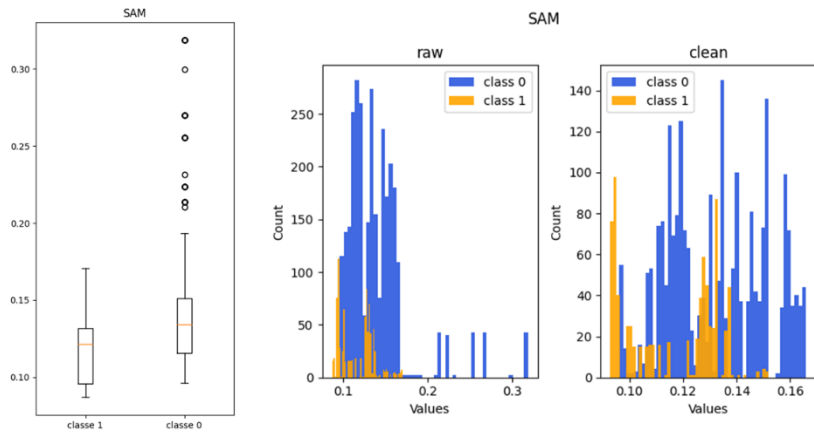


Figura 60. Distribuzioni dei valori: SAM.

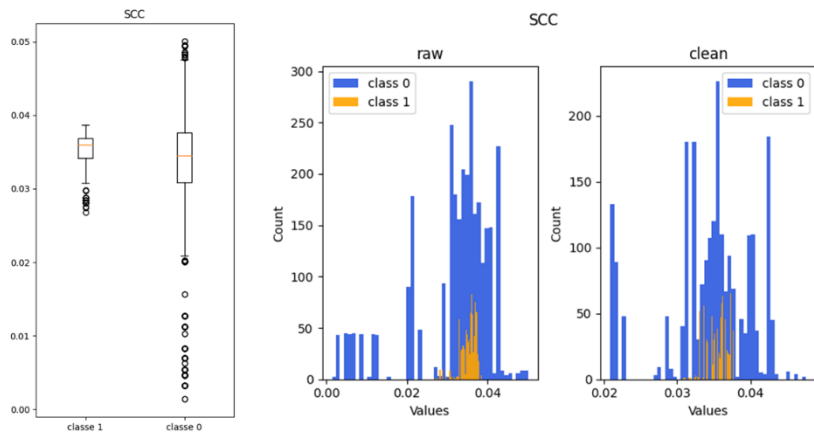


Figura 61. Distribuzioni dei valori: SCC.

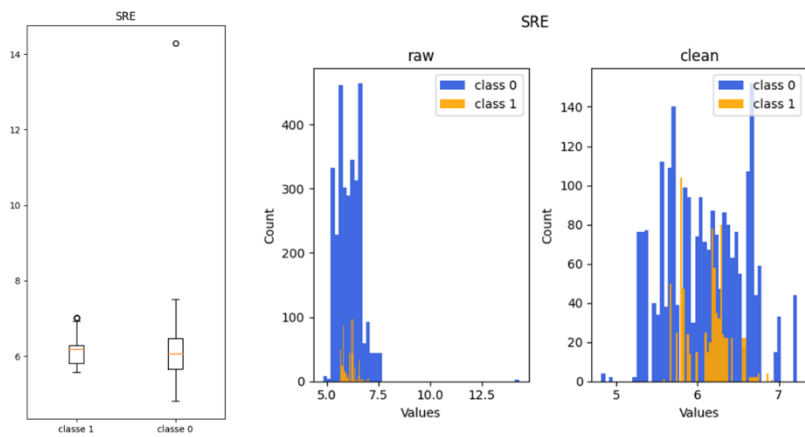


Figura 62. Distribuzioni dei valori: SRE.

