

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Biomedica

Tesi di Laurea Magistrale

Confronto di algoritmi di Deep Learning
per l'individuazione del midollo osseo
attivo in immagini CT



Relatrice:

Prof.ssa Samanta Rosati

Corelatrice:

Prof.ssa Gabriella Balestra

Candidato:

Salvatore Tullio Sepe

Anno Accademico 2020/2021

Abstract

Il midollo osseo è composto da un tessuto adiposo, detto midollo osseo giallo (YM), e da un tessuto connettivo chiamato midollo osseo rosso (RM). Quest'ultimo è responsabile della produzione di cellule staminali ematopoietiche. Il RM può essere soggetto a lesioni da tumori maligni, che comportano una diminuzione nella produzione di cellule ematiche e trombociti. Per curare queste patologie vengono prescritte sessioni di radioterapia che possono comportare nel paziente danni al tessuto midollare a causa delle radiazioni. Nasce quindi la necessità di individuare il midollo osseo attivo con precisione per effettuare una radioterapia focalizzata esclusivamente nei punti dove sono presenti le lesioni. Ad oggi, lo stato dell'arte nell'individuazione del RM prevede l'utilizzo di sistemi di imaging come RM, PET e soluzioni ibride tra queste o con CT. La seguente trattazione propone un metodo alternativo per l'individuazione di RM basato su algoritmi di deep learning, da applicare su immagini CT. La CT è infatti una metodologia molto più accessibile e meno costosa rispetto agli altri metodi di imaging precedentemente citati, ed è un esame di routine nel workflow ospedaliero.

Sono state utilizzate immagini CT di 50 pazienti, di cui si avevano anche le maschere RM ottenute tramite PET per effettuare dei confronti. Le immagini dei primi 40 pazienti sono state utilizzate per addestrare le reti, mentre quelle degli ultimi 10 sono state utilizzate come validation set. Da queste immagini, si sono ricavate delle ROI di dimensioni 5x5 che i modelli hanno classificato come RM o YM. Nel lavoro svolto sono stati applicati due algoritmi di deep learning per effettuare la segmentazione: le deep neural network (DNN) e le convolutional neural network (CNN). Le prime sono reti neurali che classificano le ROI analizzando delle features numeriche che sono state estratte in fase di pre-processing. Le seconde, invece, classificano le ROI valutando i valori numerici dei pixel. Per addestrare le reti sono stati creati due training set diversi: il primo è stato costruito selezionando le ROI in maniera randomica; per il secondo è stato utilizzato un algoritmo di clustering (SOM). Successivamente si è svolto un tuning dei parametri delle reti per trovare la configurazione ideale. Dai risultati è emerso che il training set ottenuto tramite clustering porta le reti ad una miglior classificazione delle ROI, grazie al fatto di essere più rappresentativo del dataset. Dopo aver trovato le reti che hanno restituito le migliori prestazioni, si sono ottenute le maschere RM predette sia con DNN che con CNN. Da queste si è svolto un confronto paziente per paziente con le maschere PET utilizzando come indici di valutazione il Dice (DSC), la Precision (Pr) e la Recall (Re). Analizzando i risultati si nota che entrambe le reti tendono a sottosegmentare il midollo osseo, generando valori molto alti di Pr (intorno all'85%), ma molto bassi di Re (40-50%). Inoltre si nota una difficoltà da parte della rete nel segmentare immagini dove vi è una quantità ridotta di midollo osseo, o dove la morfologia midollare è più frastagliata e meno regolare. Infine, si è svolto un confronto tra le due reti, che ha evidenziato una miglior capacità di classificazione e minor tendenza a sottosegmentare della CNN: per molti pazienti si è notato un incremento del DSC e della Re del 10%, mentre la Pr è rimasta relativamente alta per entrambe le reti.

Indice

1. Introduzione	1
1.1 Il midollo osseo	1
1.2 Stato dell'arte nell'individuazione del midollo attivo.....	2
1.2.1 MRI	2
1.2.2 PET	4
1.2.3 PET/CT – PET/MRI	6
1.3 Radiomica	7
1.4 Obiettivo	8
2. Dataset.	9
2.1 Popolazione	9
2.2 Training set e Test set.	11
3. Deep Neural Networks (DNN)	13
3.1 Introduzione alle DNN	13
3.2 Tuning dei parametri	16
3.3 Risultati	19
3.4 DNN con architettura diversa	21
3.5 Risultati con architettura diversa.....	22
3.6 Creazione maschere RM tramite DNN	30
3.7 Risultati segmentazione con DNN.....	31
4 Reti Neurali Convoluzionali (CNN)	50
4.1 Introduzione alle CNN	50
4.2 Tuning dei parametri	53
4.3 Risultati	54
4.4 Creazione maschere RM tramite CNN e risultati della segmentazione.....	60

5	Confronto risultati DNN / CNN	68
5.1	Confronti maschere create con DNN e CNN.....	68
5.2	Confronto dei risultati segmentazione DNN e CNN.....	81
	Conclusione	85
	Bibliografia	86
	Indice Figure	88
	Indice Grafici	90
	Indice Equazioni	91

1. Introduzione

1.1 Il midollo osseo

Il midollo osseo è un tessuto molle che si trova nei canali delle ossa lunghe e nella parte centrale delle ossa piatte. Prima di approfondire il ruolo fondamentale che ha nel processo di ematopoiesi, è opportuno rilevare che questo tessuto si distingue in due tipologie (Figura 1):

- RM (*red marrow*): midollo osseo rosso, anche detto midollo osseo attivo o ematopoietico e formato, appunto, da cellule ematopoietiche. Si presenta in quantità variabile nel corso del tempo: nei primi anni di vita è consistente, ma diminuisce progressivamente, sostituito dal midollo osseo giallo.
- YM (*yellow marrow*): midollo osseo giallo, il cui colore si deve al fatto che è formato principalmente da tessuto adiposo. Al suo interno, gli adipociti conservano i lipidi che verranno poi usati come fonte di energia.

Si è detto che il midollo osseo è fondamentale per l'ematopoiesi: le cellule staminali ematopoietiche prodotte si distinguono in due sotto-famiglie di cellule ematiche, ovvero le cellule linfoidi (cellule T, B e NK) e le cellule mieloidi (monociti, macrofagi, neutrofili, basofili, eosinofili, megacariociti e eritrociti).

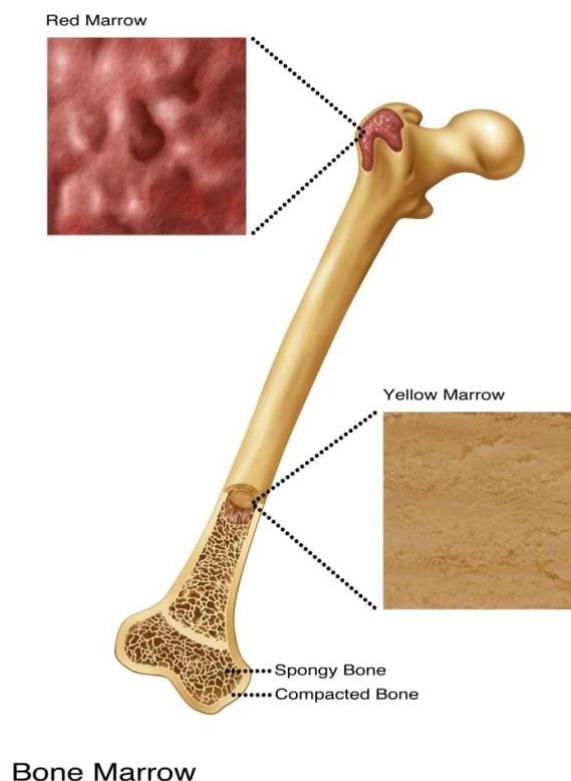


Figura 1. Fonte: Illustration Poster Print by Gwen Shockey/Science Source, Tessuto midollare

Individuare l'area di RM è molto importante per il trattamento clinico di numerose malattie, tra cui il tumore del midollo osseo pelvico (d'altronde, più della metà di RM presente nel corpo umano di un adulto si trova tra il bacino e la spina lombare). I malati oncologici spesso devono sottoporsi a numerose sessioni programmate di radio-chemioterapia. È possibile, tuttavia, che durante questo trattamento il midollo osseo rosso venga compromesso dall'esposizione alle radiazioni, causando una tossicità ematologica. Per esempio, potrebbe manifestarsi la neutropenia (una forma di leucopenia caratterizzata dalla riduzione dei neutrofili sotto il valore di soglia) [1] [2] [3] [4] [5]. Questa evenienza può essere dannosa per il trattamento di radio-chemioterapia, perché può comportare infezioni rilevanti, con conseguente ricovero del paziente; può anche determinare la necessità di spostare (o addirittura interrompere) le sessioni del trattamento, compromettendone inevitabilmente l'efficacia.

Nel tentativo di trovare una soluzione a questo problema, è emersa la proposta di effettuare una radio-chemioterapia di intensità focalizzata, di modo che né l'area di RM né l'efficacia della terapia stessa possano risultare compromesse nel corso del trattamento [6].

1.2 Stato dell'arte nell'individuazione del midollo attivo

Ad oggi vi sono tre principali metodologie applicabili per l'individuazione del midollo osseo attivo, ovvero la biopsia, l'imaging a Risonanza Magnetica (MRI) e la tomografia funzionale ad emissione di positroni (fPET). La prima rappresenta senza dubbio il metodo più accurato, in quanto consiste nell'estrazione di un campione dal midollo osseo che verrà successivamente analizzato in laboratorio attraverso degli strumenti come il microscopio. Con questa analisi è possibile distinguere le tipologie di midollo presenti. Tuttavia questa strada non è percorribile quando è necessario pianificare una radioterapia in quanto non è possibile ottenere informazioni spaziali sui tessuti adiacenti al campione estratto e inoltre si tratta di un metodo dall'alta invasività.

1.2.1 MRI

L'imaging a risonanza magnetica è un ulteriore strumento utilizzabile per l'individuazione di midollo osseo attivo. Per eseguire una MRI si sottopone il paziente ad un forte campo magnetico statico, dell'ordine del tesla. Quando immersi in un campo magnetico statico, gli spin dei protoni contenuti nei tessuti tendono ad allinearsi alle linee di forza del campo magnetico in modo parallelo o antiparallelo. Dal momento che gli spin allineati in modo parallelo sono in maggioranza, si viene a generare una magnetizzazione dei tessuti. In realtà l'allineamento non è mai perfetto, infatti gli spin dei protoni effettuano una precessione, ovvero una rotazione intorno alla direzione del campo magnetico. La frequenza di precessione è detta frequenza di Larmor, la quale negli atomi di idrogeno, che sono quelli più largamente utilizzati come fonte di segnale, per un campo di 1 T vale 42 MHz. Applicando sul paziente un campo magnetico variabile a questa stessa frequenza di risonanza (da qui il nome risonanza magnetica) si può ruotare il vettore magnetizzazione dei protoni di un angolo arbitrario che viene variato in base al tipo di immagine che si vuole ottenere. Questo angolo è detto

flip angle. A seguito dell'impulso gli spin dei protoni tenderanno a riallinearsi al campo magnetico statico. Ciò avviene con due costanti di tempo distinte e indipendenti l'una dall'altra. La prima, chiamata tempo di rilassamento spin-reticolo ed indicata per convenzione con t_1 , descrive il tempo necessario alla ristabilizzazione della magnetizzazione lungo la direzione del campo magnetico statico e dipende dall'interazione tra protoni e le molecole circostanti. La seconda, detta tempo di rilassamento spin-spin è indicata con t_2 , ed esprime l'intervallo di tempo a seguito del quale si viene ad annullare la componente trasversale della magnetizzazione. Misurando dunque tramite una bobina ricevente l'intensità di magnetizzazione nel piano perpendicolare al campo magnetico principale è possibile determinare il tempo di rilassamento e grazie a quest'ultimo è possibile distinguere i vari tessuti, poiché tessuti diversi avranno in generale tempi di rilassamento diversi.

L'imaging a risonanza magnetica dimostra un'alta sensibilità nel distinguere tessuti aventi diverse quantità d'acqua e grassi. Il midollo attivo possiede il 40% di tessuto adiposo mentre nel midollo giallo questa percentuale sale al 95%. Inoltre, il midollo osseo attivo è maggiormente irrorato dai vasi sanguigni e capillari che trasportano ossigeno e nutrienti alle cellule ematopoietiche. Ciò comporta una marcata differenza in termini di presenza di acqua e grassi nei due tessuti midollari. I protoni contenuti nei grassi sono legati a gruppi metilene (CH_2) che causano un tempo di rilassamento spin-reticolo molto bassi e quindi un segnale di magnetizzazione di intensità molto elevata nei tessuti adiposi. Gli stessi protoni presentano invece tempi di rilassamento spin-spin lunghi e quindi un segnale di intensità bassa. Al contrario, i protoni legati a molecole d'acqua presentano tempi t_2 brevi. La conseguenza di ciò è che nelle immagini T1-weighted il tessuto grasso appare brillante mentre le cavità più acquose risultano scure. Nelle immagini T2-weighted accade il contrario. [7,8].

Gli studi condotti nel 2007 da Blebeal et al. [7] e nel 2017 da Andreyvchenko et al. [9] hanno dimostrato la possibilità di una segmentazione automatica del midollo osseo attivo partendo da immagini MRI. Utilizzando le due immagini T1-weighted e T2-weighted si procede nel calcolo del parametro fat fraction, indicato come FF, che indica la percentuale di grassi presente nel tessuto analizzato (equazione 1).

$$FF = \frac{\text{fat image}}{\text{fat image} + \text{water image}} \quad (1)$$

dove fat image è l'immagine T1 e water image è l'immagine T2. Il contenuto di grassi è proprio ciò che permette di discriminare i due tessuti midollari, per tanto è possibile segmentare il midollo osseo attivo implementando una doppia soglia del valore del pixel come segue

$$\text{if } S_{min} < P_{value} < S_{max}, \quad \text{then } P \in RM$$

dove S_{min} e S_{max} rappresentano le soglie inferiore e superiore, P indica il pixel ed RM il midollo attivo. Utilizzando questa tecnica appena descritta è possibile ottenere una maschera del midollo osseo attivo utilizzabile dopo un controllo da parte di un radiologo nella pianificazione della radioterapia per la riduzione dell'emotossicità della terapia stessa.

L'imaging a risonanza magnetica presenta quindi un'elevata sensibilità e risoluzione spaziale, oltre ad avere l'importante vantaggio di utilizzare radiazioni elettromagnetiche non ionizzanti che

non comportano effetti dannosi al paziente. Questa metodologia tuttavia è in grado di riportare solo informazioni di tipo strutturali ma non fornisce informazioni di tipo funzionale sul livello di attività del midollo osseo attivo. Infatti, è possibile che vi siano dei tessuti con lo stesso livello di fat fraction del tessuto in questione ma che non sono responsabili dell'attività ematopoietica e che potrebbero quindi essere erroneamente coinvolti nella radioterapia.

1.2.2 fPET

La tomografia ad emissione di positroni è una tecnica di imaging della medicina nucleare. A differenza delle altre tipologie di imaging che restituiscono informazioni di tipo morfologico, la PET è capace di fornire informazioni di tipo fisiologico in grado di mappare i processi funzionali e metabolici nel paziente. Il procedimento che porta all'ottenimento della bioimmagine inizia con l'iniezione di un radiofarmaco, composto da un radio-isotopo ad emivita breve legato ad una molecola attiva metabolicamente nel processo che si vuole analizzare. Il radiofarmaco raggiunge quindi il sito anatomico interessato ed inizia a decadere, emettendo positroni. Questi ultimi vengono catturati con l'ausilio di una gamma camera ed andranno a formare l'immagine.

Vi sono diversi tipi di meccanismi di captazione utilizzabili per ottenere un'immagine del midollo osseo attivo, ciascuno dei quali utilizza radiofarmaci diversi. Alcuni esempi sono [10] :

- Captazione delle cellule granulopoietiche: basate sul labeling di anticorpi e cellule facenti parte del sistema immunitario. I radiofarmaci utilizzabili a questo scopo sono molto complessi da produrre e sono altamente captabili da parte di reni e fegato, cosa che comporta una riduzione di quantità di radiofarmaco che raggiunge il sito desiderato.
- Captazione delle cellule eritropoietiche. I radiofarmaci non vengono captati da reni o fegato, ma il costo è molto elevato. Un esempio è il citrato di Fe-52.
- Captazione da parte del reticolo endoteliale (RES) del midollo osseo attivo. Rimane il problema dell'assorbimento di radiofarmaco da parte di reni e fegato. Un esempio è il SC (colloide zolfo-Tc-99m).

Il radiofarmaco più largamente utilizzato è il radioisotopo 2-fluoro-2-deossi-glucosio (FDG) [11] che viene assorbito dalle cellule ematopoietiche del midollo attivo molto più di quanto non accada per il midollo giallo. Infatti la maggior parte dei processi metabolici midollari avviene proprio nel midollo attivo, ed il farmaco in questione, essendo un analogo del glucosio, viene captato ma non catabolizzato nei processi metabolici e rimane quindi nella sua forma fino al decadimento del fosforo-18, che è il radioisotopo responsabile dell'emissione dei positroni. L'utilizzo dell'FDG permette quindi l'individuazione del tessuto midollare in base al grado di attività metabolica e non in base alle caratteristiche strutturali del tessuto stesso.

Sono stati effettuati diversi studi [12] che hanno mostrato la possibilità di una segmentazione automatizzata e semi-automatizzata del midollo osseo attivo a partire da bioimmagini FDG-PET. Uno di questi prevede il calcolo del parametro Standardized Uptake Value (SUV) che è definito come la

concentrazione di attività divisa per il peso del paziente e la quantità di attività somministrata al paziente. Alcune varianti a questo parametro sono:

- SUV_{mean} che rappresenta la media del valore di SUV in un volume specificato
- SUV_{max} , il massimo del valore di SUV in un volume analizzato
- SUV_{peak} che è calcolato come la media del valore di SUV in un volume sferico di 1.2 cm di diametro.
- SUV_{LBM} equivalente al valore di concentrazione di attività normalizzata per il valore di massa magra del paziente.

Utilizzando la SUV è possibile effettuare la segmentazione del midollo osseo attivo andando ad implementare una semplice operazione di thresholding, ovvero viene classificato come midollo attivo quella porzione di tessuto che presenta un valore di SUV maggiore della soglia impostata [13].

La PET, tuttavia, è una tecnica di imaging che presenta diverse criticità. La prima riguarda le caratteristiche intrinseche dell'immagine generata. Infatti la PET viene eseguita solitamente su tutto il corpo del paziente, per cui ne risulta un'immagine che restituisce informazioni sul consumo di glucosio di tutti i tessuti. Questo porta alla presenza di un elevato rumore di fondo, che può inevitabilmente pregiudicare la possibilità di una segmentazione automatica e l'interpretazione dei risultati. E' inoltre doveroso ricordare che nelle immagini ottenute tramite PET non vi sono informazioni di tipo morfologico e strutturale. Un'ulteriore problematica è data dalla presenza di effetti di assorbimento e scattering del segnale, che si traducono in una bassa risoluzione spaziale la quale, unitamente all'imperfezione degli algoritmi di ricostruzione, portano ad una quantificazione del radiofarmaco non ottimale. Ciò comporta un peggioramento della precisione dell'immagine ed un basso grado di standardizzazione dei risultati quando vengono adoperati macchinari ed algoritmi differenti. Il Partial Volume Effect è un esempio di effetto negativo. Questo effetto è legato alla bassa risoluzione spaziale del sistema di imaging e causa la perdita di informazione di porzioni di tessuto che non sono almeno 2-3 volte più grandi della risoluzione del macchinario. Il PVE può essere parzialmente corretto adoperando degli algoritmi di ricostruzione che tentano di recuperare l'informazione perduta, ma mai completamente eliminato [12].

Un altro aspetto negativo della PET è che sono richiesti radiofarmaci a bassa emivita (vanno dunque utilizzati immediatamente dopo la sintesi) e che sono fabbricabili solo attraverso l'utilizzo di ciclotroni. E' dunque necessario che il macchinario per la PET sia locato nelle vicinanze dei ciclotroni; questo aspetto ha ridotto enormemente la diffusione di questa tecnologia a livello territoriale dal momento che non tutti gli ospedali possono investire la quantità di denaro necessaria all'acquisto di ciclotroni. L'esame PET è infatti molto costoso rispetto agli altri sistemi di imaging. Inoltre si tratta di una procedura invasiva per il paziente con possibili effetti collaterali.

1.2.3 PET/CT – PET/MRI

Come è stato detto, nelle immagini PET vi è la mancanza di informazioni di tipo strutturale-morfologico, mentre nelle immagini MRI a mancare sono le informazioni funzionali e metaboliche. Alcune tecniche sono state ideate per sopperire alle mancanze dei singoli metodi di imaging; queste prevedono l'aggiunta all'esame tramite PET un'immagine strutturale come la MRI o la tomografia computerizzata (CT). Grazie a questa tecnologia ibrida è possibile ottenere dati più precisi e dettagliati che favoriscono una segmentazione automatica del midollo osseo attivo ottimale.

Il quantitative total marrow imaging (QTMI) sviluppato da Perk et al. [14] è un esempio di applicazione. Questa procedura consiste nell'utilizzo di immagini PET/CT per quantificare in maniera automatizzata l'attività midollare. Per l'esecuzione è prevista l'acquisizione in parallelo di una immagine PET e di una CT. Quest'ultima viene utilizzata per segmentare il tessuto osseo del paziente sfruttando un algoritmo di tipo atlas-based, ovvero un algoritmo che sfrutta delle maschere pre-registrate di segmenti ossei da sovrapporre all'immagine del paziente che vengono deformate fino alla corretta individuazione di tutte le regioni scheletriche. Nell'altra immagine ottenuta, la PET, viene calcolato il parametro SUV descritto precedentemente. Successivamente viene effettuata una moltiplicazione tra l'immagine PET e la maschera ottenuta tramite CT in modo tale da evidenziare sulla PET le sole regioni di tessuto osseo. Infine viene applicato il thresholding sul parametro SUV allo scopo di lasciare sull'immagine il solo midollo osseo attivo.

Nella tecnologia ibrida PET/CT, l'utilizzo della CT consente quindi di effettuare una precisa segmentazione delle ossa del corpo. Essa può inoltre offrire una maggiore risoluzione spaziale, la quale è una delle principali limitazioni della PET, oltre a poter essere utilizzata per attenuare le problematiche indotte dall'effetto di scattering fotonico. Dall'immagine CT, infatti, vengono estratti una serie di parametri correttivi, che consentono di correggere parzialmente gli effetti di scattering. Le correzioni di queste degradazioni sono generalmente già implementate nei software delle macchine: le immagini CT, infatti, forniscono una mappa dei coefficienti di attenuazione dei tessuti, direttamente legati ai fenomeni di scattering.

Se da un lato si riesce ad avere una precisa segmentazione del midollo osseo, i macchinari PET/CT sono macchinari molto costosi, aventi tutti gli stessi problemi delle sole PET, e necessitano di un importante investimento da parte degli ospedali, a cui si aggiunge un grande aumento delle dosi di radiazioni somministrate ai pazienti: sia CT che PET separatamente hanno un alto livello di radiazioni, e la tecnologia ibrida aumenta ulteriormente tale dose, diventando uno dei metodi d'indagine con più alto livello di radiazioni.

Oltre a questa tecnica appena descritta, è stata sviluppata un'ulteriore procedura che consiste nell'utilizzo combinato di immagini PET ed MRI. Questa tecnica combina alla capacità di individuazione di attività metabolica da parte della PET la caratterizzazione dei tessuti tipica dell'MRI [11]. Gli step da effettuare per ottenere la segmentazione del midollo sono i medesimi descritti per la PET/CT: viene acquisita l'immagine MRI che evidenzia e segmenta i tessuti midollari interessati, con la maschera così ottenuta si effettua una moltiplicazione con l'immagine PET e si ha in uscita un'immagine che mappa l'attività metabolica del solo midollo osseo attivo. E' doveroso sottolineare che la tecnologia PET/MRI sottopone il paziente ad una dose di radiazioni nettamente

inferiore rispetto alla PET/CT e che le informazioni che restituisce sono di migliore qualità e con una soglia di imprecisione più bassa. Tuttavia, il costo estremamente elevato di questa tecnica ne ha limitato l'utilizzo per quanto riguarda l'individuazione di midollo attivo per pianificazione di radioterapia. La metodologia più conveniente ed utilizzata rimane quella della sola MRI.