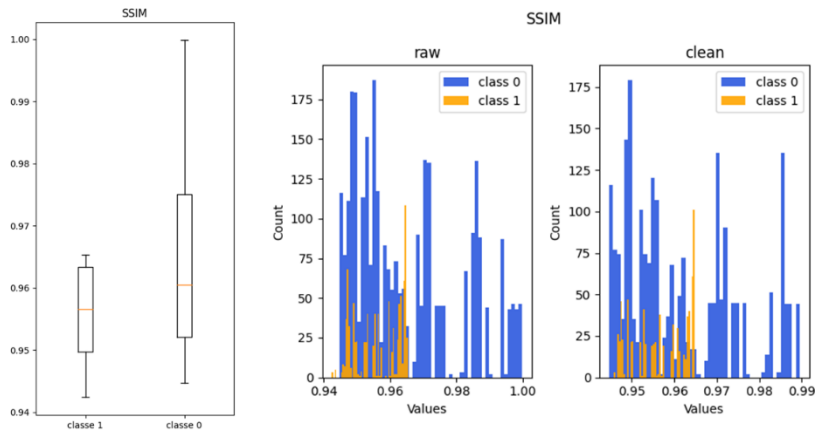


Figura 63. Distribuzioni dei valori: SSIM.

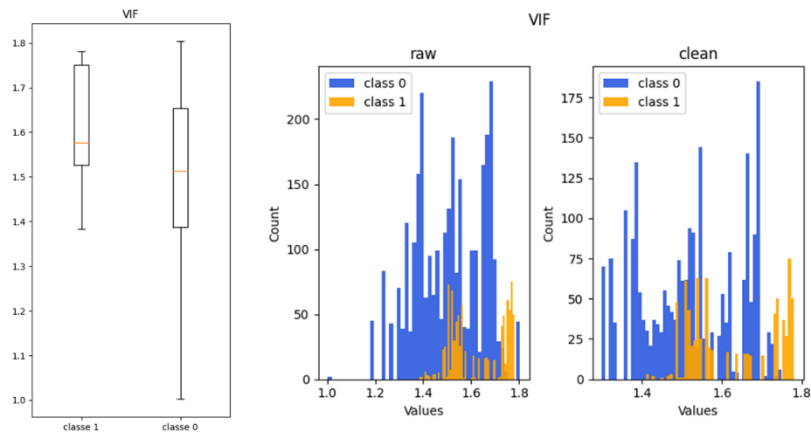


Figura 64. Distribuzioni dei valori: VIF.

Bibliografia

- [1] Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. *Int J Cancer*. 2021;149(4):778–89.
- [2] Culp MBB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur Urol* [Internet]. 2020;77(1):38–52. Available from: <https://doi.org/10.1016/j.eururo.2019.08.005>
- [3] Hamdy FC, Donovan JL, Lane JA, Metcalfe C, Davis M, Turner EL, et al. Fifteen-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *N Engl J Med*. 2023;388(17):1547–58.
- [4] Kulik A, Dąbkowski M. Radioterapia raka gruczołu krokowego. *Współczesna Onkol*. 2011;15(5):317–29.
- [5] Dang A, Kupelian PA, Cao M, Agazaryan N, Kishan AU. Image-guided radiotherapy for prostate cancer. *Transl Androl Urol*. 2018;7(3):308–20.
- [6] Murgić J, Gregov M, Mrčela I, Budanec M, Krengli M, Fröbe A, et al. Mri-Guided Radiotherapy for Prostate Cancer: a New Paradigm. *Acta Clin Croat*. 2022;61:65–70.
- [7] Ariyaratne H, Chesham H, Pettingell J, Alonzi R. Image-guided radiotherapy for prostate cancer with cone beam CT: dosimetric effects of imaging frequency and PTV margin. *Radiother Oncol* [Internet]. 2016;121(1):103–8. Available from: <http://dx.doi.org/10.1016/j.radonc.2016.07.018>
- [8] Boulanger M, Nunes JC, Chourak H, Largent A, Tahri S, Acosta O, et al. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. *Phys Medica* [Internet]. 2021;89(August):265–81. Available from: <https://doi.org/10.1016/j.ejmp.2021.07.027>