1.3 Radiomica

La radiomica è un approccio innovativo all'imaging medico, che mira a migliorare i dati esistenti a disposizione dei medici mediante analisi matematiche avanzate. Attraverso l'estrazione matematica della distribuzione spaziale delle intensità del segnale e delle interrelazioni tra i pixel, la radiomica quantifica le informazioni della texture utilizzando metodi di analisi dal campo dell'intelligenza artificiale [15]. Il concetto alla base del processo è che le immagini cliniche sia morfologiche che funzionali contengono informazioni qualitative e quantitative, che possono riflettere la fisiopatologia sottostante di un tessuto. Nascosti tra i valori numerici dei pixel, infatti, vi sono delle proprietà che non sono visibili attraverso una semplice analisi visiva da parte dei radiologi e che possono diventare strumento per lo sviluppo di supporto decisionale per i medici.

Diversi tipi di features possono essere derivati da immagini cliniche. Le features semantiche qualitative sono comunemente usate nel lessico radiologico per descrivere le lesioni [16]. Le caratteristiche quantitative sono descrittori estratti dalle immagini mediante software che implementano algoritmi matematici [17]. Queste presentano diversi livelli di complessità ed esprimono proprietà in primo luogo della forma della lesione e dell'istogramma dell'intensità del pixel, secondariamente della disposizione spaziale dei valori di intensità a livello del pixel. Possono essere estratti direttamente dalle immagini o dopo aver applicato diversi filtri o trasformazioni. Le features quantitative sono generalmente suddivise nei seguenti sottogruppi:

- *Features di forma*: descrivono la forma della regione di interesse tracciata (ROI) e le sue proprietà geometriche come volume, diametro massimo lungo diverse direzioni ortogonali, superficie massima, compattezza del tumore e sfericità. Ad esempio, il rapporto superficie-volume di un tumore spiccolato mostrerà valori più alti di quello di un tumore rotondo di volume simile.
- *Features statistiche del primo ordine*: descrivono la distribuzione dei singoli valori pixel senza preoccuparsi delle relazioni spaziali. Si tratta di proprietà basate su istogrammi che riportano i valori medi, mediani, massimi e minimi delle intensità dei pixel sull'immagine, nonché la loro skewness (asimmetria), kurtosis (piattezza), uniformità e casualità (entropia).
- *Features statistiche del secondo ordine*: includono le cosiddette textural features, che si ottengono calcolando le interrelazioni statistiche tra pixel vicini [18]. Forniscono una misura della disposizione spaziale delle intensità dei pixel e quindi dell'eterogeneità intra-tissutale.
- *Features statistiche di ordine superiore*: sono ottenute con metodi statistici dopo l'applicazione di filtri o trasformazioni matematiche alle immagini; ad esempio, con l'obiettivo di identificare schemi ripetitivi o non ripetitivi, sopprimere il rumore o evidenziare i dettagli. Questi includono l'analisi frattale, la trasformazione wavelet e le trasformazioni laplaciane di immagini filtrate gaussiane, che possono estrarre aree con modelli di texture sempre più grossolani.

Considerando che molti parametri possono essere regolati dall'utente, centinaia di variabili possono essere generate da una singola immagine. Inoltre, l'analisi radiomica può essere eseguita su tutte le tipologie di immagini mediche, consentendo un approccio multimodale integrato utilizzando il potenziale valore additivo delle informazioni di imaging estratte, ad esempio, dalla MRI, dalla TC o dalla PET.

La possibilità di estrarre informazioni nascoste da immagini mediche digitali è di particolare interesse in quanto può potenziare le capacità predittive degli algoritmi di segmentazione automatici esistenti. L'estrazione di nuove informazioni, precedentemente nascoste, può essere sfruttata per la segmentazione automatica delle immagini.

1.4 Obiettivo

A causa delle sue caratteristiche, il midollo osseo attivo non è distinguibile dal midollo osseo giallo con una semplice analisi visiva di un'immagine CT. D'altro canto, utilizzando un approccio di radiomica sull'immagine CT è possibile ottenere informazioni che possono portare alla segmentazione del midollo osseo attivo senza ricorrere alla PET o alla MRI. Questo rappresenta un enorme vantaggio, in quanto tutti i pazienti con patologie oncologiche sono sottoposti alla CT, trattandosi di un esame standard nel workflow ospedaliero e nella pianificazione della radioterapia. Essere in grado di utilizzare queste stesse immagini, che verosimilmente sono state già acquisite per altri scopi, porterebbe ad un notevole risparmio in termini economici (essendo la CT molto meno costosa e più accessibile del MRI e della PET) ed in termini di tempo, oltre a causare una semplificazione della procedura di pianificazione della radioterapia rispetto ai metodi attualmente utilizzati, ovvero le tecniche di imaging precedentemente citate.

La seguente trattazione vuole approcciare al problema dell'individuazione del midollo osseo attivo attraverso metodologie di radiomica, e punta a valutare le performance di una segmentazione automatizzata a partire da immagini CT basata su algoritmi di deep learning. Per la sperimentazione sono state utilizzate immagini CT acquisite da 50 pazienti. Verranno analizzati i risultati con l'utilizzo di due tipologie di reti neurali differenti ed al variare di diversi parametri. La prima rete è la deep neural network, che classifica le ROI ricavate dalle immagini effettuando un'analisi delle features numeriche estratte dalle ROI stesse, mentre la seconda, la convolutional neural network, effettua una predizione sull'appartenenza o meno delle ROI al tessuto midollare attivo sfruttando i valori numerici dei pixel delle ROI. Una volta effettuata una valutazione delle performance al variare di diversi parametri, sono state scelte delle reti, attraverso le quali si sono ottenute le maschere del midollo osseo attivo. Queste ultime sono state confrontate con le maschere ottenute tramite PET ed attraverso alcuni indici, che verranno illustrati in seguito, si è valutata la bontà della segmentazione adoperata tramite algoritmi di deep learning.

2. Dataset

2.1 Popolazione

Le immagini sottoposte ad analisi per questa trattazione sono state acquisite da una popolazione di 50 pazienti, affetti da carcinomi a cellule transizionali e/o carcinomi a cellule squamose in cura presso il dipartimento oncologico dell'ospedale Molinette di Torino. Questi sono stati sottoposti sia ad una CT che ad una FDG-PET. La prima è stata eseguita su tutti i pazienti, in posizione supina e con dei supporti per ginocchia e caviglia, utilizzando uno scanner CT Philips "BigBore", che presenta un pixel spacing di 0.93 mm e una distanza tra le varie slice di 3 mm. L'esame FDG-PET è stato condotto adoperando una Philips Gemini PET/CT Tomography.

Sulle immagini PET è stata effettuata una segmentazione semi-automatica del RM utilizzando l'indice Standardized Uptake Value (SUV) descritto nel capitolo precedente. Nello specifico, è stato calcolato il suddetto valore ed è stato normalizzato con il valore SUV trovato nelle pelvi e nel fegato di ogni paziente. Successivamente è stato applicato un thresholding, impostando come soglia il valore SUV_{mean} del paziente, e sono stati classificati come appartenenti al midollo attivo tutti quei pixel che hanno presentato un valore di SUV maggiore della soglia.

Per le immagini CT, invece, è stata adoperata una segmentazione manuale del bacino da parte di un radioterapista oncologico, che ha separato le immagini in tre regioni anatomiche:

1. Midollo osseo lombosacrale (LSBM): comprende l'area tra il bordo somatico superiore della quinta vertebra lombare (L5) fino alla parte inferiore del coccige;
2. Porzione bassa del midollo osseo pelvico (LPBM): include il pube bilaterale, l'ischio, l'acetabolo e il femore prossimale;
3. Midollo osseo iliaco (IBM): ne fanno parte le creste iliache ed il bordo superiore.

Delle sopra citate tre regioni, è stata considerata la sola regione di midollo osseo lombosacrale LSBM. Le immagini CT che compongono il dataset sono state sottoposte ad una fase di pre-processing, nel quale è stata adoperata la rimozione del tavolino medico e dei tessuti non ossei. In aggiunta, è stata effettuata la rimozione dell'osso compatto in quanto, come noto dall'anatomia, in esso non vi è contenuto il midollo osseo che si vuole identificare. Grazie a questo processo è stato possibile ottenere le maschere binarie contenenti midollo osseo (BM), oltre che a quelle contenenti il solo midollo attivo (RM). Dalla sottrazione delle due si è risaliti alle maschere che identificano la porzione d'immagine contenente midollo giallo (YM). E' possibile osservare il risultato di tale operazione nelle seguenti figure (figure 2, 3, 4, 5).

Slice originale

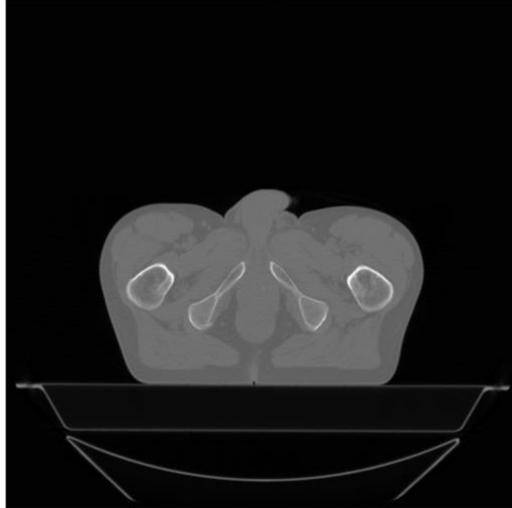


Figura 2, Slice originale

Mask BM BONE



Figura 3,
Maschera BM
con osso
compatto

Mask RM BONE

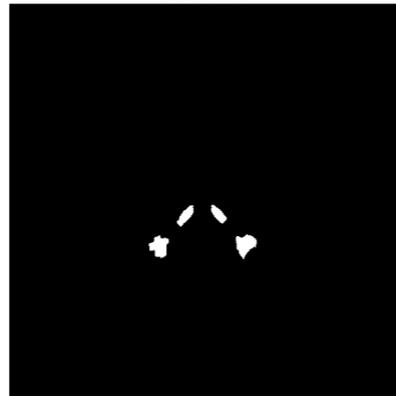


Figura 4,
Maschera RM
con osso
compatto

Mask BM no Bone



Figura 5, Maschera
BM senza osso
coompatto

Mask RM no Bone



Figura 6, Maschera RM
senza osso compatto

2.2 Training set e Test set

Le reti neurali che sono state utilizzate per questo studio, come già accennato, sono state progettate per predire l'appartenenza o meno al midollo attivo di ROI di dimensione 5 x 5. Per addestrare una rete neurale e valutarne le performance sono necessari un training set ed un test set. Il primo contiene tutte quelle ROI che vengono utilizzate in fase di addestramento, mentre il secondo contiene ROI che non sono state sottoposte in fase di addestramento e i risultati che la rete restituisce su queste indicano la capacità di generalizzare della rete stessa. La fase di addestramento e la valutazione delle performance verranno illustrate nel dettaglio nei capitoli seguenti.

La prima rete, la deep neural network - DNN, accetta in ingresso delle features numeriche. Pertanto, sono state individuate, in tutte le ROI poste sotto analisi, 36 features:

- 4 features statistiche del primo ordine: media, deviazione standard, skewness e kurtosis dei pixel della regione
- 32 features statistiche del secondo ordine: di questi, 22 sono state estrapolate dalla matrice di co-occorrenza (GLCM), 5 sono state calcolate dal Grey Level Difference Method (GLDM) e le rimanenti 5 attraverso il Grey Level Run Length Method (GLRLM)

Il risultato di questa operazione sono matrici Nx41, dove N indica il numero di ROI, mentre le 41 colonne indicano, in ordine, l'id del paziente dalla cui immagine è stata prelevata la ROI, il numero di slice, le coordinate x e y del pixel centrale della ROI, seguiti dalle 36 features prima descritte e, nell'ultima colonna, un valore che indica l'appartenenza della ROI al RM o al YM (0 = appartenenza al YM, 1 = appartenenza al RM). Di seguito, un esempio delle matrici così ottenute.

ID PAZ.	N. SLICE	X CENTR	Y CENTR	FEATURES				CLASSE
1	102	273	281	133	0.30	48	...	1
1	105	201	288	138	0.42	32	...	1
2	62.5	245	274	145	0.74	55	...	0
3	78	331	199	158	0.87	42	...	1
3	81	278	275	199	0.45	98	...	0
4	65.5	240	288	161	0.36	74	...	1
4	65.5	233	290	155	0.91	45	...	1
4	68.5	212	234	143	0.87	58	...	1
5	71.5	212	235	128	0.09	72	...	0

Tabella 1, esempio di matrice

Successivamente, le features sono state tutte normalizzate utilizzando la formula del min-max scaling (equazione 2), trovando il minimo ed il massimo per ciascuna delle 36 features:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Dove x' è la feature normalizzata. Per la seconda rete, la convolutional neural network, invece, è stato necessario isolare e salvare le ROI stesse. Queste sono state separate in due cartelle diverse in base all'appartenenza o meno al RM, in quanto il software (MATLAB R2020b) utilizza la cartella di appartenenza per etichettare il dataset. Successivamente, i valori dei pixel di tutte le ROI sono stati normalizzati anch'essi con il metodo min-max scaling ma, in questo caso, è stato trovato un valore di minimo e di massimo per ogni paziente su tutto il volume, ovvero considerando tutte le slice appartenenti al paziente.

Dei 50 pazienti che compongono il dataset, le ROI scelte per il construction set sono state prelevate dalle immagini acquisite dai primi 40 pazienti, mentre i rimanenti 10 pazienti sono stati scelti per la validazione. Nello studio effettuato sono stati adoperati due dataset diversi (quindi, due training set e due test set).

Il primo è stato realizzato andando a selezionare le ROI dalle immagini mediche in maniera random. Nello specifico, per tutti i 40 pazienti, si sono selezionate 5 slice dove vi era la presenza di almeno 20 ROI appartenenti al RM e di altrettante appartenenti al YM. Si sono ottenute quindi 200 slice e, da queste, sono state estratte 5000 ROI che appartengono al RM e 5000 che appartengono al YM in maniera random, in modo tale da avere un training set di 10000 elementi bilanciato e che rappresentasse dunque le due classi in egual modo. Tutte le altre ROI facenti parte delle 200 slice sopra menzionate, che non sono state selezionate per costruire il training set, sono state utilizzate per formare il test set.

Per il secondo, invece, è stato adoperato un algoritmo di clustering che ha permesso di comporre il training set con ROI presentanti caratteristiche diverse tra loro. Per ottenere questo risultato si è adoperato l'algoritmo delle Self-Organizing Map (SOM), il quale si basa su un principio simile a quello delle reti neurali. Nelle SOM, i neuroni sono disposti su di una mappa solitamente 2D. Quando viene proposto alle rete SOM un input, i vari neuroni che compongono la rete entrano in "competizione" tra loro, e ne risulta vincitore quel neurone che presenta le caratteristiche più adatte per riconoscere quello specifico input. In quel momento avviene la fase di apprendimento, ovvero vi è una modifica dei pesi del neurone stesso che nel caso delle reti SOM coinvolge anche il vicinato del neurone vincitore (la fase di apprendimento delle reti verrà illustrata nel dettaglio nel capitolo seguente). Nell'applicazione effettuata in questo caso si è scelta, a seguito di diverse prove, una rete SOM di dimensione 12 e con vicinato 1. Anche in questo caso, sono state selezionate dai 40 pazienti 5 slice con le stesse condizioni adoperate per l'ottenimento del primo training set. Le ROI appartenenti a queste slice sono state quindi sottoposte alla rete SOM, che attraverso il meccanismo di competizione ha separato le ROI in 12 cluster. Successivamente, si è preso un numero di ROI da ogni cluster, proporzionato alla numerosità del cluster stesso, per formare il training set. Una volta determinato quante ROI estrarre da ogni cluster, queste sono state prese in maniera random, fino ad ottenere 10000 ROI, di cui 5000 rappresentanti il YM e 5000 raffiguranti il RM. Le restanti sono state adoperate per la composizione del test set.

3. Deep Neural Network (DNN)

3.1 Introduzione alle DNN

Le deep neural network (reti neurali profonde - DNN) sono una delle tipologie di modelli utilizzati nell'ambito del deep learning [19]. Quest'ultimo rappresenta una branca del più ampio machine learning, il quale racchiude tutti quegli algoritmi in cui vi è una fase di apprendimento, ovvero una modifica iterativa di alcuni parametri (che verranno esplicitati successivamente) con lo scopo di ridurre l'errore tra la verità e la predizione che la rete genera da dei dati di addestramento. Nello specifico, le reti neurali sono modelli matematici ispirati al complesso tessuto neurale biologico formati da nodi (o neuroni) e connessioni tra loro. Negli ultimi anni, le Neural Network - NN hanno trovato ampio utilizzo in diversi settori ingegneristici e non, tra cui la biomedica[20] [21] [22] [23] [24] [25]. Le applicazioni includono problemi di ottimizzazione, simulazione di processi, classificazione (quella adoperata in questo elaborato). I neuroni sono disposti in strati (layers), e tutti i neuroni di uno strato sono collegati a tutti i neuroni appartenenti allo strato precedente e allo strato successivo, senza però essere collegati con quelli dello stesso strato. Equivalentemente al potenziale d'azione presente nei neuroni biologici, i nodi delle NN possono attivarsi se ricevono un input adeguato. Ogni neurone è una funzione matematica, ed è caratterizzato da una funzione di attivazione, un bias ed un peso nelle connessioni con gli altri neuroni ad esso collegati. Ciò è osservabile nella seguente equazione (equazione 3):

$$a_j^i = \sigma(\sum_k w_{jk}^i a_k^{i-1} + b_j^i) \tag{3}$$

a_j^i rappresenta l'attivazione del j-esimo neurone presente nel i-esimo layer, σ è una funzione di attivazione, w_{jk}^i definisce il peso nella connessione tra il neurone stesso ed il k-esimo neurone nel layer precedente e b_j^i indica il bias associato al neurone. La struttura del singolo neurone, secondo il modello di McCulloch e Pits, è rappresentata nella seguente figura (figura 7).

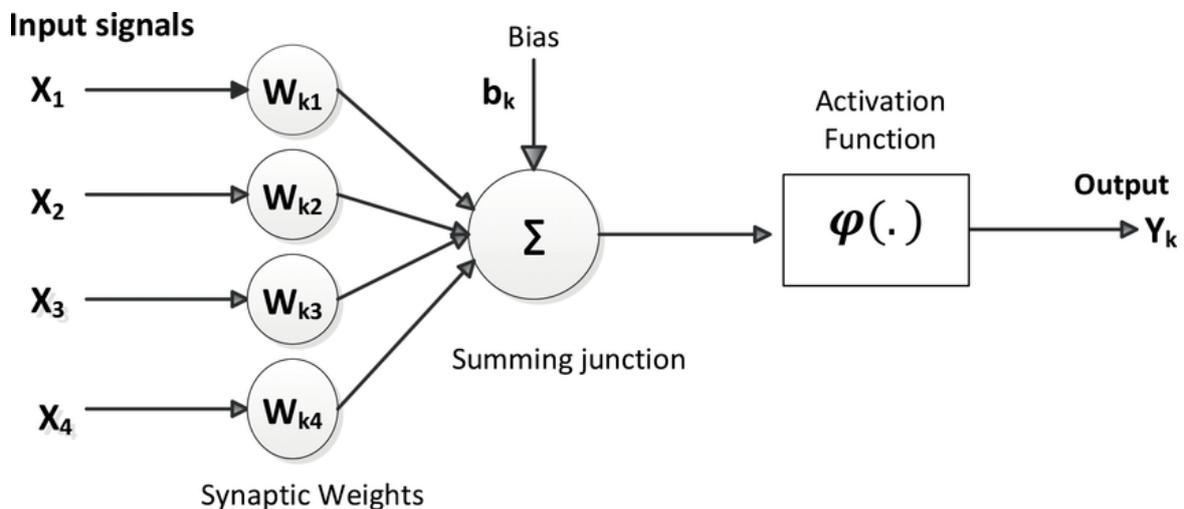


Figura 7, Modello di Neurone di McCulloch e Pits

Come già accennato, una NN è composta da un certo numero di neuroni disposti su strati paralleli. L'input che ricevono i neuroni nel primo layer, detto appunto input layer, è proprio l'input numerico che si vuole processare (nel caso in esame, sono le features numeriche estratte dalle ROI, descritte nel capitolo precedente). L'output che questi neuroni generano viene trasmesso come input ai neuroni dello strato successivo (layers nascosti), fino ad arrivare all'output layer, dove l'output che è generato dai neuroni presenti in questo layer è l'output finale restituito dalla rete.

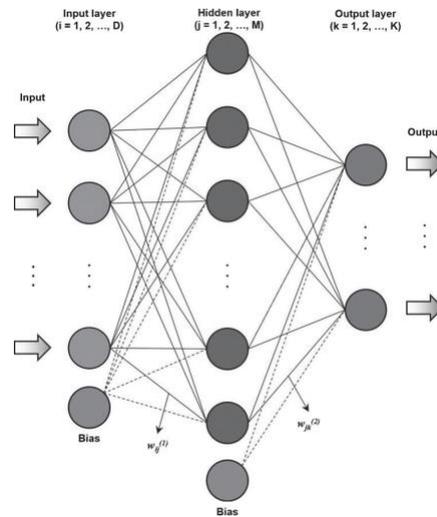


Figura 8, Esempio di struttura NN

Una rete è considerata profonda (DNN) quando presenta nella sua architettura un elevato numero di hidden layers.

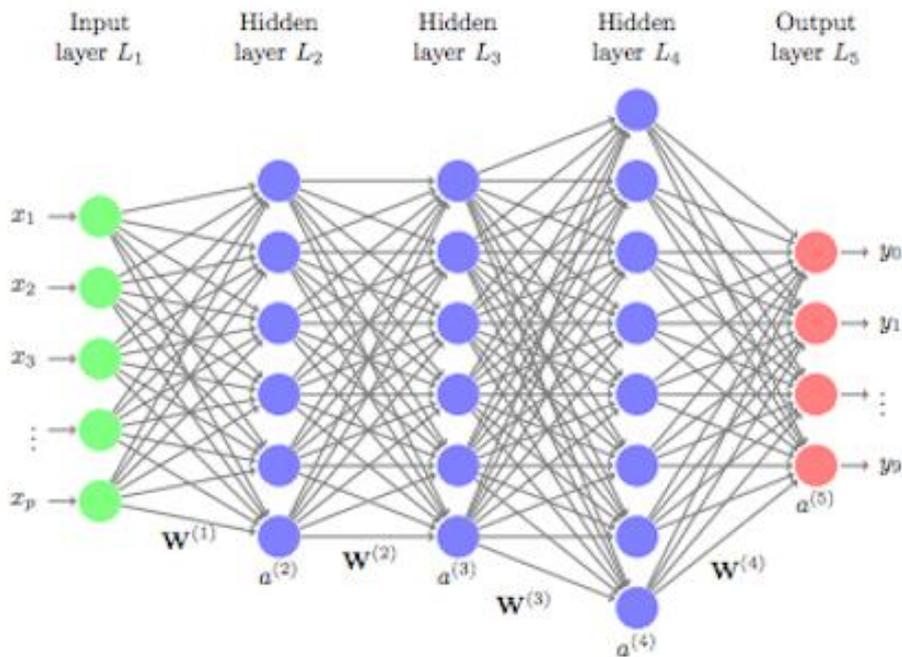


Figura 9, Esempio di struttura DNN

Si è detto che le NN subiscono un processo di training dove vengono stabiliti i valori ottimali di peso e bias. Vi sono diversi tipi di addestramento: quello utilizzato in questa trattazione è l'apprendimento supervisionato. Questo si basa sul fornire alle reti gli input del training set insieme all'output desiderato, ovvero la classe di appartenenza reale (YM o RM, in questo caso). All'inizio, vengono dati ai neuroni e alle connessioni interneuroni valori di peso e bias in maniera random. Successivamente, durante la fase di addestramento, viene calcolato l'errore da una funzione di costo (loss function), e la rete modifica iterativamente i suoi parametri per minimizzare il costo generato dalla funzione. Una delle loss function utilizzabili è quella della cross-entropy (equazione 4):

$$C = -\frac{1}{n} \sum_x [y \ln(a) + (1 - y) \ln(1 - a)] \quad (4)$$

Dove a è l'output generato dal neurone, n è il numero totale di elementi facenti parte del training set ed y è l'output desiderato. Il processo di addestramento può essere ripetuto un numero prestabilito di volte, oppure può essere terminato in anticipo una volta che la rete raggiunge stabilmente un livello di errore prestabilito per un certo numero di epoche, ovvero arriva a convergenza. Per i neuroni vi sono diverse tipologie di funzioni di attivazione che possono essere applicate.

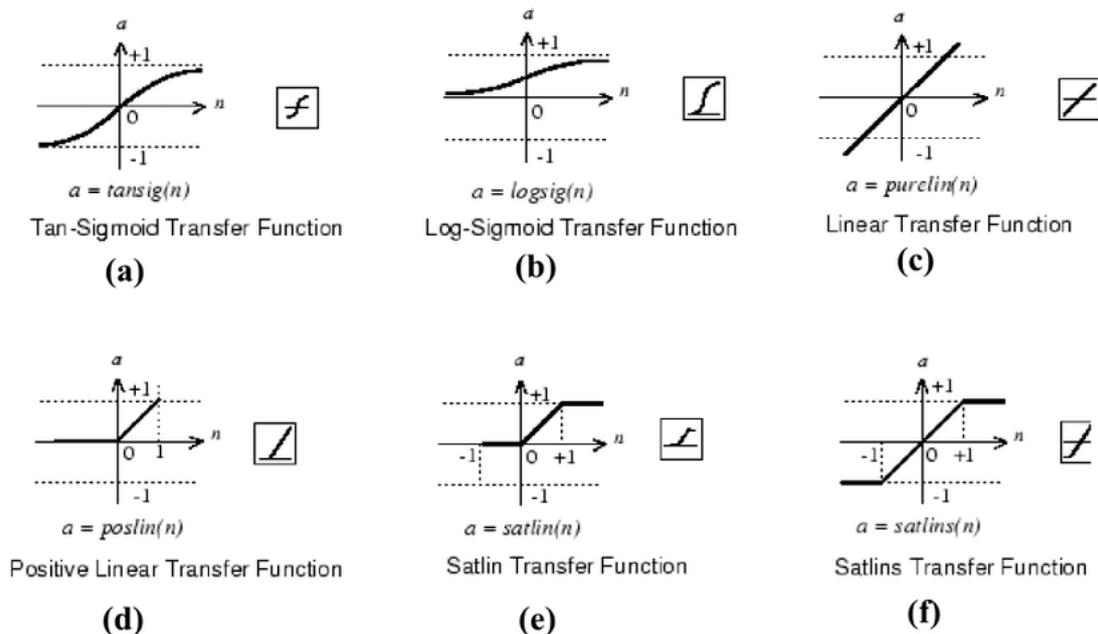


Figura 10, Possibili funzioni di attivazione dei neuroni

Come è evidente, quando si progetta una rete neurale profonda è necessario considerare un elevato numero di parametri, come architettura, funzioni di attivazione, funzioni di costo, tipologie di algoritmi di addestramento, e altro. A causa di ciò, per trovare l'assetto ideale che la rete deve assumere per risolvere il problema, è necessario effettuare un tuning dei parametri, ovvero si addestra la rete un elevato numero di volte (a causa della randomicità con cui vengono preliminarmente assegnati i valori di peso e bias), andando a cambiare di volta in volta un parametro, e valutandone l'effetto. Questo processo è stato effettuato, ed è descritto nel paragrafo seguente.

3.2 Tuning dei parametri

L'architettura scelta per la DNN presenta una struttura triangolare, con 36 neuroni (come il numero di features processate) nel layer di input, un neurone in meno per ogni layer nascosto successivo (34 in totale) fino ad arrivare al layer di output con un unico neurone. Ogni neurone ha come funzione di attivazione la sigmoide, e come loss function è stata adoperata la cross-entropy. Nella figura seguente è possibile osservare tale struttura.

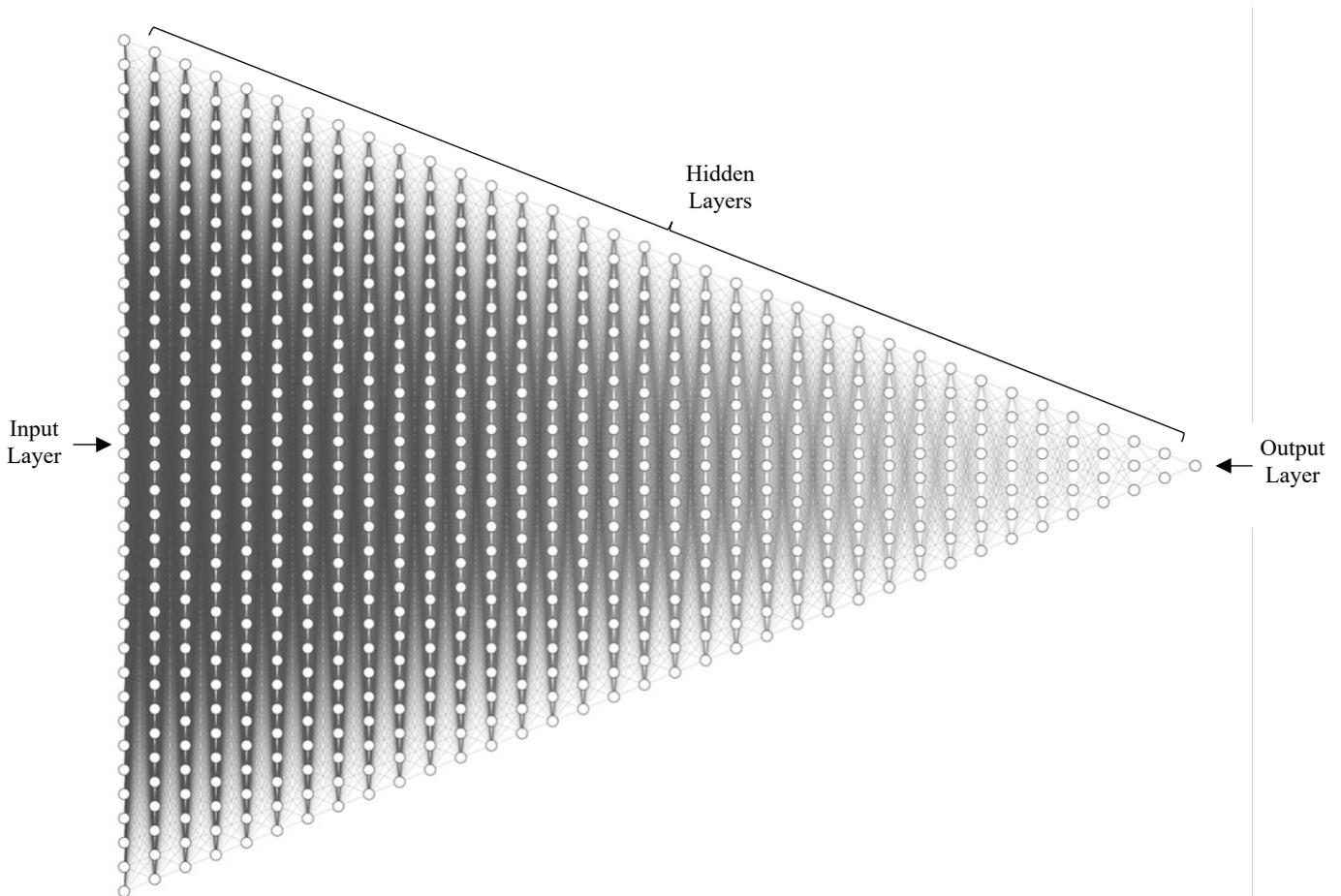


Figura 11, Struttura Triangolare DNN

Per trovare l'assetto ideale dei vari settaggi della deep neural network è stato effettuato un tuning dei parametri, ovvero si è ripetuto il training della rete e si sono valutate le sue performance diverse volte, cambiando alcuni parametri. Nello specifico, si è fatta variare la funzione di attivazione del layer di output tra la sigmoide e la lineare pura (nel software utilizzato, MATLAB_R2020b, sono nominate rispettivamente 'tansig' e 'purelin'). Inoltre, si sono valutate le performance con i due diversi training set descritti nel capitolo precedente. Infine si sono valutate due opzioni per la binarizzazione dell'output: per la prima, si è utilizzata una soglia fissa a 0.5, mentre per la seconda è stata adoperata una soglia variabile ad ogni addestramento, calcolata tramite curva ROC. Tutti

le ROI che hanno restituito un output minore della soglia, sono state classificate come YM (valore 0), le altre, con valori maggiori o uguali, sono state classificate come RM (valore 1).

Per valutare le performance sul training set e sul test set, sono stati utilizzati quattro indici ricavabili dalle confusion matrix. Queste sono delle matrici i cui elementi indicano i risultati della classificazione, come è possibile vedere nella seguente figura d'esempio (figura 7):

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Figura 12, Struttura confusion matrix

Dove “true positive” indica i veri positivi, ovvero il numero di quelle ROI che sono state classificate come RM e che realmente ne fanno parte. Il valore “true negative” ha lo stesso significato, ma per le ROI YM. I “false negative” sono le ROI che sono state classificate dalla rete come YM ma che in realtà raffigurano il RM. L’ultimo dato, “false positive”, rappresenta ovviamente il caso opposto a quest’ultimo. I quattro indici sopra citati sono così calcolabili (equazioni 5, 6, 7, 8):

$$accuracy = \frac{true\ positives + true\ negatives}{true\ positives + true\ negatives + false\ positives + false\ negatives} \quad (5)$$

$$specificity = \frac{true\ negatives}{true\ negatives + false\ positives} \quad (6)$$

$$sensitivity = \frac{true\ positives}{true\ positives + false\ negatives} \quad (7)$$

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \quad (8)$$

E' interessante comprendere il funzionamento della curva ROC. La curva ROC (Receiver Operator Characteristics) viene utilizzata in problemi di classificazione binaria per ottimizzare le prestazioni del classificatore. La curva ROC viene creata tracciando sulle ordinate il valore del True Positive Rate (TPR, frazione di veri positivi) e sulle ascisse il False Positive Rate (FPR, frazione di falsi positivi) a varie impostazioni. Il primo è anche noto come sensitivity, mentre il secondo si può calcolare come $1 - \text{specificity}$. Una curva ROC è il grafico dell'insieme delle coppie (FPR, TPR) al variare di un parametro del classificatore. In questo caso, il parametro da ottimizzare è la soglia da applicare all'output per classificare le ROI. Si è dunque calcolata la frazione di veri positivi e quella di falsi positivi per ogni possibile valore della soglia; tutti i punti così ottenuti nello spazio FP-TP descrivono la curva ROC. Il miglior metodo di previsione possibile produrrebbe un punto nell'angolo superiore sinistro o coordinata (0,1) dello spazio ROC, che rappresenta il 100% di sensibilità (senza falsi negativi) e il 100% di specificità (senza falsi positivi). Il punto (0,1) è anche chiamato punto di classificazione perfetta. Nell'immagine seguente (figura 8) viene illustrato il procedimento che porta al tracciamento della curva ROC. La retta nera è la soglia che separa i negativi dai positivi. Spostandola e calcolano ogni volta il TPR e il FPR si determinano i punti della curva ROC. La soglia ideale è quella per la quale la curva ROC raggiunge il punto più lontano dalla retta diagonale nello spazio ROC.

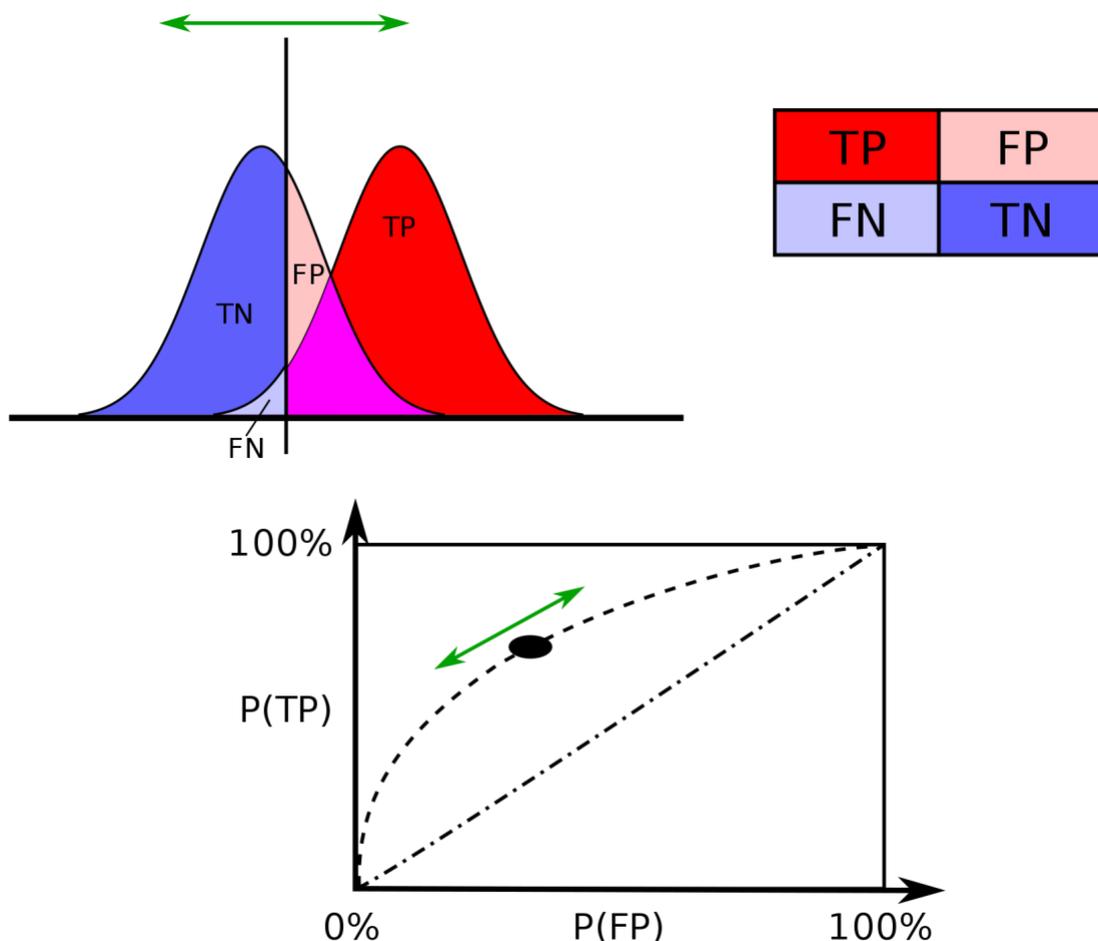


Figura 13, principio di funzionamento della curva ROC

3.3 Risultati

Per ogni combinazione di parametri descritti nel paragrafo precedente, sono stati effettuati 20 addestramenti. I valori di accuracy, sensitivity, specificity e precision che si possono vedere nei seguenti grafici rappresentano la media dei quattro indici ottenuta a seguito delle 20 prove.

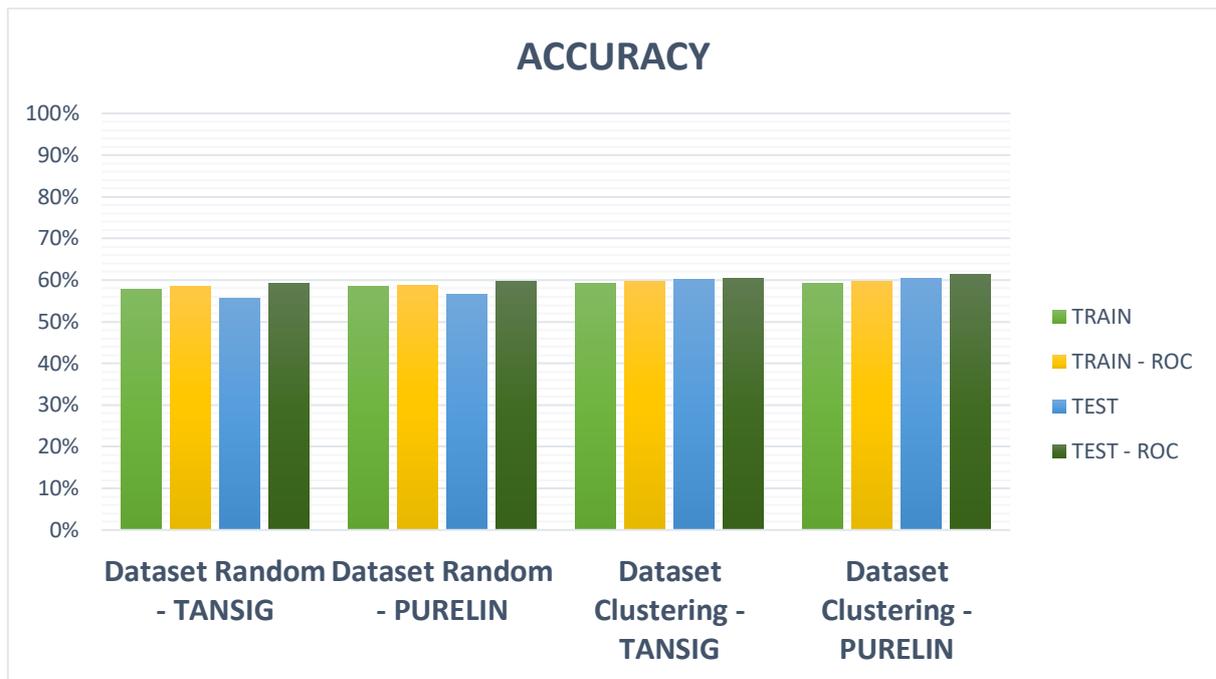


Grafico 1, Risultati rete triangolare, accuracy

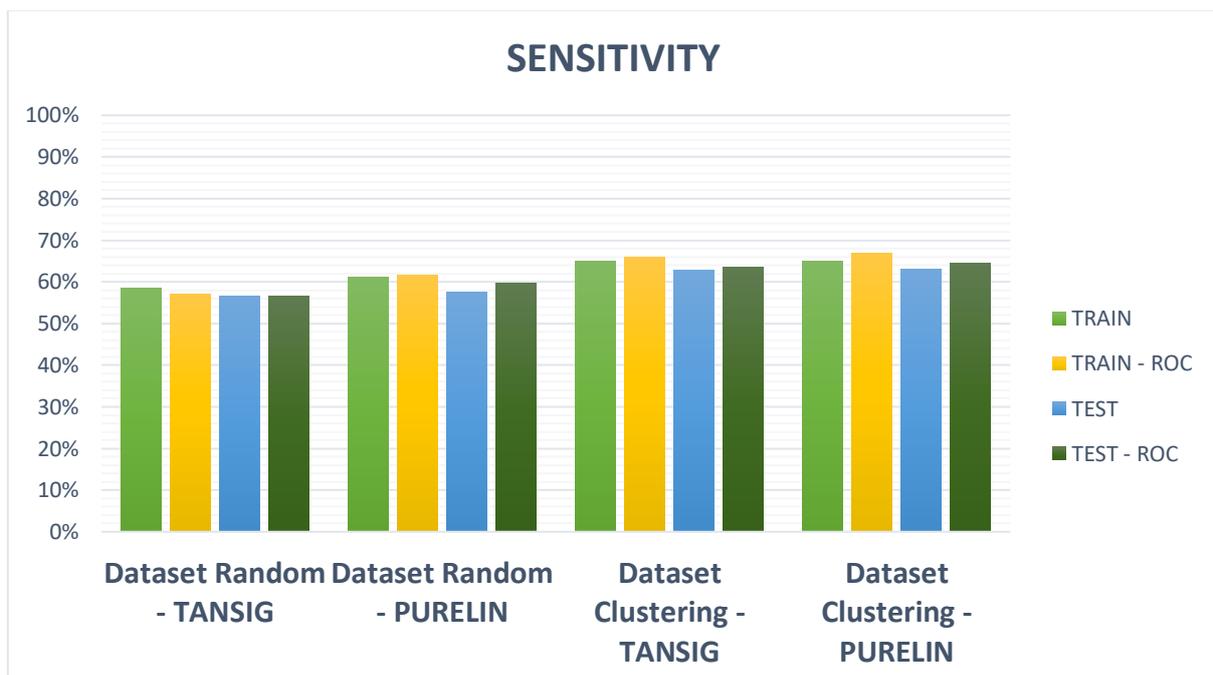


Grafico 2, Risultati rete triangolare, sensitivity

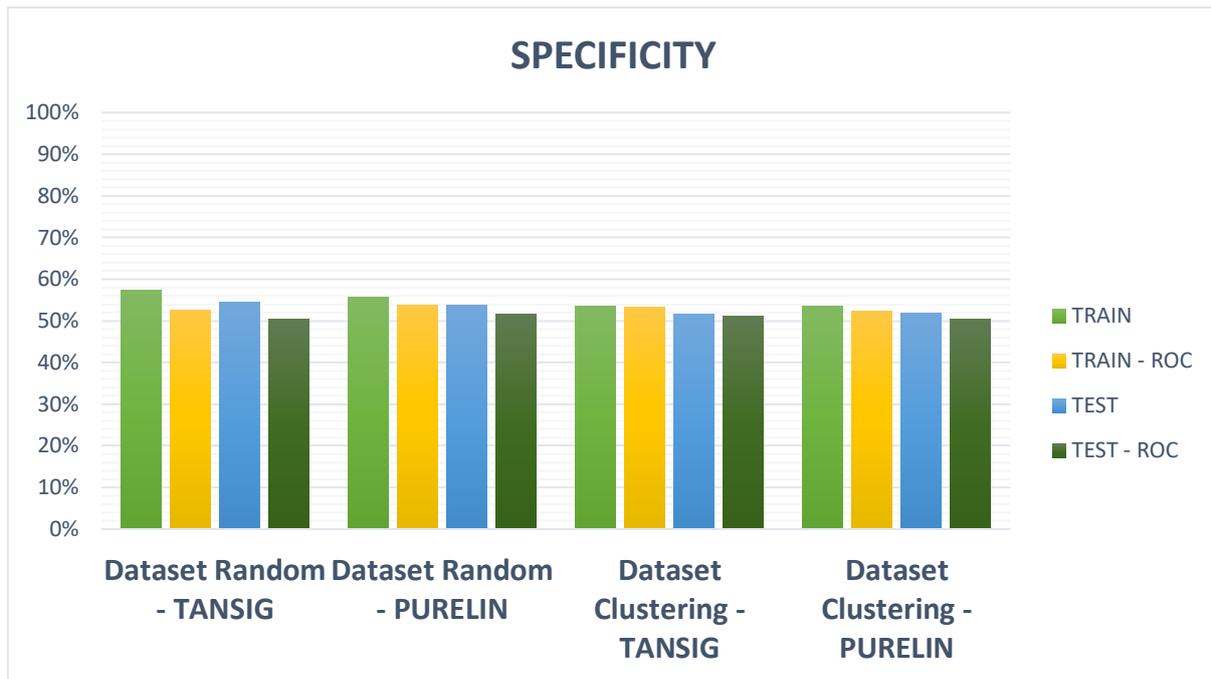


Grafico 3, Risultati rete triangolare, specificity

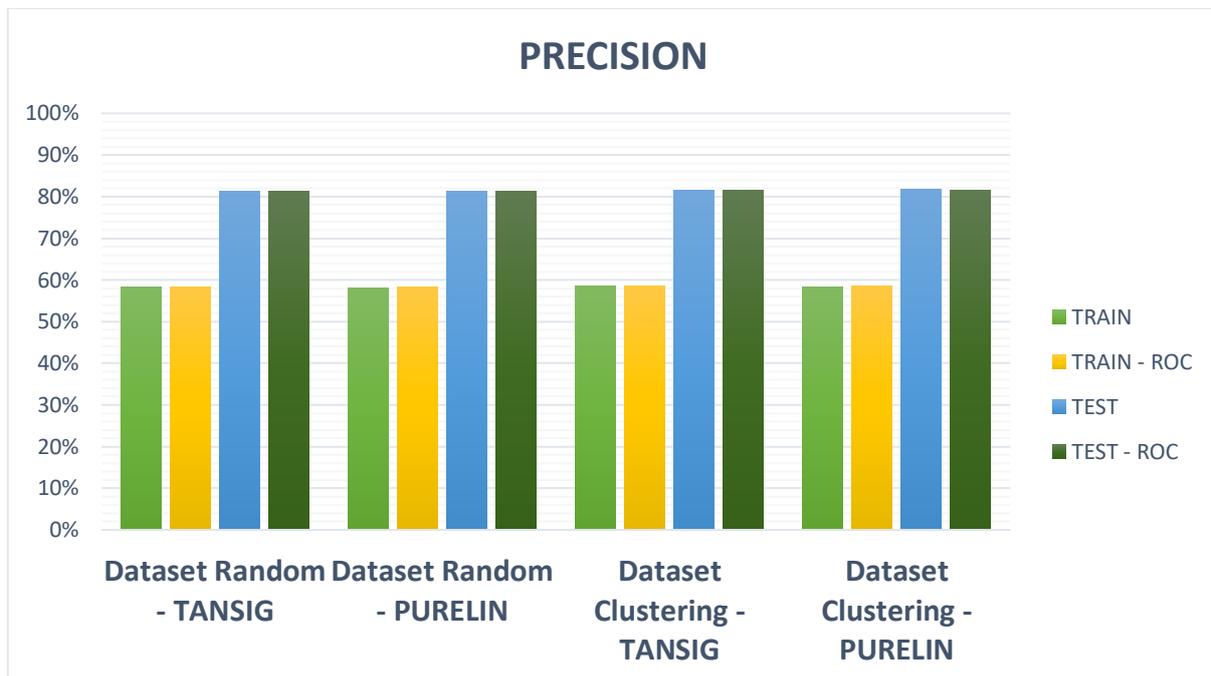


Grafico 4, Risultati rete triangolare, precision

Come è possibile notare dai grafici riportati, non si sono ottenuti mediamente risultati ottimali per nessun indice studiato, con l'unica eccezione data dalla Precision calcolata nel test set che si mantiene costantemente alta, al di sopra dell'80%, per entrambe le funzioni di attivazione, con entrambi i dataset ed indipendentemente dalla presenza o meno della soglia calcolata tramite curva ROC. Al contrario, la Specificity rimane bassa indipendentemente dalla combinazione di parametri adoperato.