- [9] Webster A, Appelt AL, Eminowicz G. Image-Guided Radiotherapy for Pelvic Cancers: A Review of Current Evidence and Clinical Utilisation. *Clin Oncol* [Internet]. 2020;32(12):805–16. Available from: <https://doi.org/10.1016/j.clon.2020.09.010>
- [10] Wong JR, Gao Z, Uematsu M, Merrick S, Machernis NP, Chen T, et al. Interfractional Prostate Shifts: Review of 1870 Computed Tomography (CT) Scans Obtained During Image-Guided Radiotherapy Using CT-on-Rails for the Treatment of Prostate Cancer. *Int J Radiat Oncol Biol Phys*. 2008;72(5):1396–401.
- [11] Byrne TE. A review of prostate motion with considerations for the treatment of prostate cancer. *Med Dosim*. 2005;30(3):155–61.
- [12] Huang K, Palma DA, Scott D, McGregor D, Gaede S, Yartsev S, et al. Inter- and intrafraction uncertainty in prostate bed image-guided radiotherapy. *Int J Radiat Oncol Biol Phys* [Internet]. 2012;84(2):402–7. Available from: <http://dx.doi.org/10.1016/j.ijrobp.2011.12.035>
- [13] Alaei P, Spezi E. Imaging dose from cone beam computed tomography in radiation therapy. *Phys Medica* [Internet]. 2015;31(7):647–58. Available from: <http://dx.doi.org/10.1016/j.ejmp.2015.06.003>
- [14] Hvid CA, Elstrøm U V., Jensen K, Grau C. Cone-beam computed tomography (CBCT) for adaptive image guided head and neck radiation therapy. *Acta Oncol (Madr)* [Internet]. 2018;57(4):552–6. Available from: <https://doi.org/10.1080/0284186X.2017.1398414>
- [15] Iliopoulos P, Simopoulou V, Simopoulos, G. K, Theodorou K. Review on Cone Beam Computed Tomography (CBCT) Dose in Patients Undergoing Image Guided Radiotherapy (IGRT). *Dosim Radiopharm - New Adv Needs* [Working Title] IntechOpen, Nov 07 [Internet]. 2023; Available from: doi: 10.5772/intechopen.1002683
- [16] Scarfe WC, Farman AG. What is Cone-Beam CT and How Does it Work? *Dent Clin North Am*. 2008;52(4):707–30.
- [17] Venkatesh, E., & Elluru SV. CONE BEAM COMPUTED TOMOGRAPHY: BASICS AND APPLICATIONS IN DENTISTRY. *Eur Oral Res*. 2017;51:102–21.
- [18] Eckl M, Hoppen L, Sarria GR, Boda-Heggemann J, Simeonova-Chergou A, Steil V, et

- al. Evaluation of a cycle-generative adversarial network-based cone-beam CT to synthetic CT conversion algorithm for adaptive radiation therapy. *Phys Medica* [Internet]. 2020;80(October):308–16. Available from: <https://doi.org/10.1016/j.ejmp.2020.11.007>
- [19] Schröder L, Stankovic U, Remeijer P, Sonke JJ. Evaluating the impact of cone-beam computed tomography scatter mitigation strategies on radiotherapy dose calculation accuracy. *Phys Imaging Radiat Oncol* [Internet]. 2019;10(November 2018):35–40. Available from: <https://doi.org/10.1016/j.phro.2019.04.001>
- [20] Oyama A, Kumagai S, Arai N, Takata T, Saikawa Y, Shiraishi K, et al. Image quality improvement in cone-beam CT using the super-resolution technique. *J Radiat Res.* 2018;59(4):501–10.
- [21] Schulze R, Heil U, Groß D, Bruellmann DD, Dranischnikow E, Schwanecke U, et al. Artefacts in CBCT: A review. *Dentomaxillofacial Radiol.* 2011;40(5):265–73.
- [22] Nagarajappa A, Dwivedi N, Tiwari R. Artifacts: The downturn of CBCT image. *J Int Soc Prev Community Dent.* 2015;5(6):440.
- [23] de Hond YJM, Kerckhaert CEM, van Eijnatten MAJM, van Haaren PMA, Hurkmans CW, Tijssen RHN. Anatomical evaluation of deep-learning synthetic computed tomography images generated from male pelvis cone-beam computed tomography. *Phys Imaging Radiat Oncol* [Internet]. 2023;25(January):100416. Available from: <https://doi.org/10.1016/j.phro.2023.100416>
- [24] Zhang Y, Yue N, Su MY, Liu B, Ding Y, Zhou Y, et al. Improving CBCT quality to CT level using deep learning with generative adversarial network. *Med Phys.* 2021;48(6):2816–26.
- [25] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol.* 2018;63(18).
- [26] Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. *Med Phys.* 2020;47(3):1115–25.