Analizzando a fondo i risultati, è emerso che il dataset ottenuto tramite algoritmi di clustering ha delle performance migliori, soprattutto nel test set. Ciò è chiaramente apprezzabile andando ad osservare l'Accuracy ottenuta sul test set, che per entrambe le funzioni di attivazione passa dal 55,5% e 56,6% con il dataset creato in maniera random al 60,2% e 60,4% nel dataset estratto tramite algoritmo di clustering (rispettivamente per la funzione di attivazione sigmoide e lineare pura). Queste osservazioni hanno confermato un risultato importante, ovvero che il training set ottenuto con le reti SOM è più rappresentativo del dataset rispetto al training set ottenuto randomicamente. Inoltre, si è constatato che non vi è una differenza degna di nota in termini di performance tra le due funzioni di attivazione applicate al neurone di output e, infine, si è notato che l'utilizzo della soglia calcolata tramite curva ROC ottimizza i risultati, soprattutto in termini di Accuracy.

3.4 Reti Neurali profonde con architettura diversa

Per confermare quanto emerso nelle prove precedenti, si sono addestrate nuove reti con una architettura diversa. Nello specifico, sono state progettate reti dalla struttura meno fitta, quindi con un numero inferiore di layers nascosti. Ciò ha portato ad un costo computazionale della fase di training e di valutazione delle prestazioni significativamente più basso rispetto a quello necessario alla rete triangolare utilizzata in principio. Sono state sottoposte ad analisi 8 tipologie di architettura diverse, andando a variare il numero di layers nascosti, mentre il layer di input con 36 neuroni ed il layer di output con un neurone sono rimasti invariati. La seguente tabella (tabella 2) mostra le strutture poste sotto analisi, dove i numeri tra parentesi quadre indicano il numero di neuroni per quel layer nascosto.

N.	Numero di neuroni per layer nascosto
#1	[34-32-30-28-26-24-22-20-18-16-14-12-10-8-6-4-2]
#2	[33-30-27-24-21-18-15-12-9-6-3]
#3	[32-28-24-20-16-12-8-4]
#4	[31-26-21-16-11-6-3]
#5	[30-24-18-12-6-3]
#6	[28-20-12-8-4]
#7	[27-18-9-5-3]
#8	[18-9-5-3]

Tabella 2, Strutture

Per la valutazione di queste reti si è proceduto seguendo gli stessi passi effettuati per la rete triangolare, ovvero sono state svolte diverse prove andando a variare dei parametri, gli stessi che sono stati analizzati per la prima DNN. Anche questa volta sono stati effettuati 20 addestramenti per ogni combinazione di parametri, per ogni struttura.

3.5 Risultati con architettura diversa

Nei seguenti grafici sono rappresentate le medie dei 4 indici ottenute dopo 20 addestramenti per ogni combinazione di parametri, utilizzando le diverse architetture descritte nel paragrafo precedente. E' possibile inoltre vedere il confronto con la prima DNN, la triangolare, indicata nei grafici come TR.

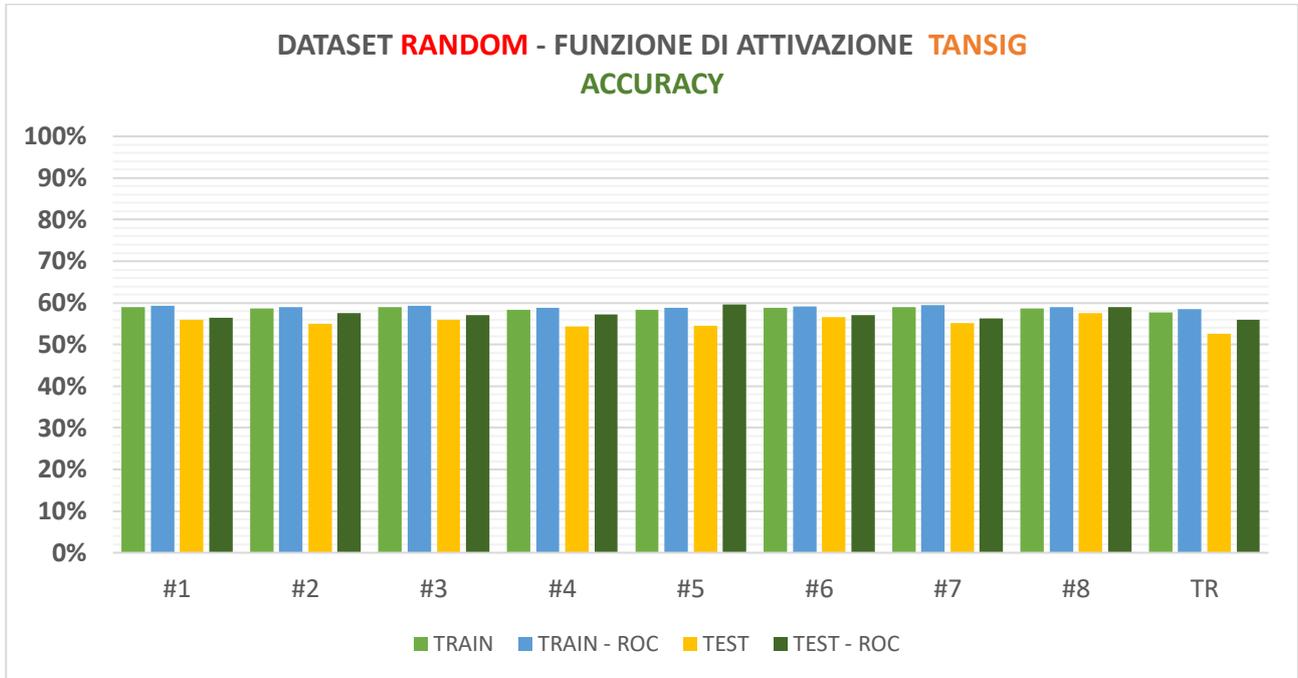


Grafico 5, Risultati 8 strutture, dataset random, funzione di attivazione tansing, accuracy

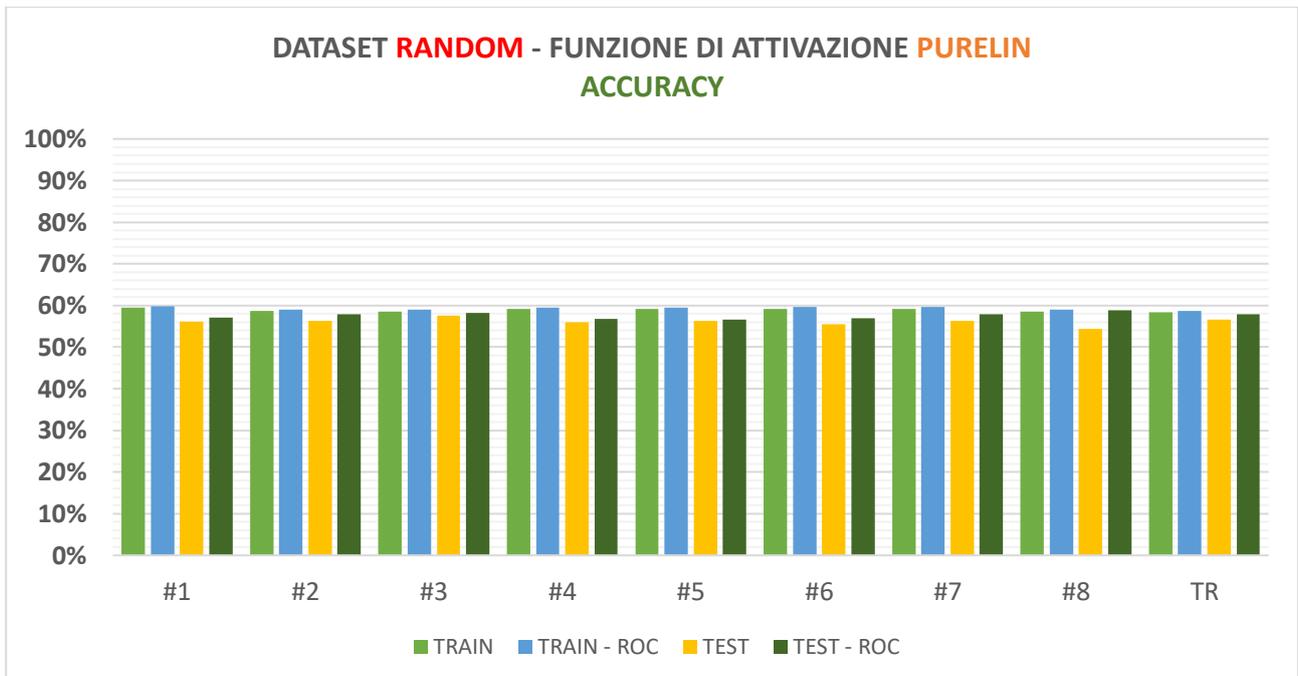


Grafico 6, Risultati 8 strutture, dataset random, funzione di attivazione purelin, accuracy

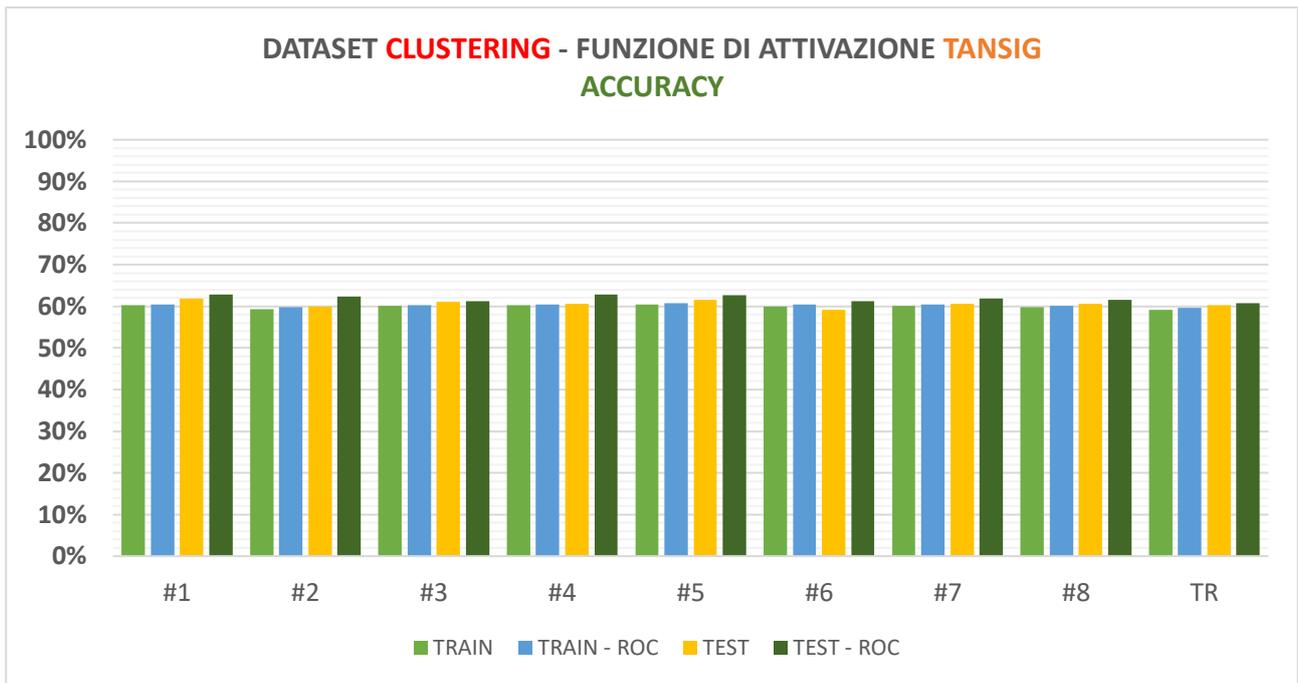


Grafico 7, Risultati 8 strutture, dataset clustering, funzione di attivazione tansig, accuracy

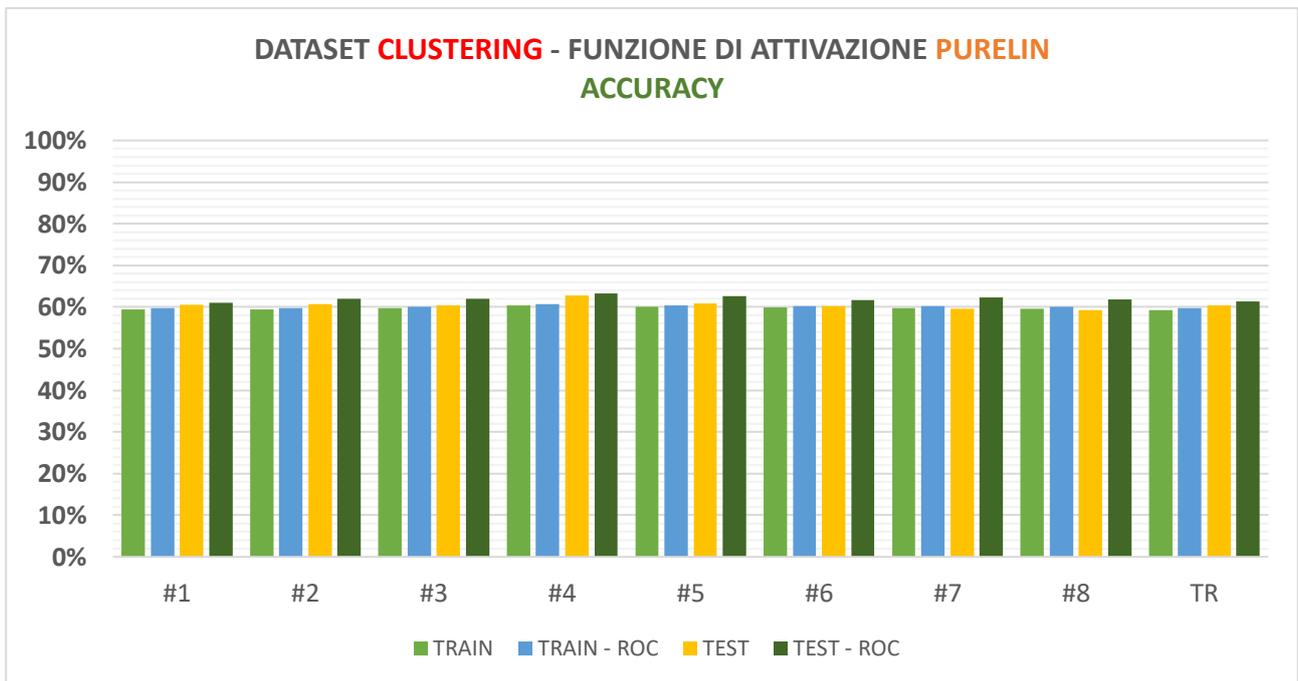


Grafico 8, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, accuracy

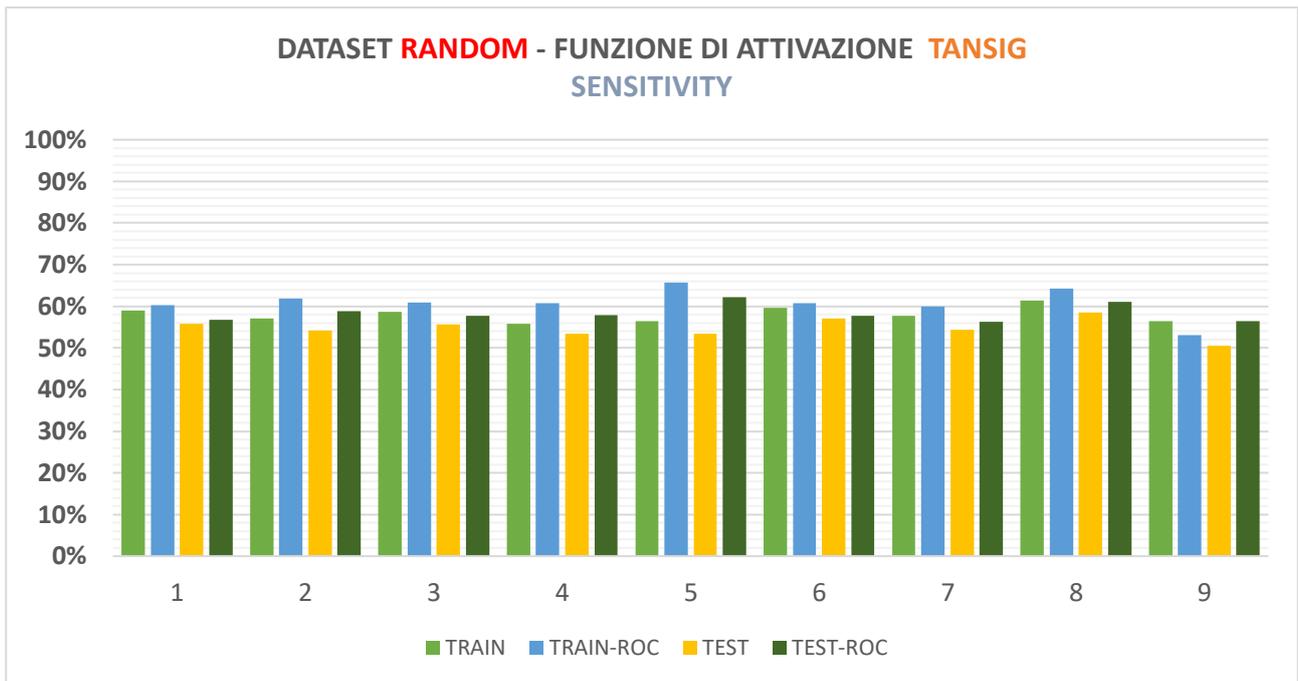


Grafico 9, Risultati 8 strutture, dataset random, funzione di attivazione tansing, sensitivity

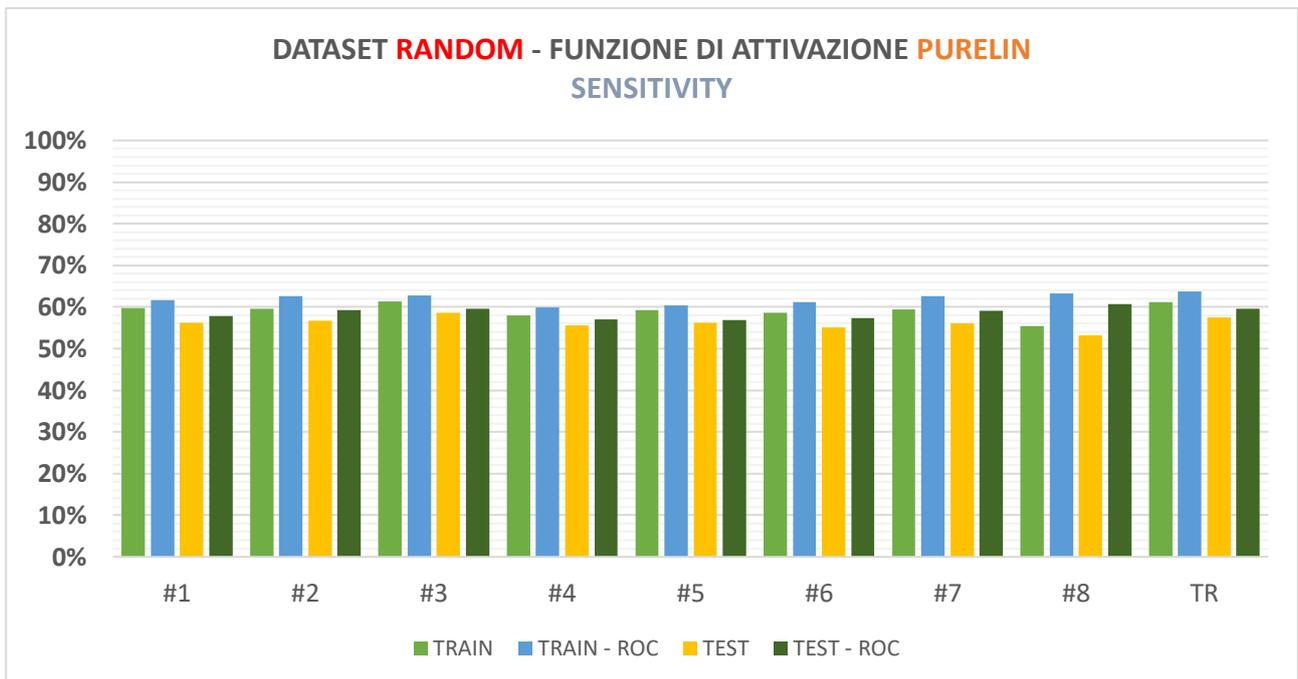


Grafico 10, Risultati 8 strutture, dataset random, funzione di attivazione purelin, sensitivity

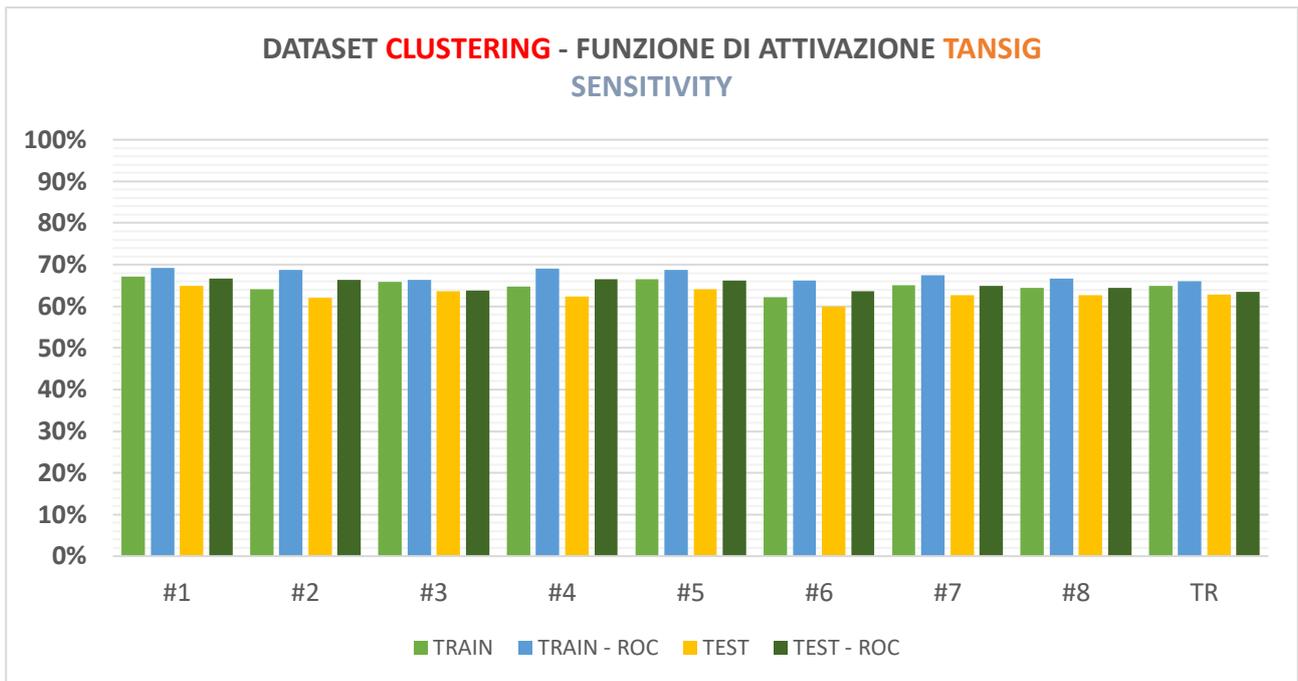


Grafico 11, Risultati 8 strutture, dataset clustering, funzione di attivazione tansig, sensitivity

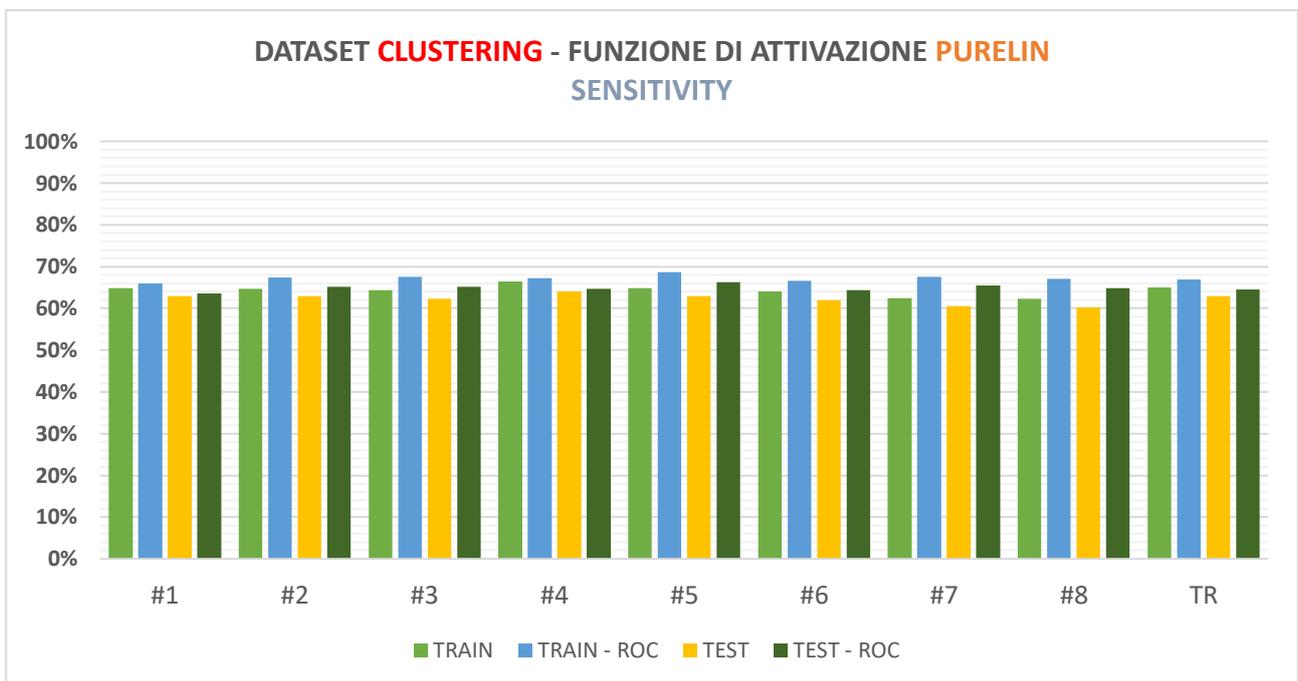


Grafico 12, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, sensitivity

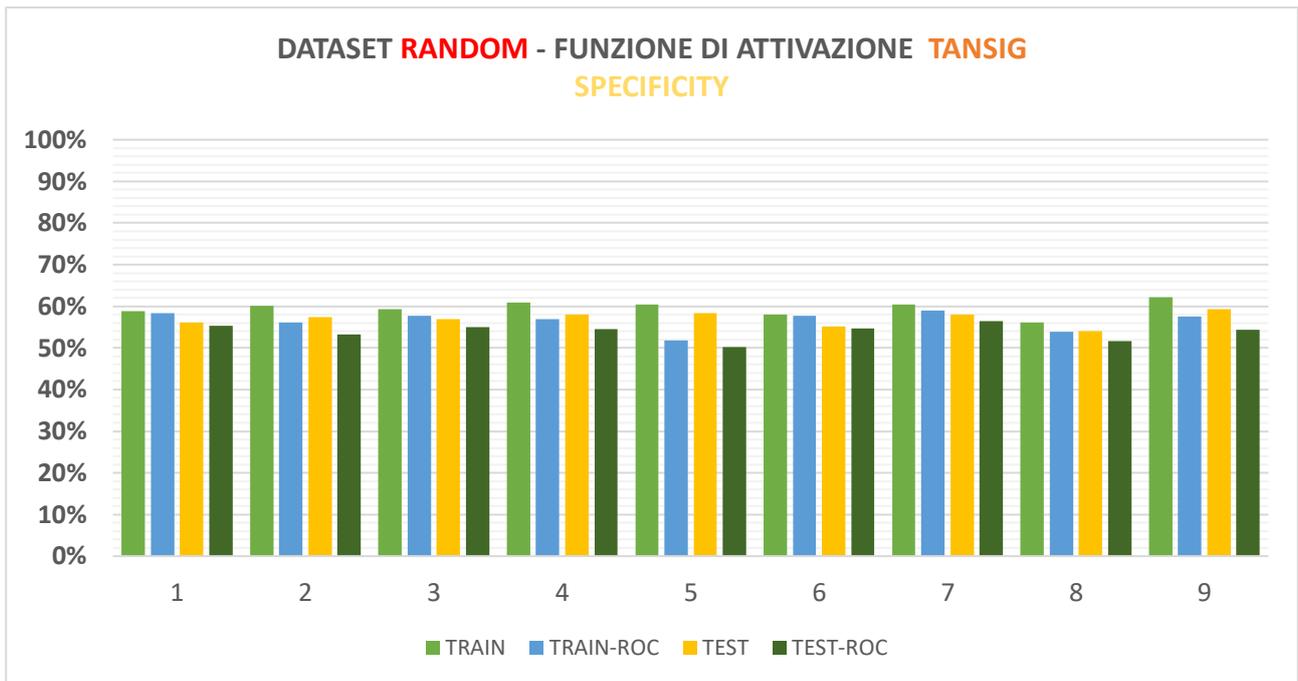


Grafico 13, Risultati 8 strutture, dataset random, funzione di attivazione tansig, specificity

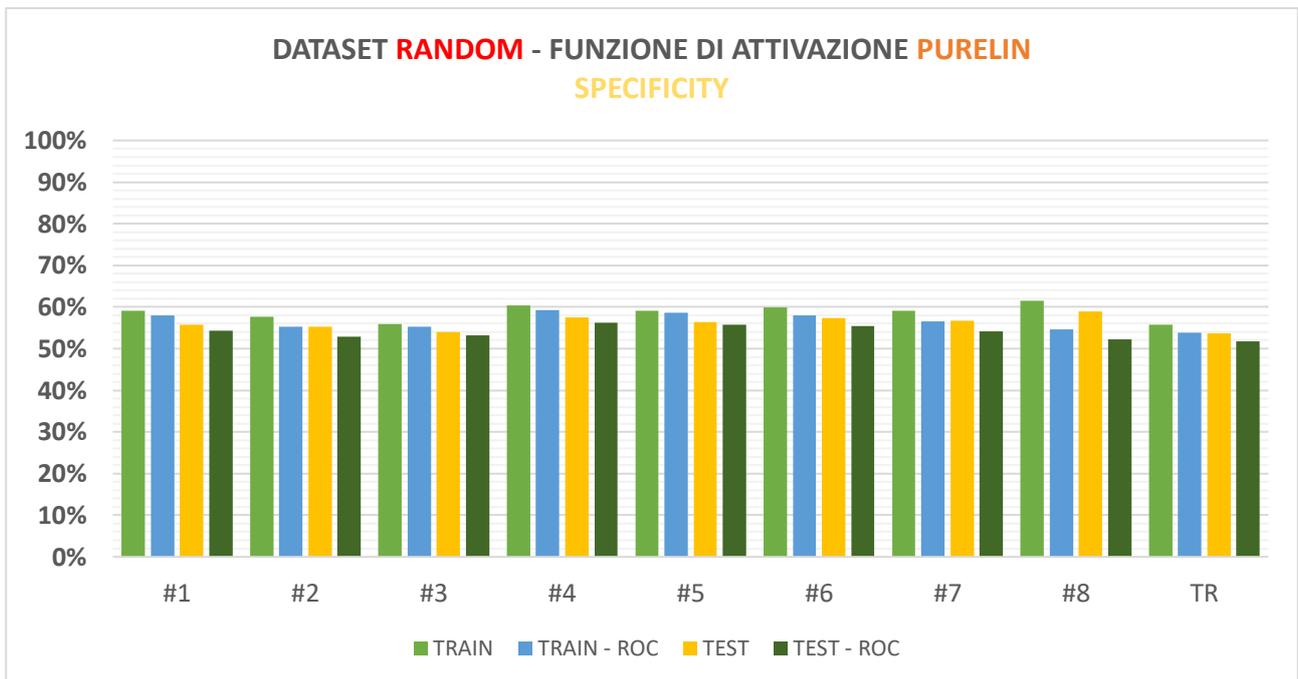


Grafico 14, Risultati 8 strutture, dataset random, funzione di attivazione purelin, specificity

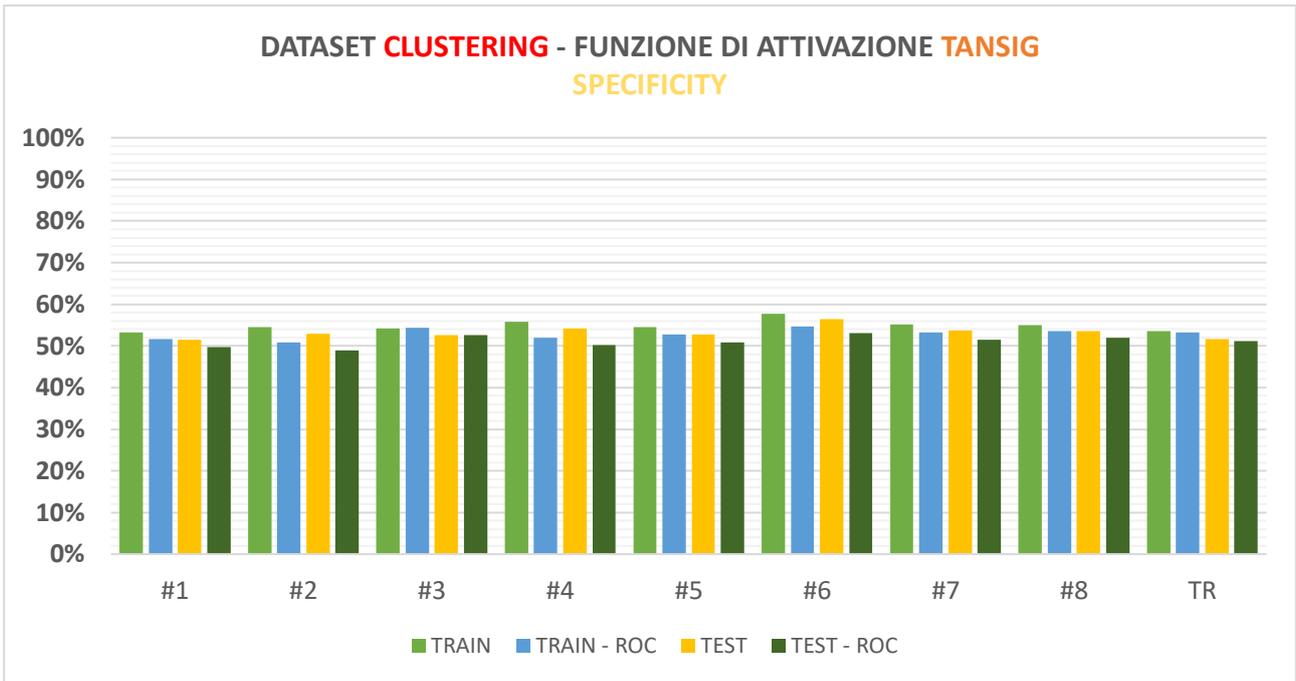


Grafico 15, Risultati 8 strutture, dataset clustering, funzione di attivazione tansig, specificity

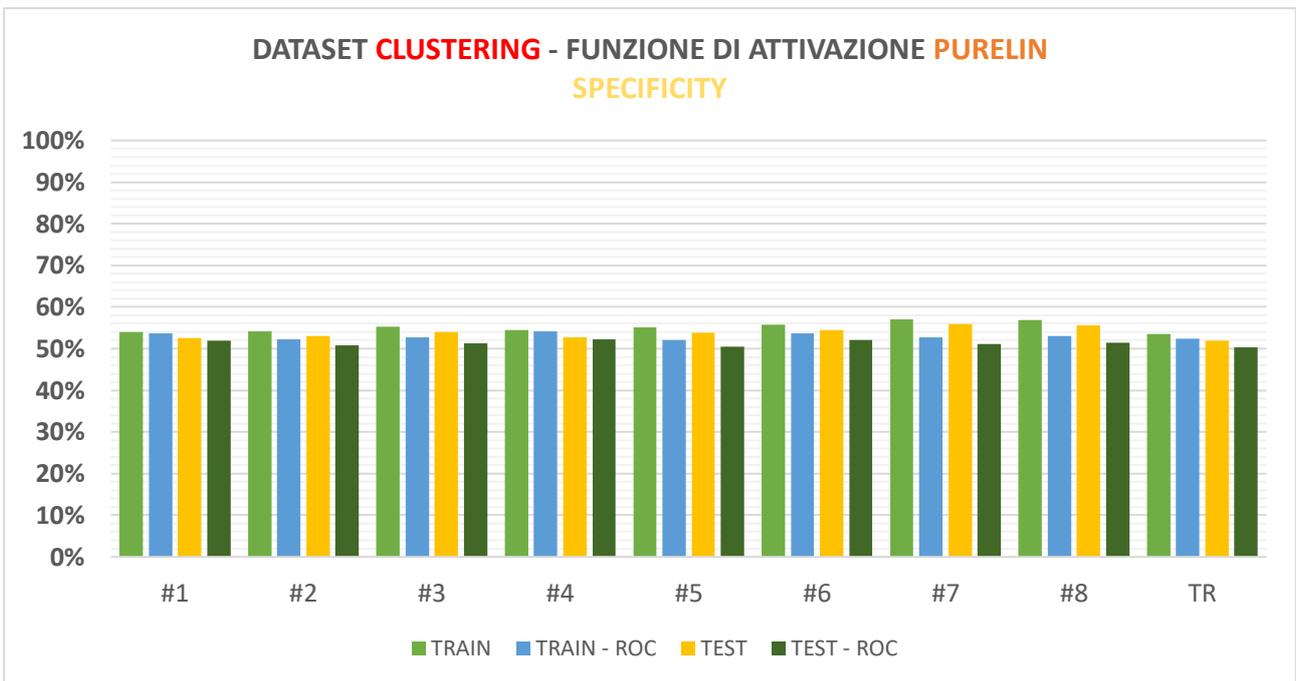


Grafico 16, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, specificity

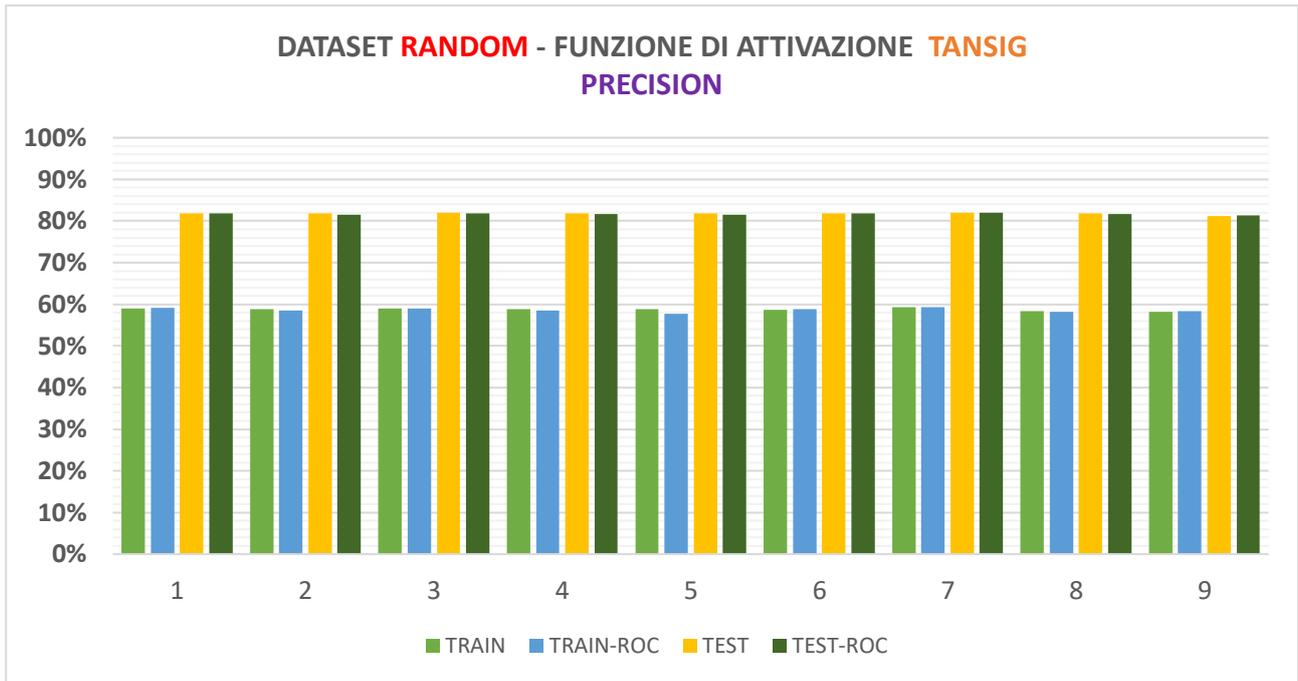


Grafico 17, Risultati 8 strutture, dataset random, funzione di attivazione tansing, precision

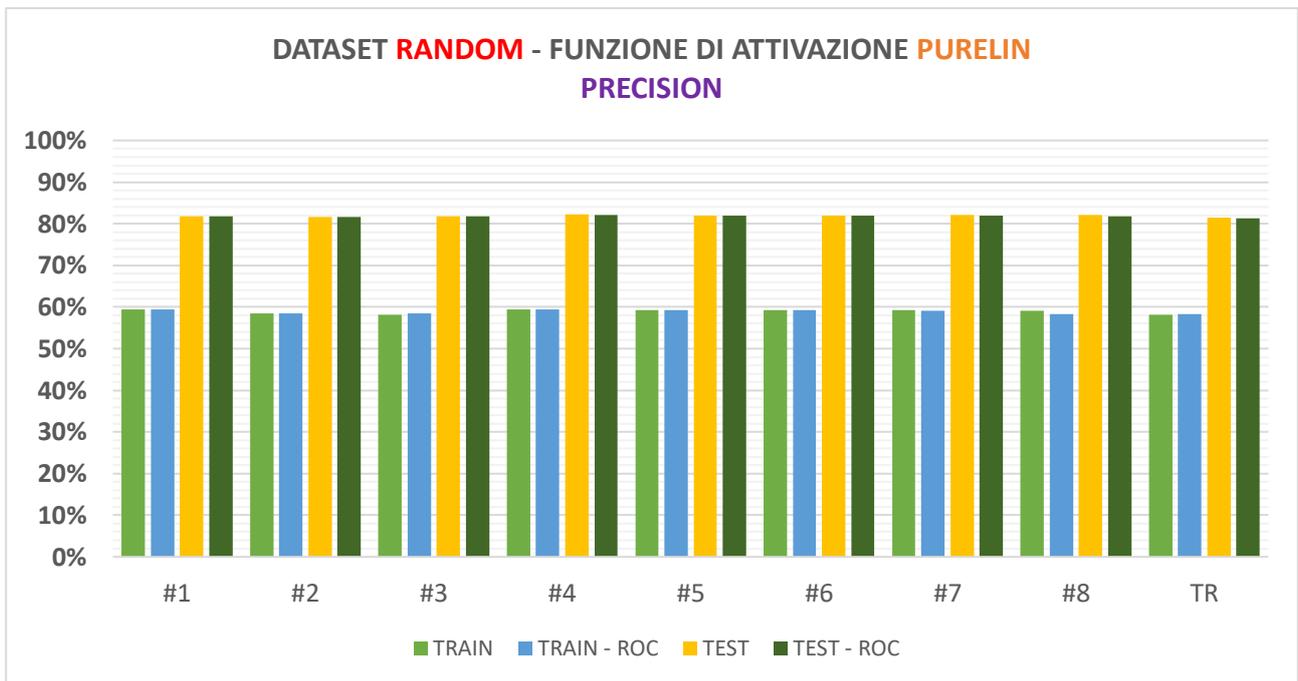


Grafico 18, Risultati 8 strutture, dataset random, funzione di attivazione purelin, precision

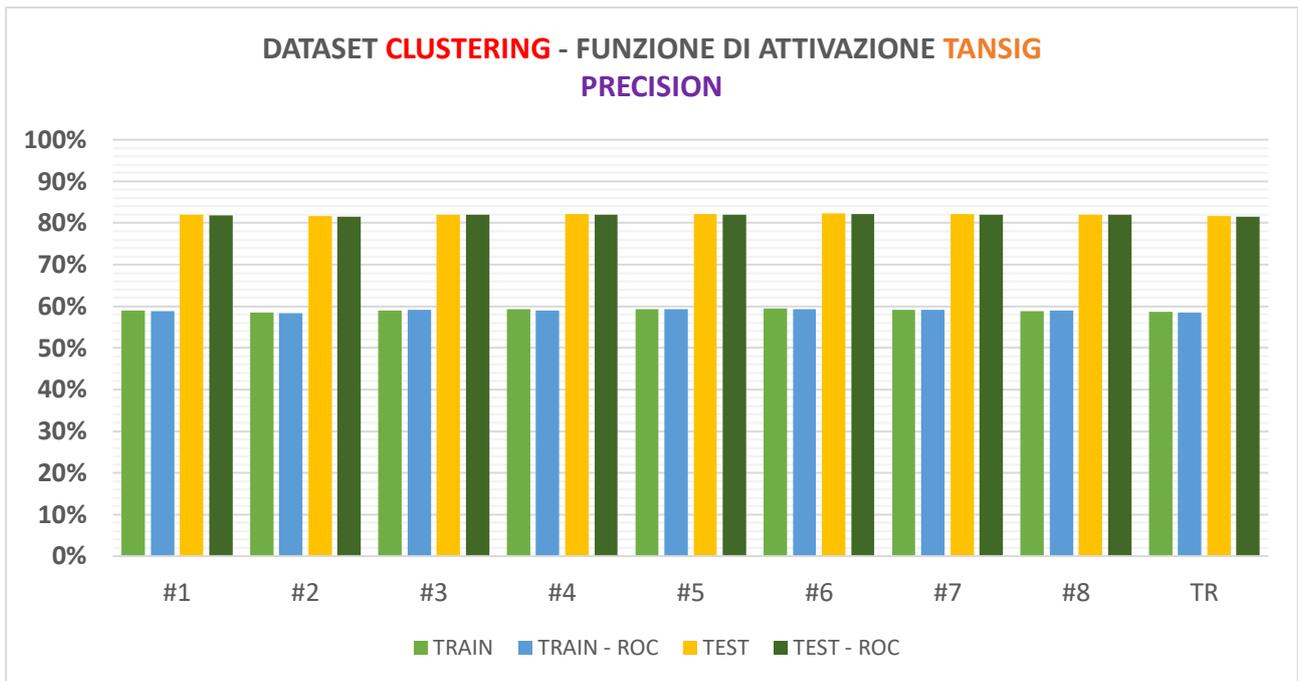


Grafico 19, Risultati 8 strutture, dataset clustering, funzione di attivazione tansig, precision

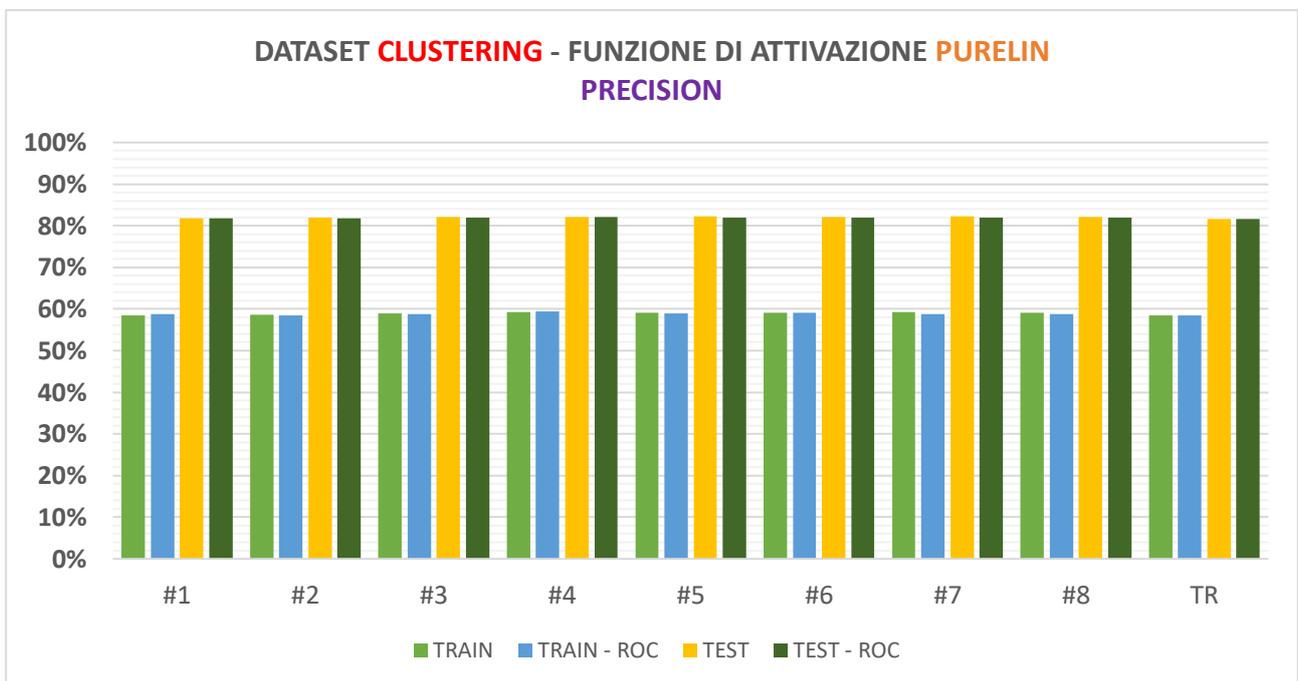


Grafico 20, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, precision

I risultati ottenuti hanno dimostrato che le varie strutture non presentano differenze sostanziali tra loro in termini di performance. Inoltre, è possibile osservare che i quattro indici hanno un andamento generale che è in linea con quello osservato per la DNN triangolare. Infatti, anche in questo caso, è emerso che le due funzioni di attivazione non presentano differenze importanti per nessuno dei quattro indici. Inoltre, anche in questo caso l'utilizzo della soglia calcolata da curva ROC ha portato ad un'ottimizzazione dei risultati. Analizzando i grafici 5, 6, 7 e 8, raffiguranti l'accuracy, è chiaro come il dataset ottenuto tramite clustering abbia prestazioni migliori rispetto all'altro dataset. Questo è una conferma di quanto constatato precedentemente, valutando i risultati restituiti dalla rete triangolare. Nonostante si sia detto che, in linea di massima, le varie strutture poste ad analisi non presentino differenze marcate, è possibile notare come l'accuracy nel test sia mediamente più alta (62,77%) per le reti con funzione di attivazione Purelin e struttura #4, addestrate col training set "clustering" (Grafico 8). Questa rete è stata quindi selezionata per il procedimento successivo, l'ottenimento delle maschere RM predette con DNN.