- [27] Maspero M, Houweling AC, Savenije MHF, van Heijst TCF, Verhoeff JJC, Kotte ANTJ, et al. A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. *Phys Imaging Radiat Oncol* [Internet]. 2020;14(April):24–31. Available from: <https://doi.org/10.1016/j.phro.2020.04.002>
- [28] Gao L, Xie K, Wu X, Lu Z, Li C, Sun J, et al. Generating synthetic CT from low-dose cone-beam CT by using generative adversarial networks for adaptive radiotherapy. *Radiat Oncol* [Internet]. 2021;16(1):1–16. Available from: <https://doi.org/10.1186/s13014-021-01928-w>
- [29] Jiang Y, Zhang Y, Luo C, Yang P, Wang J, Liang X, et al. A generalized image quality improvement strategy of cone-beam CT using multiple spectral CT labels in Pix2pix GAN. *Phys Med Biol*. 2022;67(11).
- [30] Deng L, Zhang M, Wang J, Huang S, Yang X. Improving cone-beam CT quality using a cycle-residual connection with a dilated convolution-consistent generative adversarial network. *Phys Med Biol*. 2022;67(14).
- [31] Suwanraksa C, Bridhikitti J, Liamsuwan T, Chaichulee S. CBCT-to-CT Translation Using Registration-Based Generative Adversarial Networks in Patients with Head and Neck Cancer. *Cancers (Basel)*. 2023;15(7):1–23.
- [32] Peng J, Qiu RLJ, Wynne JF, Chang CW, Pan S, Wang T, et al. CBCT-Based synthetic CT image generation using conditional denoising diffusion probabilistic model. *Med Phys*. 2023;
- [33] Rusanov B, Hassan GM, Reynolds M, Sabet M, Rowshanfarzad P, Bucknell N, et al. Transformer CycleGAN with uncertainty estimation for CBCT based synthetic CT in adaptive radiotherapy. *Phys Med Biol*. 2024;69(3).
- [34] Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. A state-of-the-art survey on deep learning theory and architectures. *Electron*. 2019;8(3):1–67.
- [35] Kumar M, Shanavas M, Sidappa A, Kiran M. Cone beam computed tomography - know its secrets. *J Int oral Heal JIOH* [Internet]. 2015;7(2):64–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25859112><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4377156>

- [36] Pauwels R, Araki K, Siewerdsen JH, Thongvigitmanee SS. Technical aspects of dental CBCT: State of the art. *Dentomaxillofacial Radiol.* 2015;44(1):1–20.
- [37] Tamás Szabó B, Dobai A, Dobo-Nagy C. Cone-Beam Computed Tomography in Dentomaxillofacial Radiology. *Nov Imaging Spectrosc IntechOpen.* 2012;
- [38] Mc Parland NA. kV-Cone Beam CT as an IGRT Tool in the Treatment of Early Stage Prostate Cancer: A Literature Review. *J Med Imaging Radiat Sci* [Internet]. 2009;40(1):9–14. Available from: <http://dx.doi.org/10.1016/j.jmir.2008.12.002>
- [39] Srnivasan K, Mohammadi M, Shepherd J. Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review. *Polish J Radiol.* 2014;79:181–93.
- [40] Groh BA, Siewerdsen JH, Drake DG, Wong JW, Jaffray DA. A performance comparison of flat-panel imager-based MV and kV cone-beam CT. *Med Phys.* 2002;29(6):967–75.
- [41] Lehmann J, Perks J, Semon S, Purdy JA. Commissioning experience with cone-beam computed tomography for image guided radiation therapy. *J Appl Clin Med Phys.* 2007;8(3):21–34.
- [42] Hatcher DC, Dugoni AA. Operational principles for cone-beam computed tomography. *J Am Dent Assoc* [Internet]. 2010;141(10 SUPPL.):3S-6S. Available from: <http://dx.doi.org/10.14219/jada.archive.2010.0359>
- [43] Kiljunen T, Kaasalainen T, Suomalainen A, Kortensniemi M. Dental cone beam CT: A review. *Phys Medica* [Internet]. 2015;31(8):844–60. Available from: <http://dx.doi.org/10.1016/j.ejmp.2015.09.004>
- [44] Sajja S, Lee Y, Eriksson M, Nordström H, Sahgal A, Hashemi M, et al. Technical Principles of Dual-Energy Cone Beam Computed Tomography and Clinical Applications for Radiation Therapy. *Adv Radiat Oncol* [Internet]. 2020;5(1):1–16. Available from: <https://doi.org/10.1016/j.adro.2019.07.013>
- [45] Pauwels R, Nackaerts O, Bellaiche N, Stamatakis H, Tsiklakis K, Walker A, et al. Variability of dental cone beam CT grey values for density estimations. *Br J Radiol.* 2013;86(1021):1–9.
- [46] Onur TO. An application of filtered back projection method for computed tomography