3.6 Creazione maschere RM tramite DNN

Come affermato nel paragrafo precedente, è stata scelta la struttura #4, con funzione di attivazione *purelin* ed addestrata con training set ottenuto tramite reti SOM, per l'ottenimento delle maschere, in quanto queste reti hanno mostrato un valore di accuracy calcolata nel test set mediamente maggiore delle altre strutture. Con questa architettura erano state addestrate 20 reti. Analizzando le prestazioni delle singole reti nel test set, sono state selezionate due reti per l'ottenimento delle maschere: la prima presenta i valori maggiori di accuracy e sensitivity (1° rete), mentre l'altra ha restituito i più alti valori di specificity e precision (2° rete). Con queste due reti si è proceduto alla classificazione delle ROI di tutte le slice di tutti i 50 pazienti.

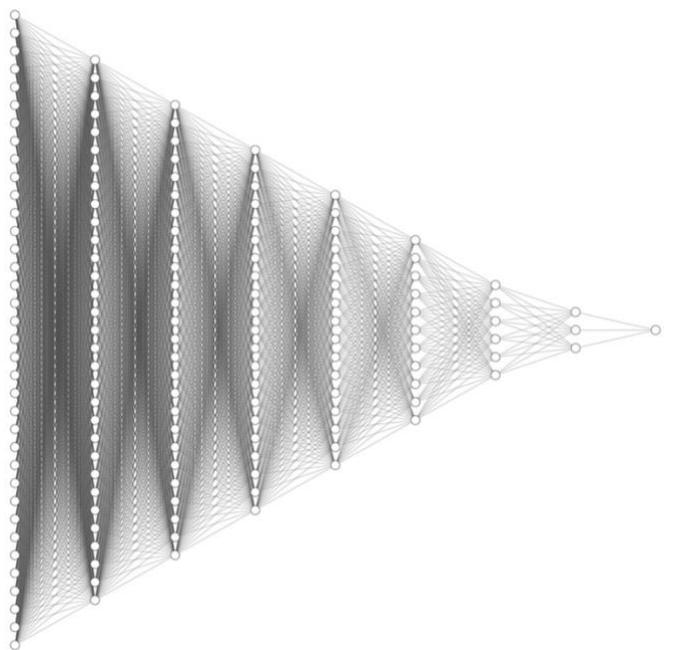


Figura 14, Struttura #4 scelta per la segmentazione

Le ROI analizzate, come già ribadito, sono di dimensione 5x5 e contengono pertanto 25 pixel al loro interno. Per classificare i singoli pixel si è adoperato l'algoritmo del voting. Infatti, molto frequentemente, lo stesso pixel è incluso in un numero diverso di ROI: per classificare quel singolo pixel l'algoritmo conteggia in quante ROI questo è compreso; successivamente procede a valutare quante di queste ROI sono state classificate dalla rete come YM e quante come RM. Il pixel viene infine classificato come RM nel caso in cui il numero di ROI, contenenti quel pixel, classificate come RM siano in maggioranza rispetto alle ROI classificate come YM. Se accade il contrario, viene etichettato come YM.

Attraverso questo procedimento si sono così ottenute le maschere binarie, predette dal classificatore, rappresentanti il midollo osseo attivo. Per valutare la bontà di questa segmentazione, si sono utilizzate come confronto le maschere ottenute dalla FDG-PET. Tra queste si sono calcolati i seguenti indici (equazioni 9, 10, 11):

$$DSC = \frac{2(true\ positives)}{2(true\ positives)+false\ positives+false\ negatives} \quad (9)$$

$$precision = \frac{true\ positives}{true\ positives+false\ positives} \quad (10)$$

$$recall = \frac{true\ positives}{true\ positives+false\ negatives} \quad (11)$$

Dove, in questo caso, true positives indica il numero di pixel classificati come RM e che realmente lo raffigurano. False positives e false negatives indicano, rispettivamente, il numero di pixel classificati come RM ma che in realtà fanno parte del YM e viceversa. DSC è il Dice Similarity coefficient.

3.7 Risultati segmentazione con DNN

Di seguito è possibile osservare degli esempi di segmentazione tramite DNN. L'immagine in scala di grigi rappresenta l'immagine CT moltiplicata per la maschera predetta dalle DNN. Il bordo viola rappresenta invece la maschera reale del RM ricavata dalla FDG-PET. In verde è raffigurata la maschera dell'intero osso compatto, che era stata rimossa precedentemente in fase di pre-processing. La stessa immagine di partenza (stesso paziente e stessa slice) viene presentata due volte: per la prima, si è ottenuta la maschera predetta con la 1° DNN, mentre la seconda rappresenta il risultato della segmentazione attraverso la 2° DNN. In alto a sinistra sono stampati i valori di Pr, DSC e Re per quella slice. Nella parte bassa dell'immagine è riportato l'ID paziente ed il numero della slice analizzata. Per ottenere le immagini è stato implementato un algoritmo che taglia l'immagine centrandola nella maschera RM reale.

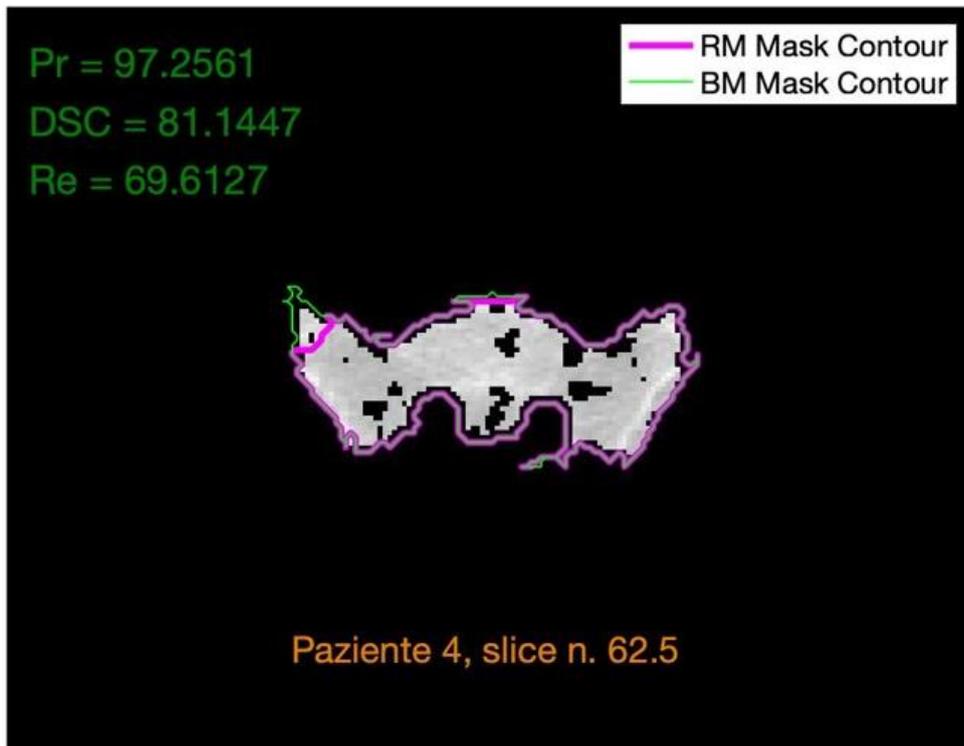


Figura 15, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

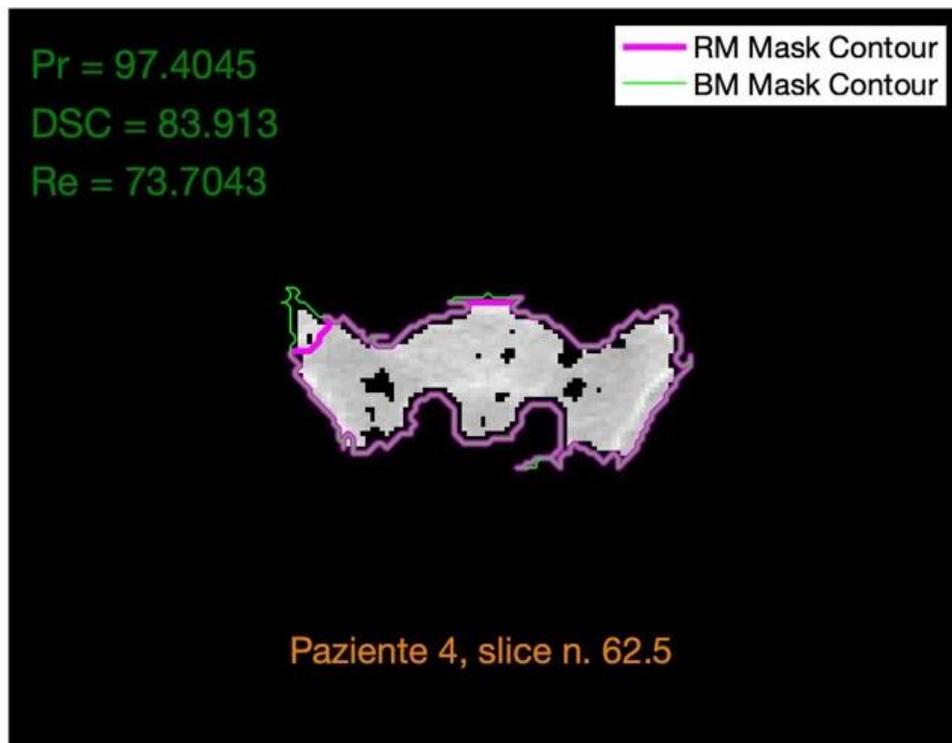


Figura 16, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

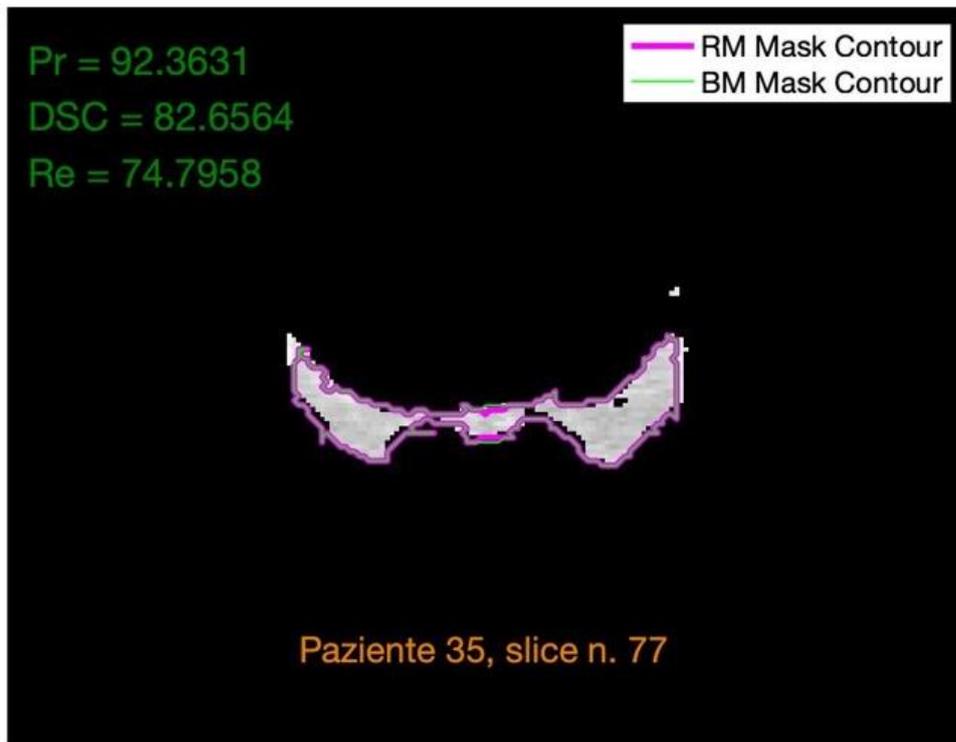


Figura 17, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

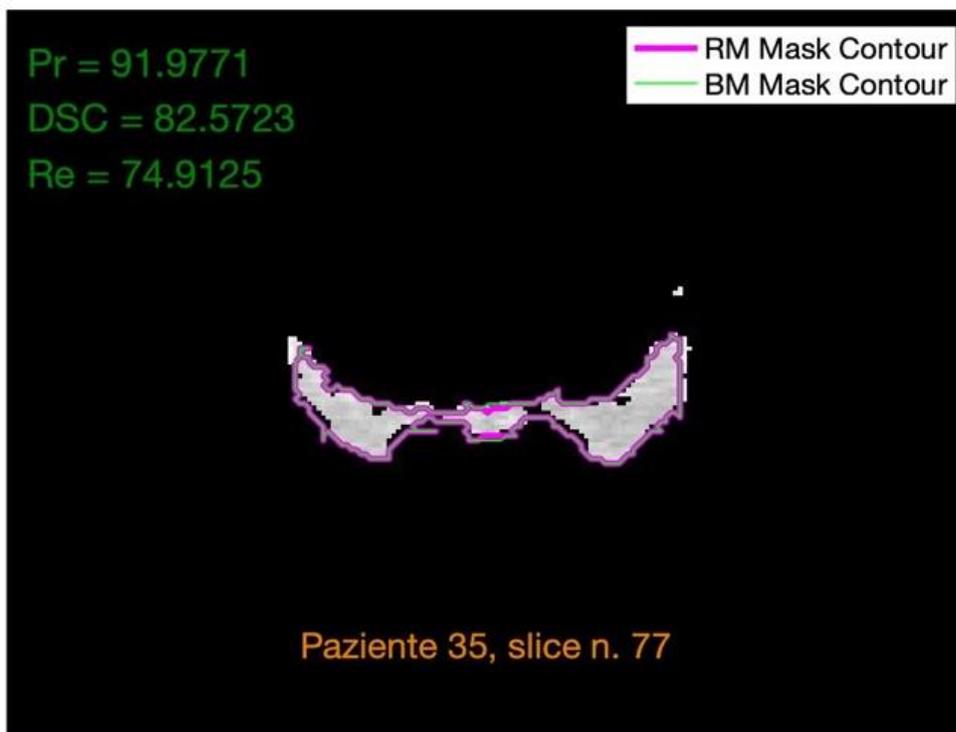


Figura 18, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

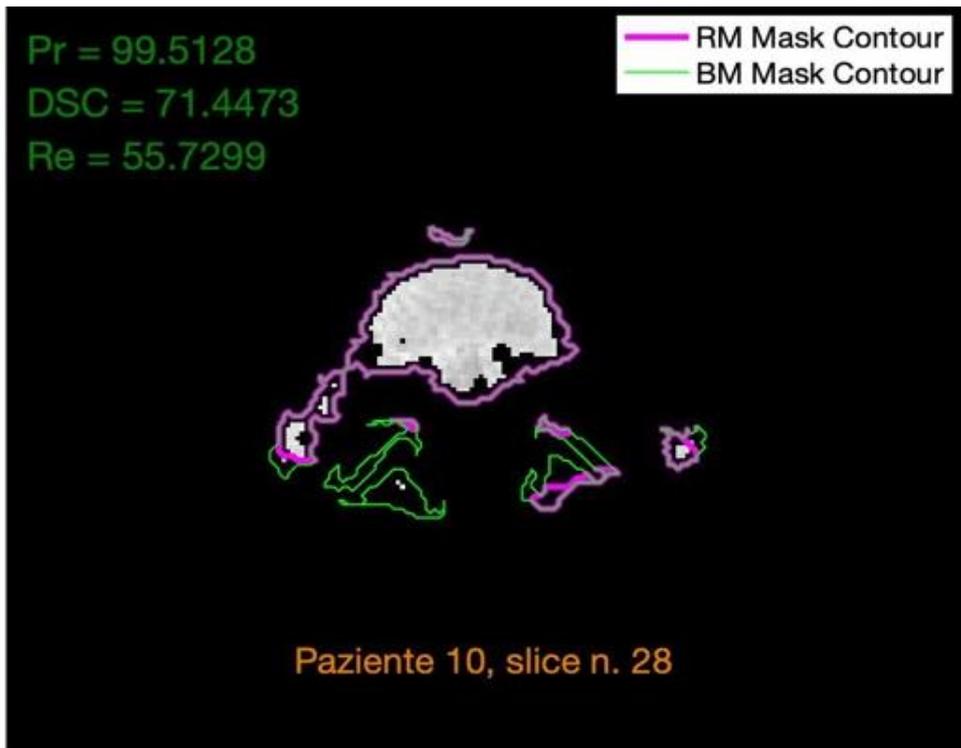


Figura 19, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

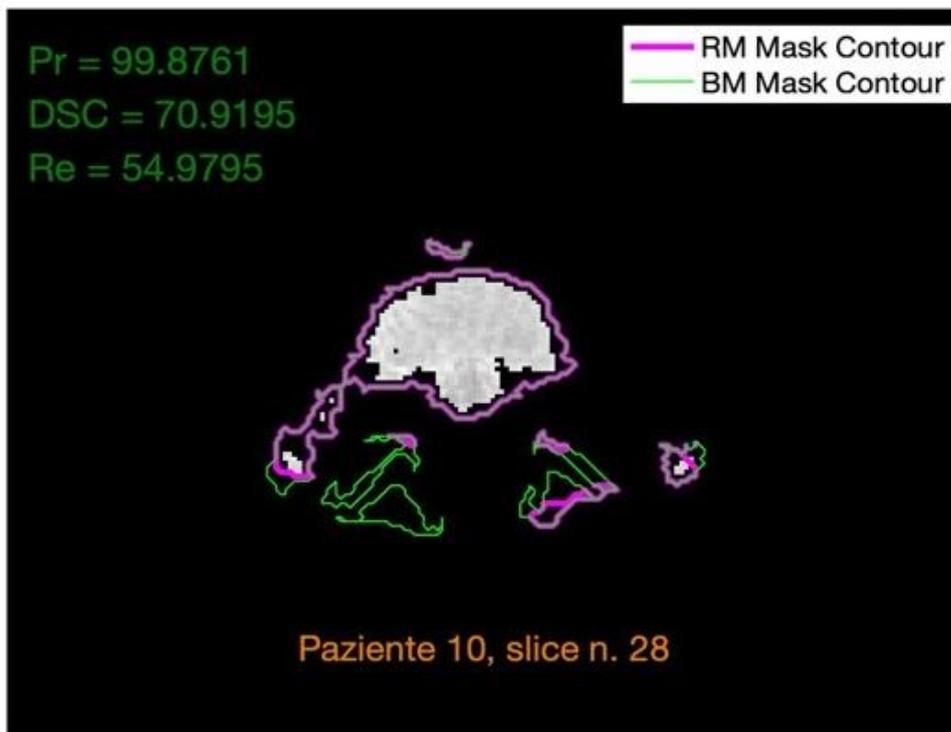


Figura 20, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

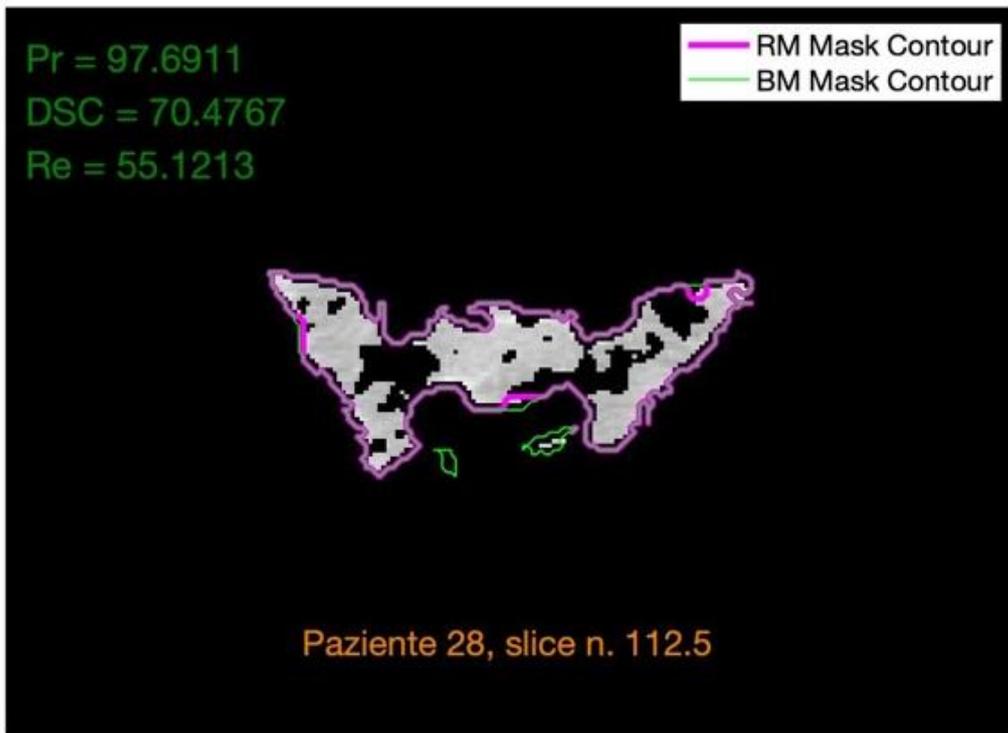


Figura 21, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

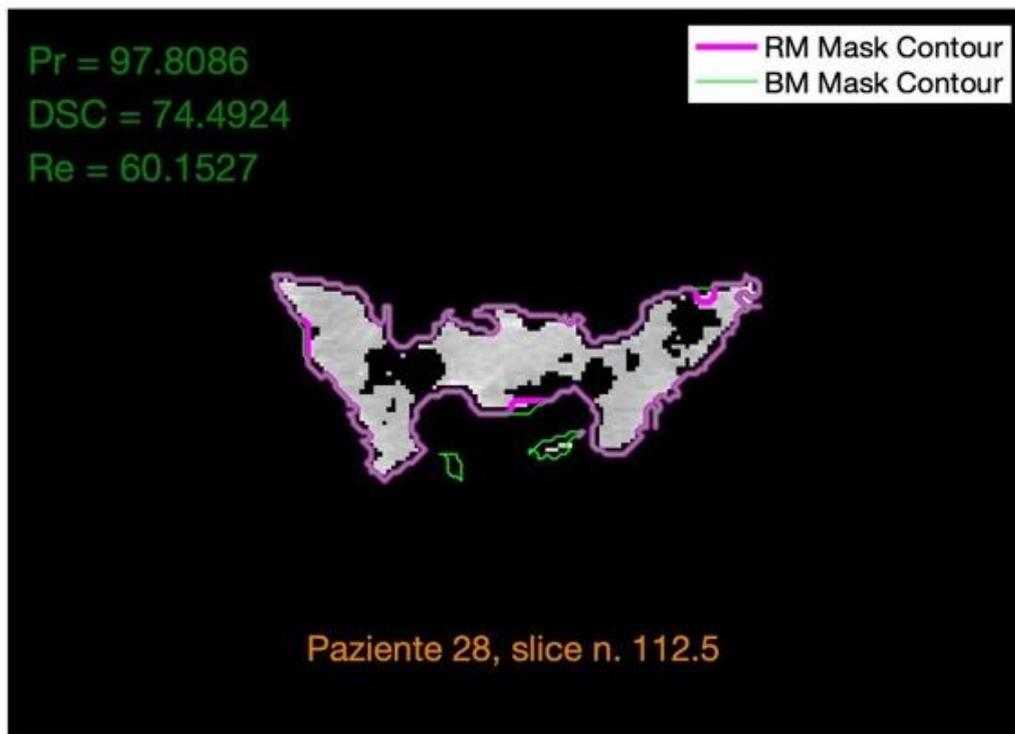


Figura 22, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

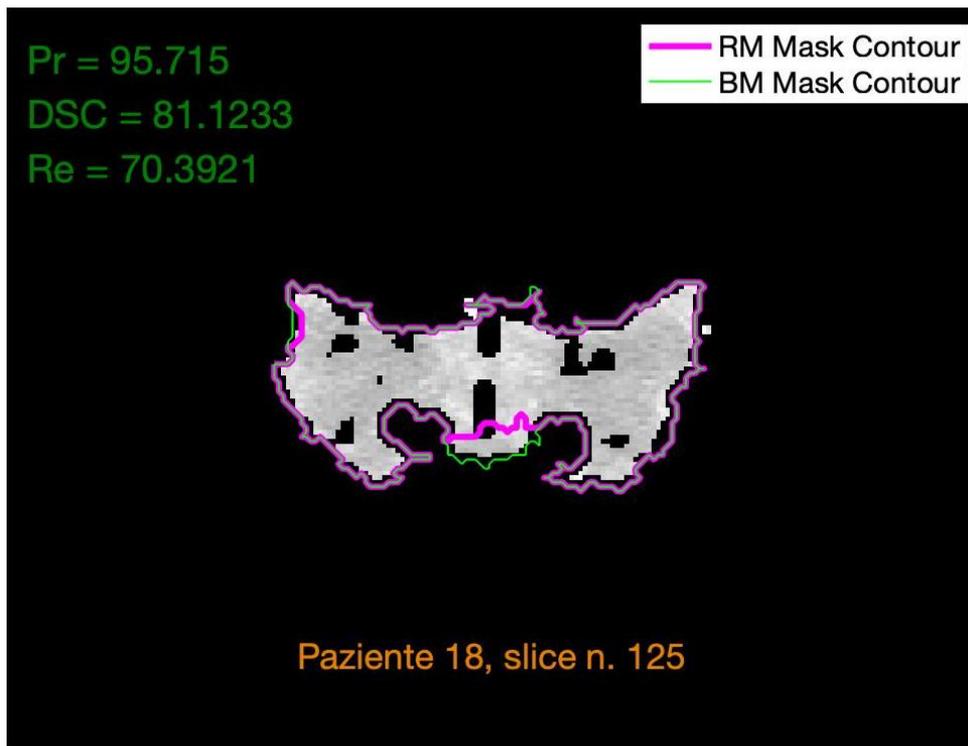


Figura 23, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

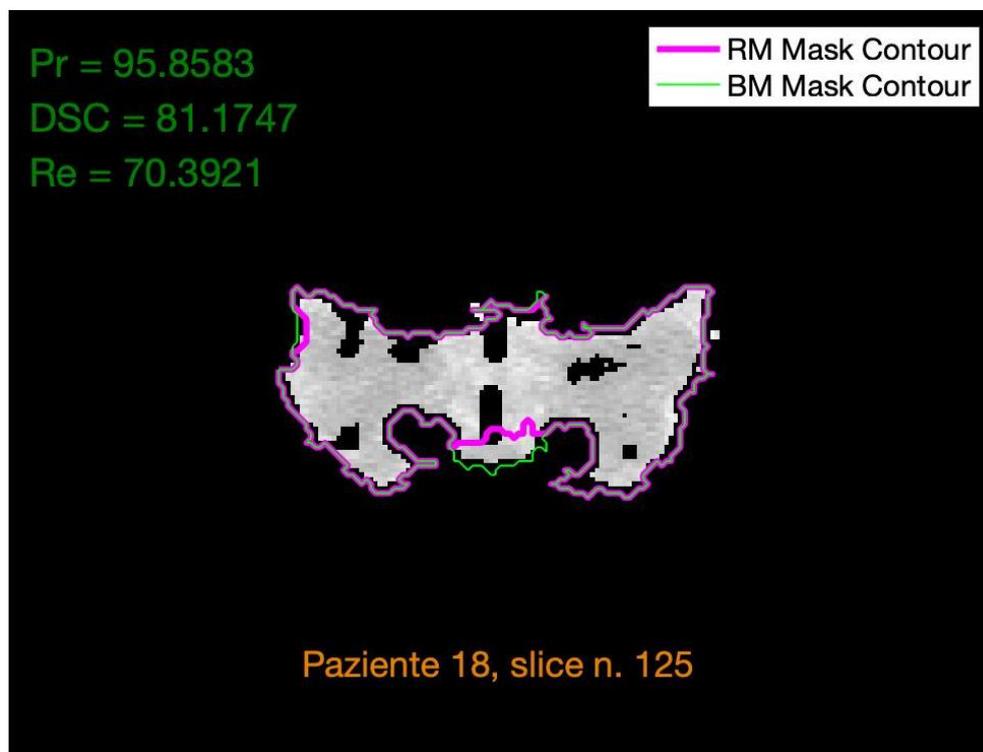


Figura 14, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

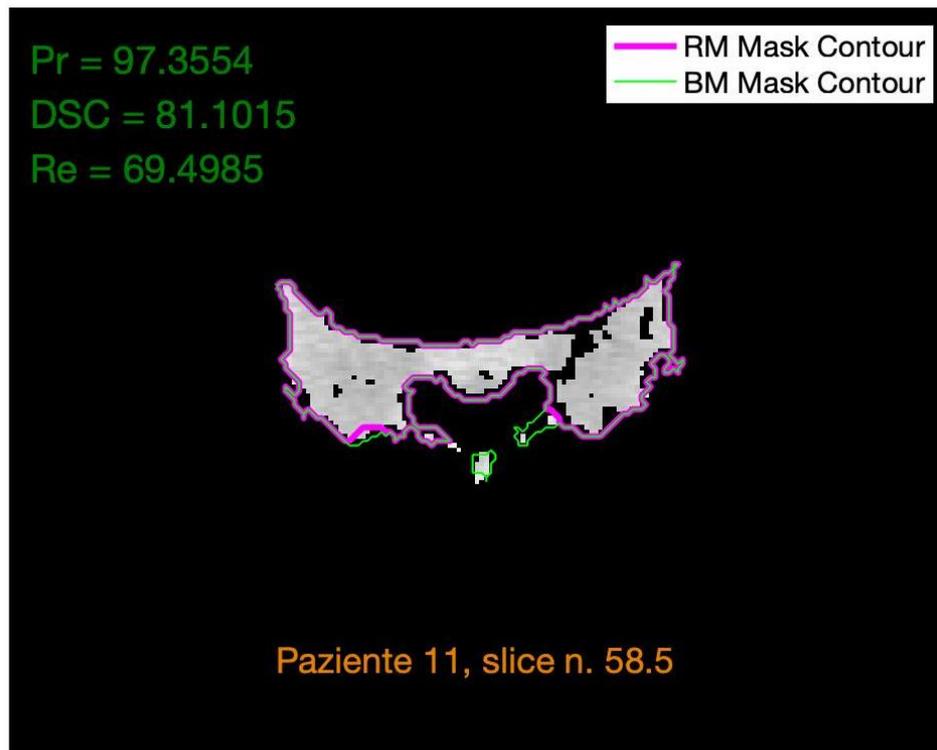


Figura 25, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

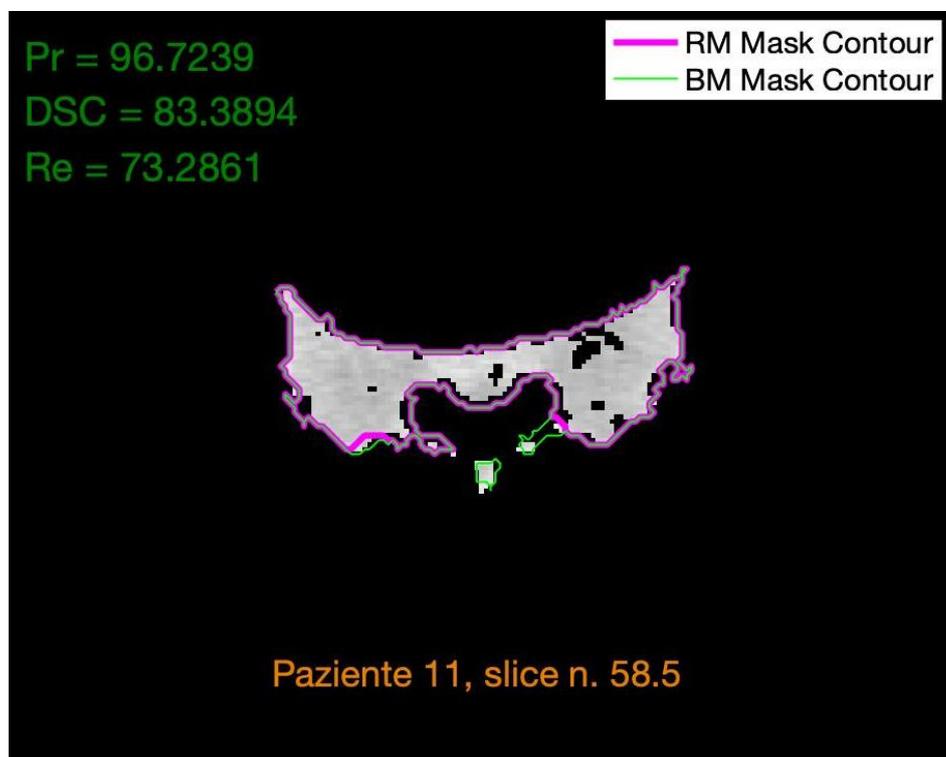


Figura 26, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

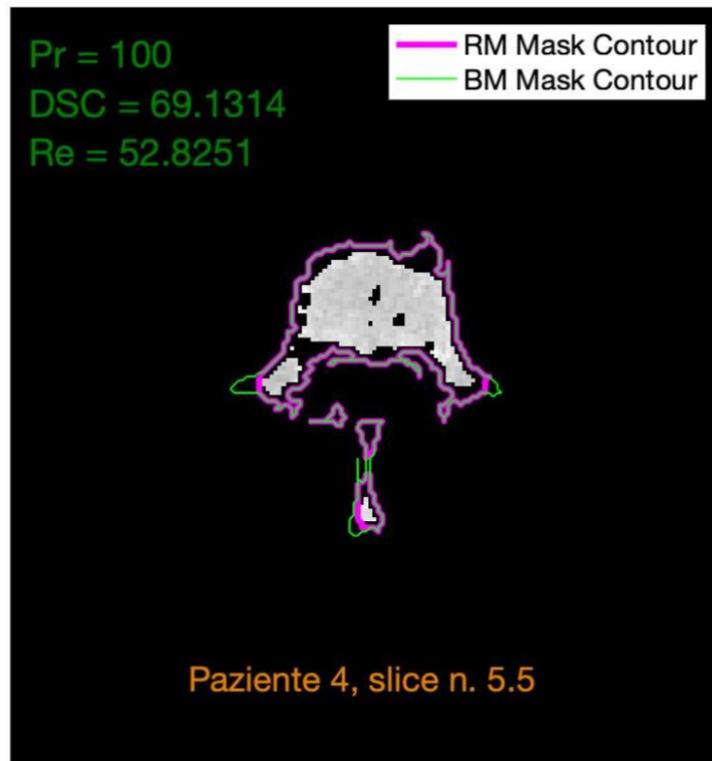


Figura 27, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

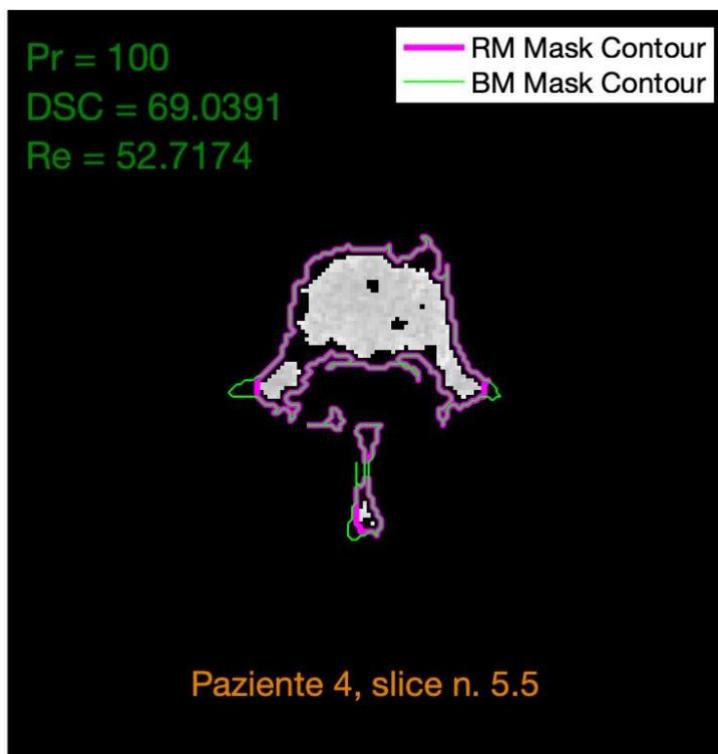


Figura 28, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

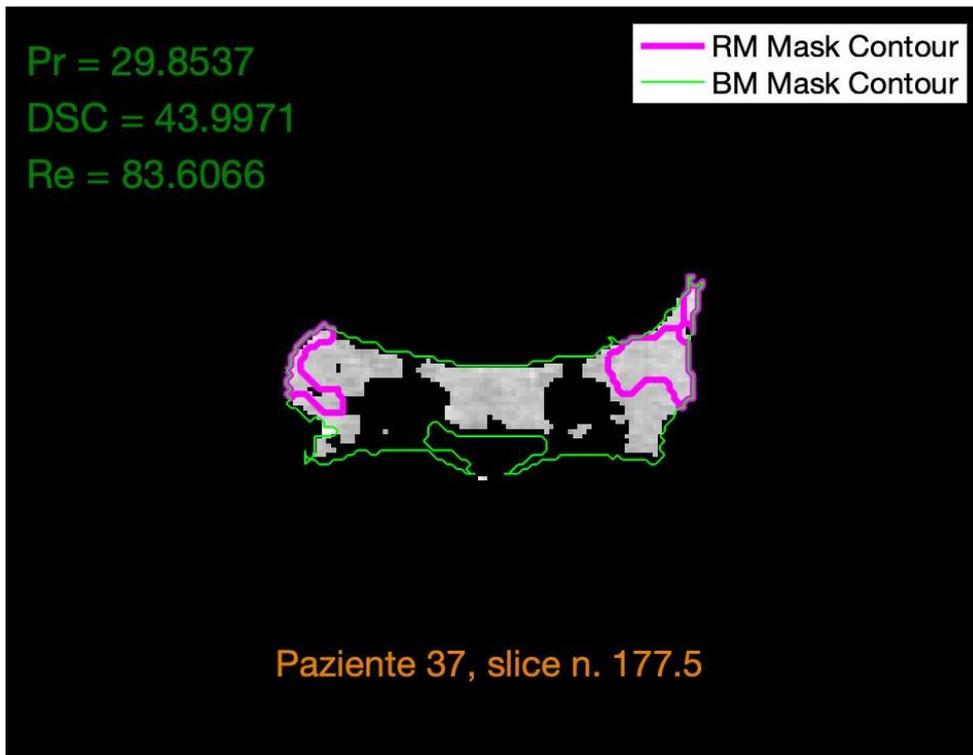


Figura 29, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

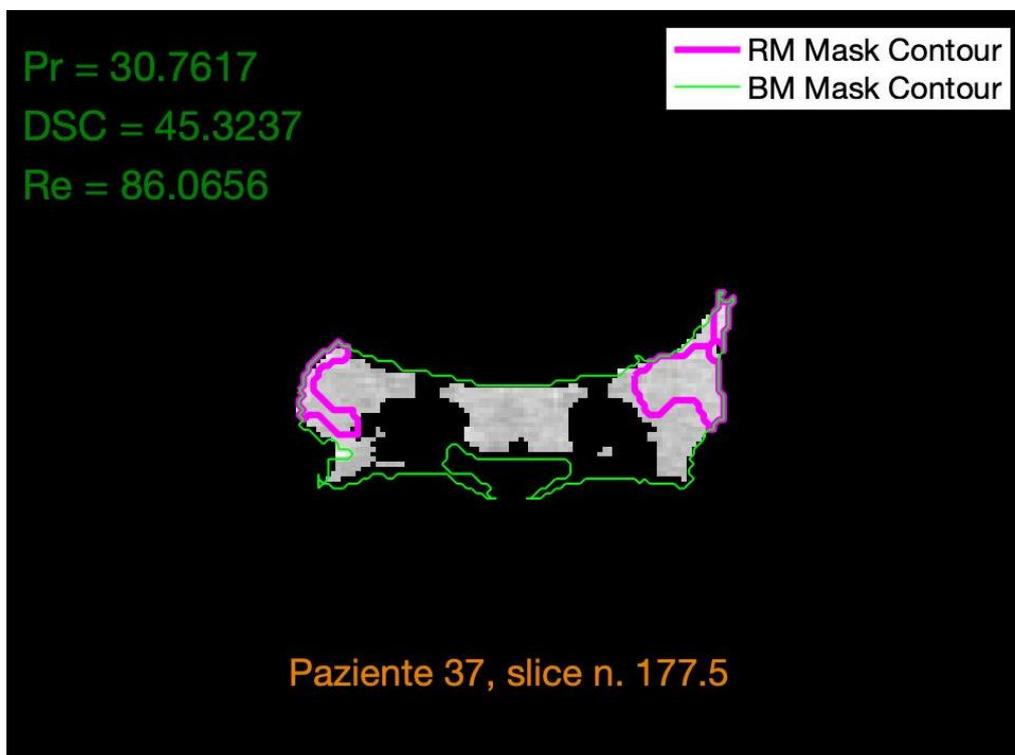


Figura 30, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

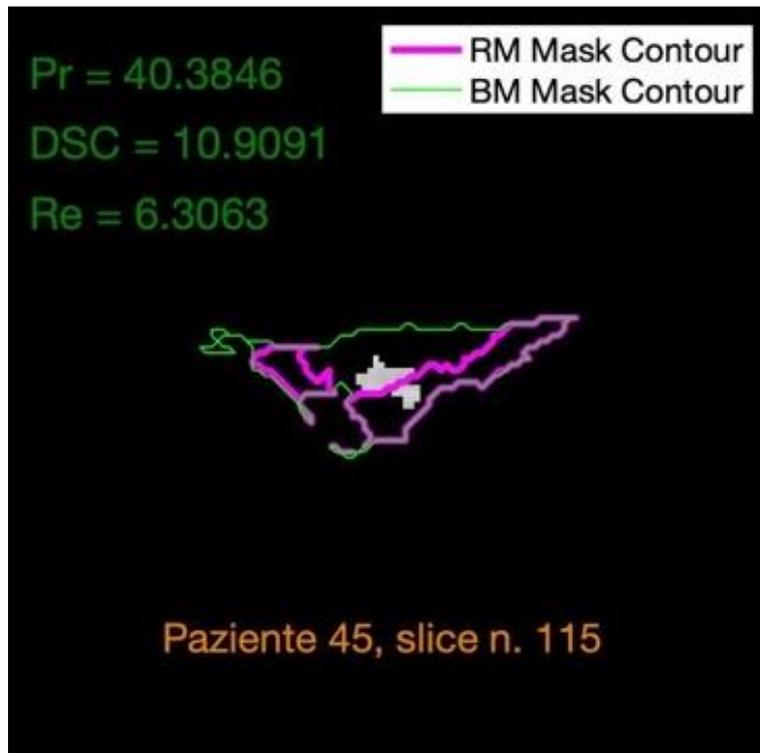


Figura 31, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

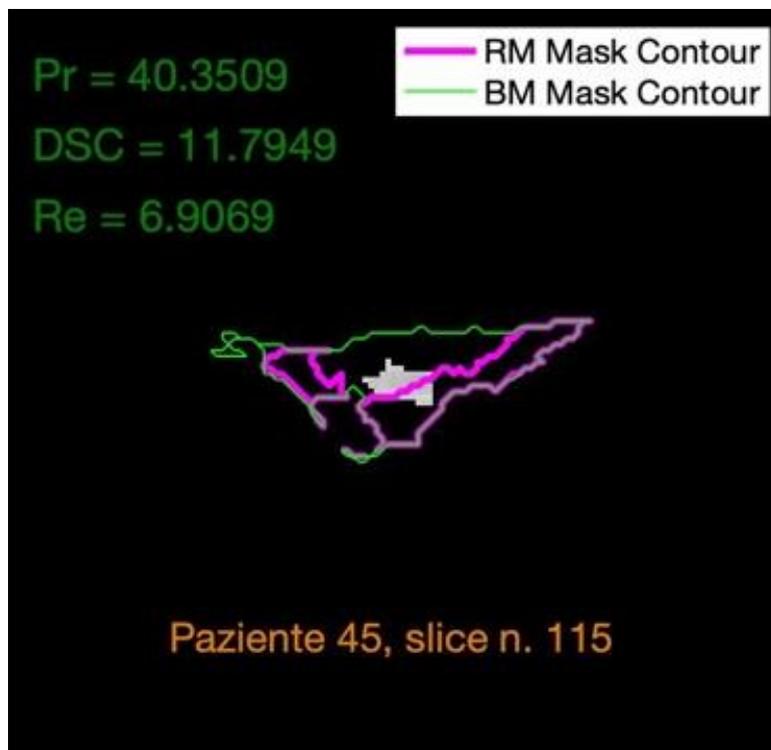


Figura 32, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

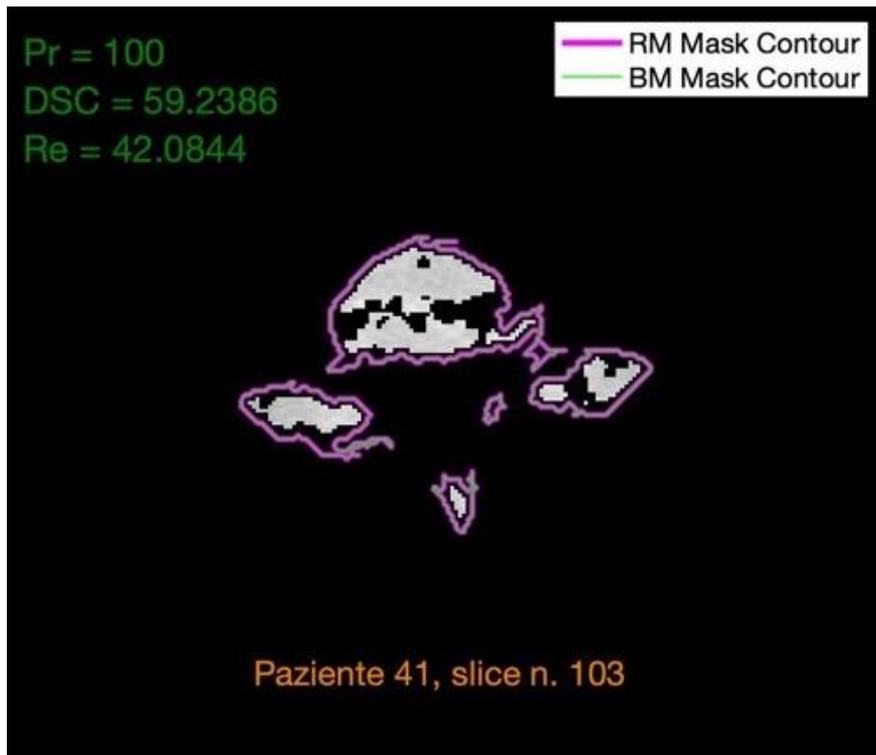


Figura 33, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

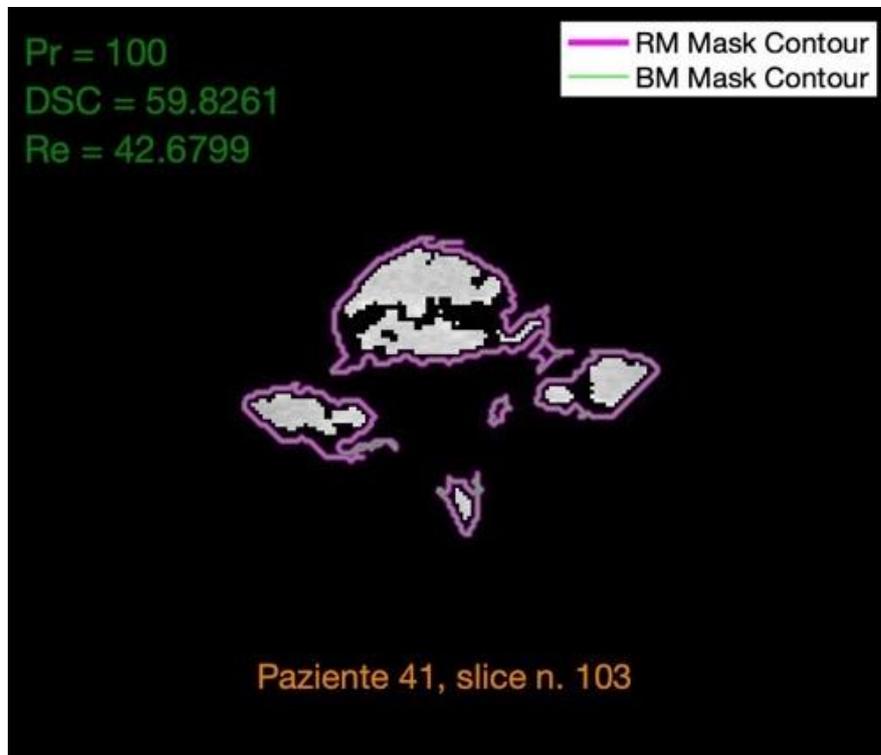


Figura 34, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

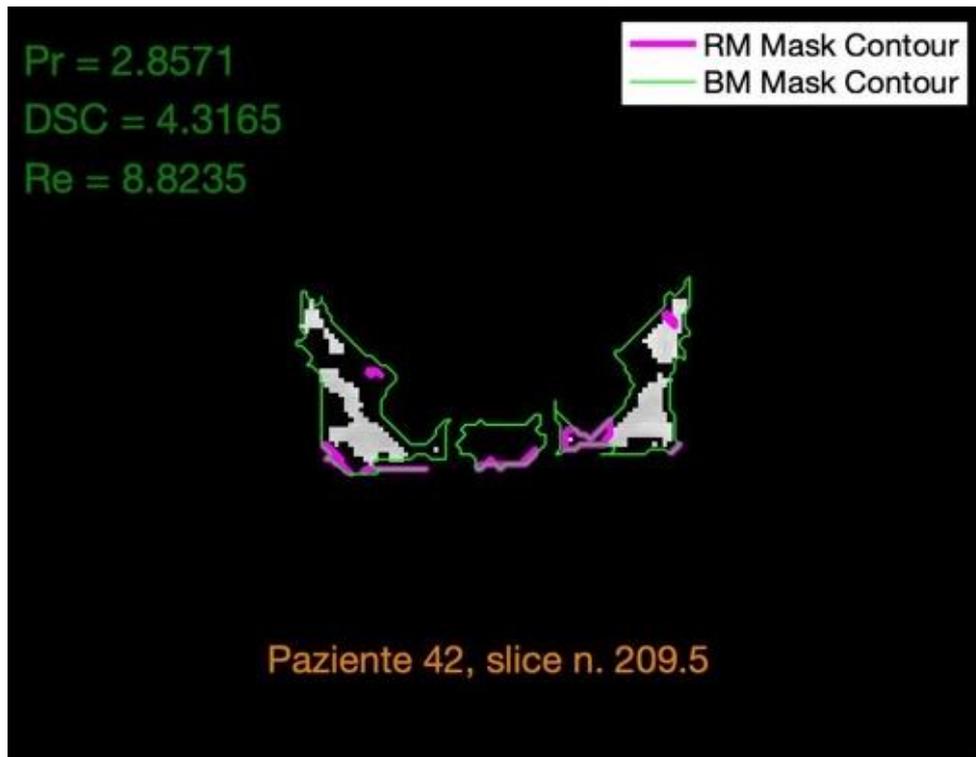


Figura 35, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