images. *Int Rev Appl Sci Eng.* 2021;12(2):194–200.

- [47] Lechuga L, Weidlich GA. Cone Beam CT vs. Fan Beam CT: A Comparison of Image Quality and Dose Delivered Between Two Differing CT Imaging Modalities. *Cureus.* 2016;8(9).
- [48] Zhang Y, Zhang L, Zhu XR, Lee AK, Chambers M, Dong L. Reducing metal artifacts in cone-beam CT images by preprocessing projection data. *Int J Radiat Oncol Biol Phys.* 2007;67(3):924–32.
- [49] Udousoro. Machine Learning: Review. *Int J Adv Res Sci Commun Technol.* 2020;02(02):251–6.
- [50] Sharifani K, Amini M. Machine Learning and Deep Learning: A Review of Methods and Applications. *World Inf Technol Eng J.* 2023;10(07):3897–904.
- [51] Khamparia A, Singh KM. A systematic review on deep learning architectures and applications. *Expert Syst.* 2019;36(3):1–22.
- [52] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions [Internet]. Vol. 8, *Journal of Big Data.* Springer International Publishing; 2021. Available from: <https://doi.org/10.1186/s40537-021-00444-8>
- [53] Talekar B. A Detailed Review on Decision Tree and Random Forest. *Biosci Biotechnol Res Commun.* 2020;13(14):245–8.
- [54] Khan MY, Qayoom A, Nizami MS, Siddiqui MS, Wasi S, Raazi SMKUR. Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complexity.* 2021;2021.
- [55] Breiman L. Random Forests. *Mach Learn [Internet].* 2001;45(1):5–32. Available from: <https://doi.org/10.1023/A:1010933404324>
- [56] Taunk K, De S, Verma S, Swetapadma A. A brief review of nearest neighbor algorithm for learning and classification. 2019 *Int Conf Intell Comput Control Syst ICCS 2019.* 2019;(Iciccs):1255–60.
- [57] Sharma S, Parmar M. Heart Diseases Prediction using Deep Learning Neural Network

Model. *Int J Innov Technol Explor Eng.* 2020;9(3):2244–8.

- [58] Somvanshi M, Chavan P, Tambade S, Shind S V. A review of machine learning techniques using decision tree and support vector machine. *Int Conf Comput Commun Control Autom.* 2016;1–7.
- [59] Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* [Internet]. 2020;408(xxxx):189–215. Available from: <https://doi.org/10.1016/j.neucom.2019.10.118>
- [60] Cardoso-Fernandes J, Teodoro AC, Lima A, Roda-Robles E. Semi-automatization of support vector machines to map lithium (Li) bearing pegmatites. *Remote Sens.* 2020;12(14).
- [61] Auria L, Moro RA. Support Vector Machines (SVM) as a Technique for Solvency Analysis. *DIW Berlin Discuss Pap.* 2008;811.
- [62] Henry J, Natalie T, Madsen D. Pix2Pix GAN for Image-to-Image Translation. 2021;(August):0–5.
- [63] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. *Med Image Anal.* 2019;58.
- [64] RaySearch Laboratories. Deformable Registration in Raystation. 2017;4. Available from: <https://www.raysearchlabs.com/media/whitepapers/>
- [65] <https://pypi.org/project/sewar/>.
- [66] Sara U, Akter M, Uddin MS. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *J Comput Commun.* 2019;07(03):8–18.