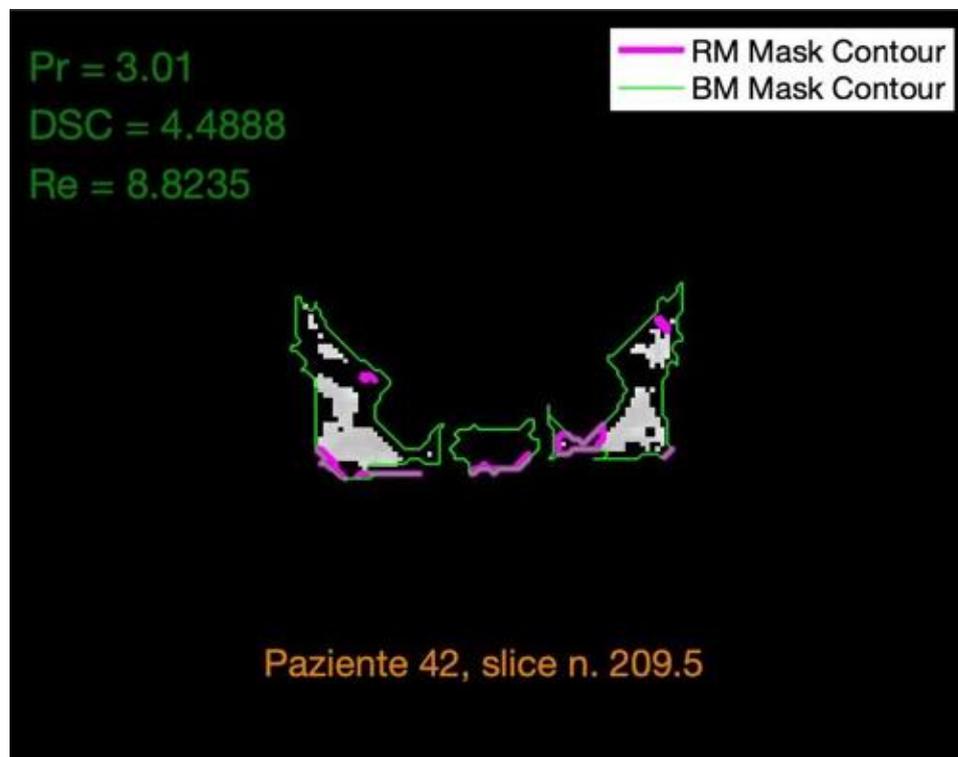


Figura 36, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

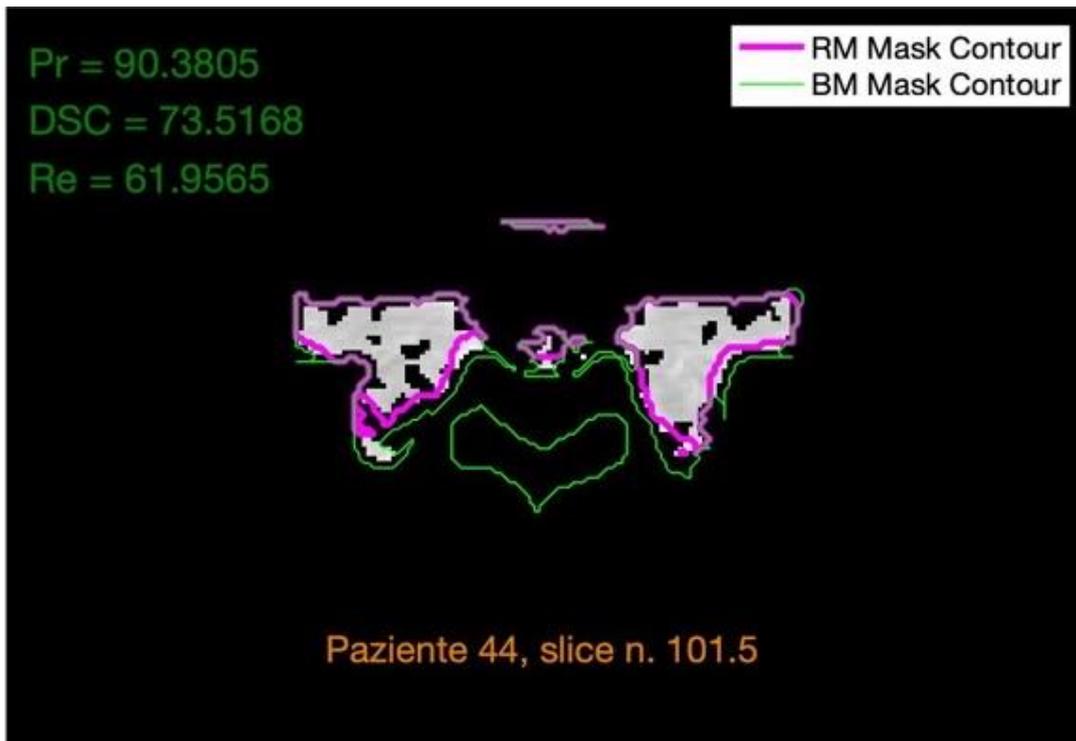


Figura 37, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

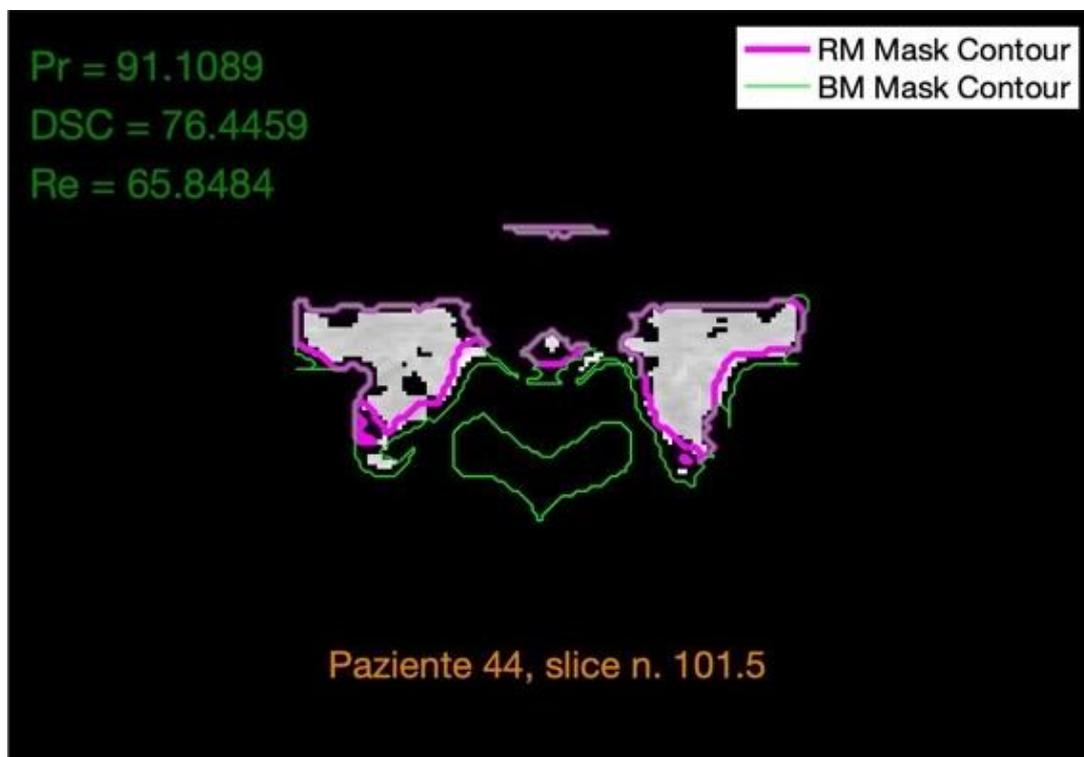


Figura 38, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

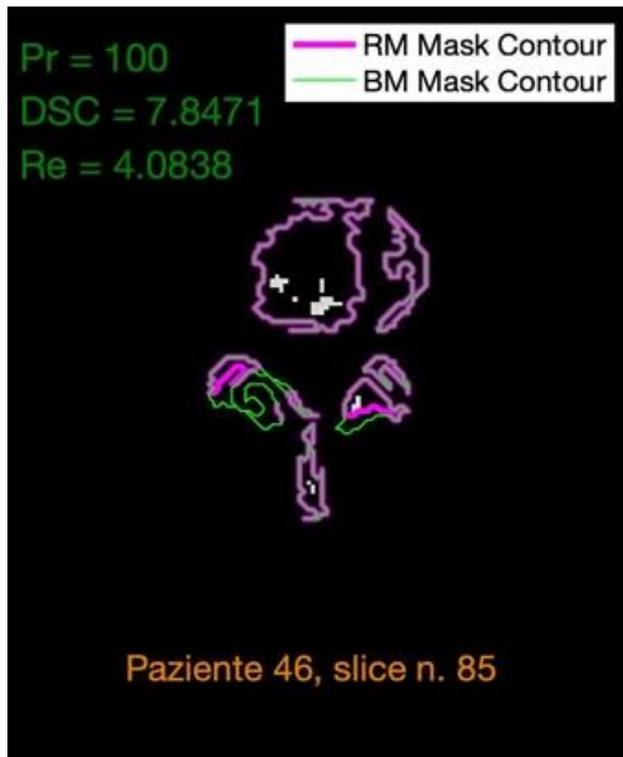


Figura 39, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

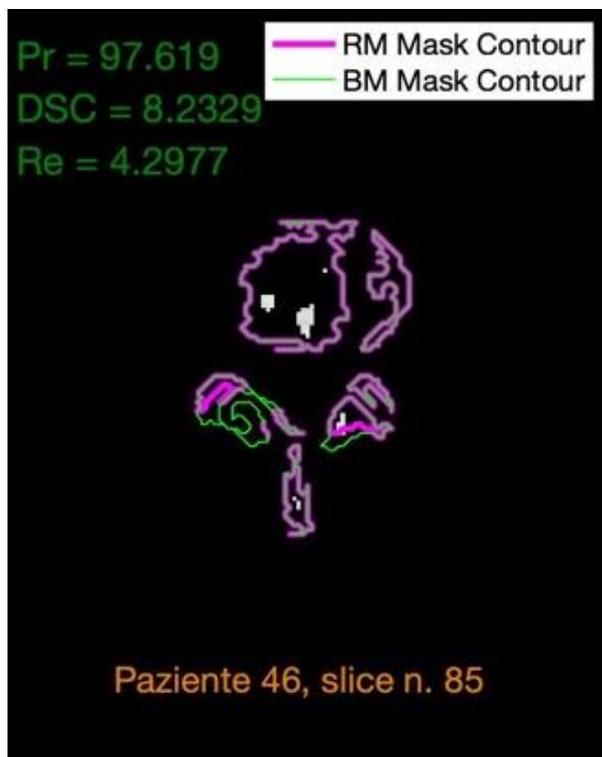


Figura 40, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

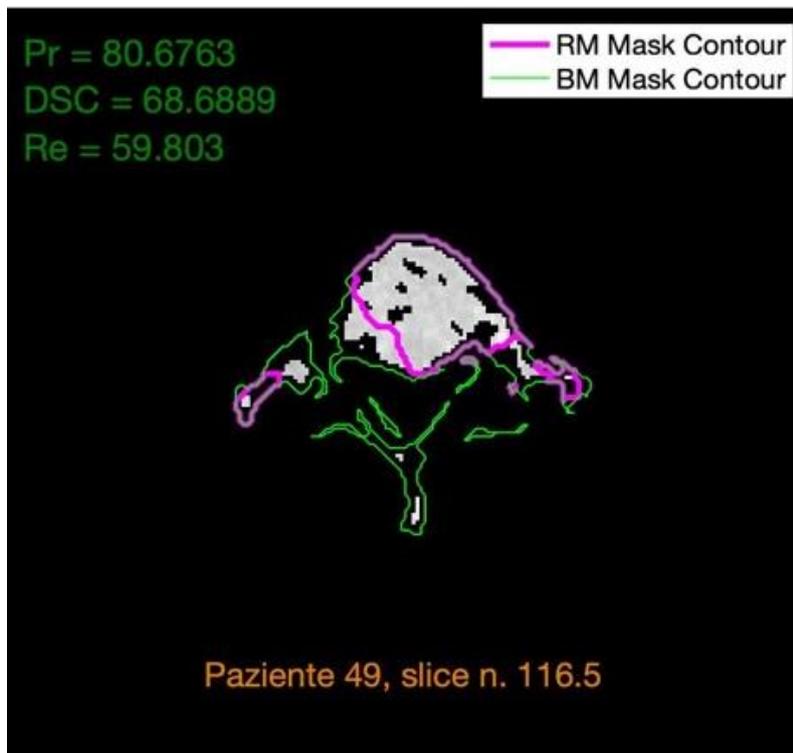


Figura 41, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

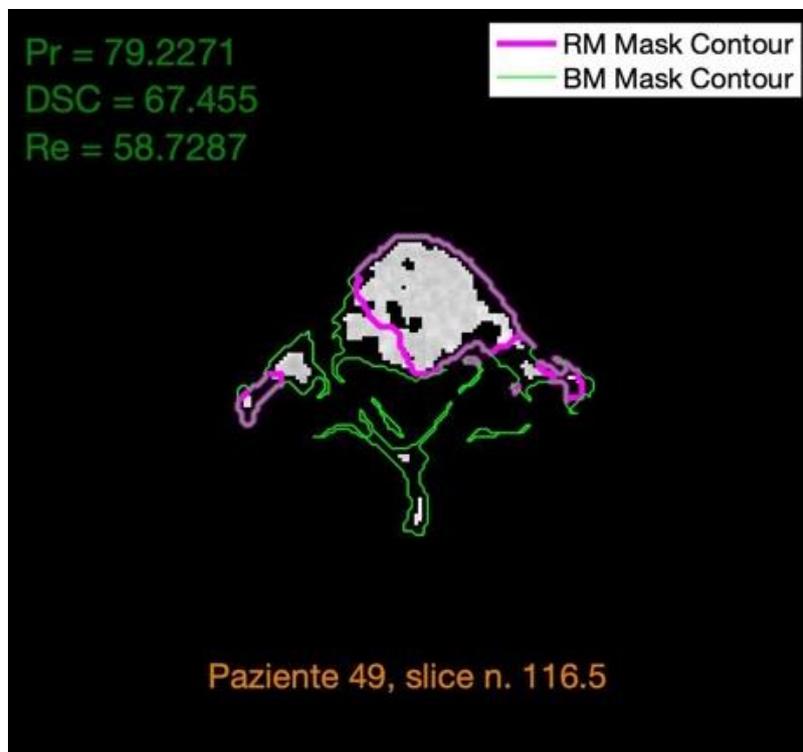


Figura 42, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

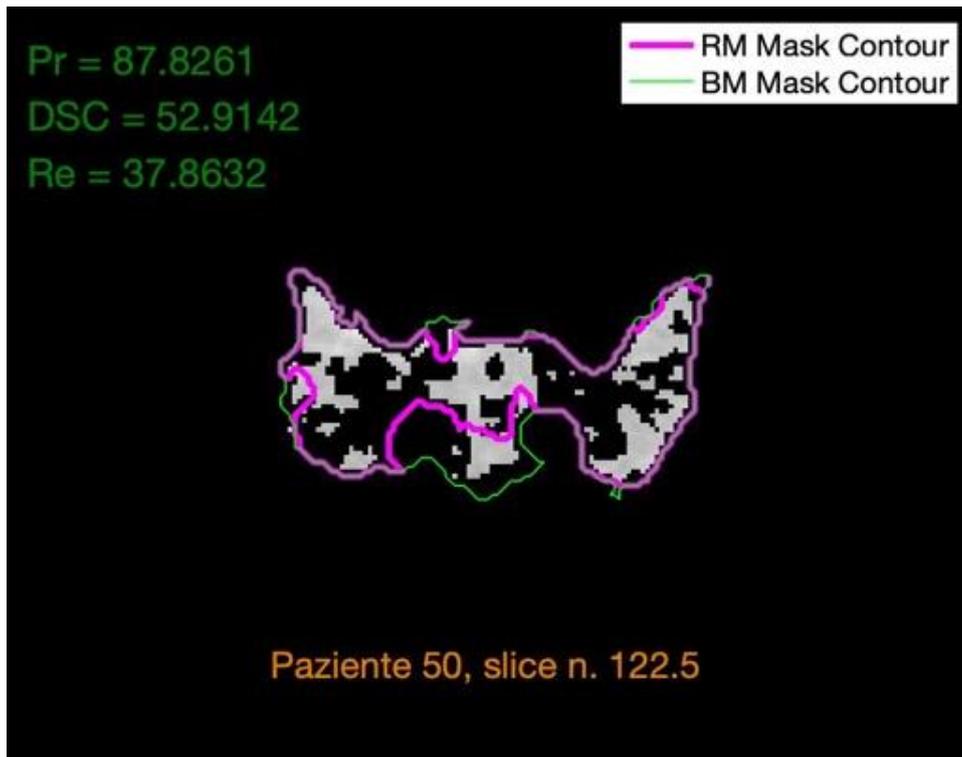


Figura 43, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

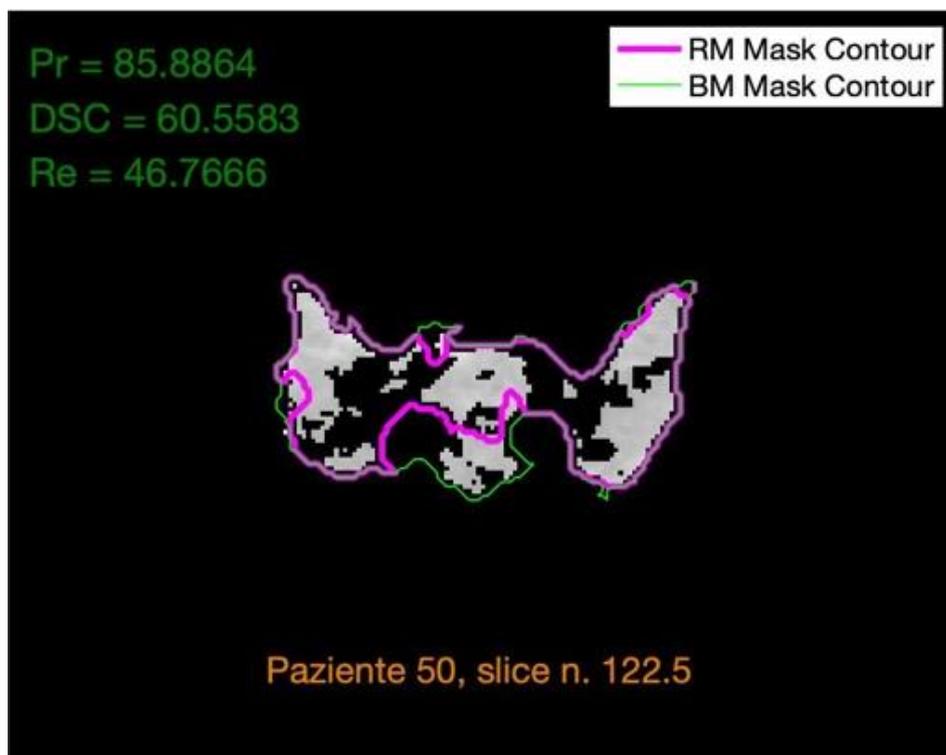


Figura 44, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

Dalle immagini precedenti emerge che le due DNN hanno un comportamento molto simile tra loro. In generale, le reti tendono a mantenere un valore di Pr molto alto, che si traduce in una classificazione corretta prevalentemente all'interno delle aree racchiuse dalle maschere reali RM. Tuttavia, vi è generalmente una sottosegmentazione all'interno delle maschere principale che si traduce visivamente nella presenza di "buchi" nelle immagini. Inoltre, la rete fatica a classificare in maniera corretta quei pixel che sono racchiusi in aree secondarie, più piccole. D'altro canto, in alcune slice (figure 29, 30), vi è una sovrasedgmentazione: vengono infatti classificati come RM pixel appartenenti in realtà all'osso compatto. Inoltre è possibile osservare come in diverse immagini prese dai pazienti non inclusi nel training set vi sia un crollo delle prestazioni (figure 39, 40, 35, 36)

Nei seguenti boxplot sono raffigurate le distribuzioni del DSC, Pr e Re per ognuno dei 50 pazienti, considerando tutte le slice a disposizione. Il lato superiore di ogni box rappresenta il 75esimo percentile, mentre quello inferiore indica il 25esimo percentile. I punti al di sopra e al di sotto dei rettangoli graficano quei valori che sono al di fuori del suddetto intervallo.

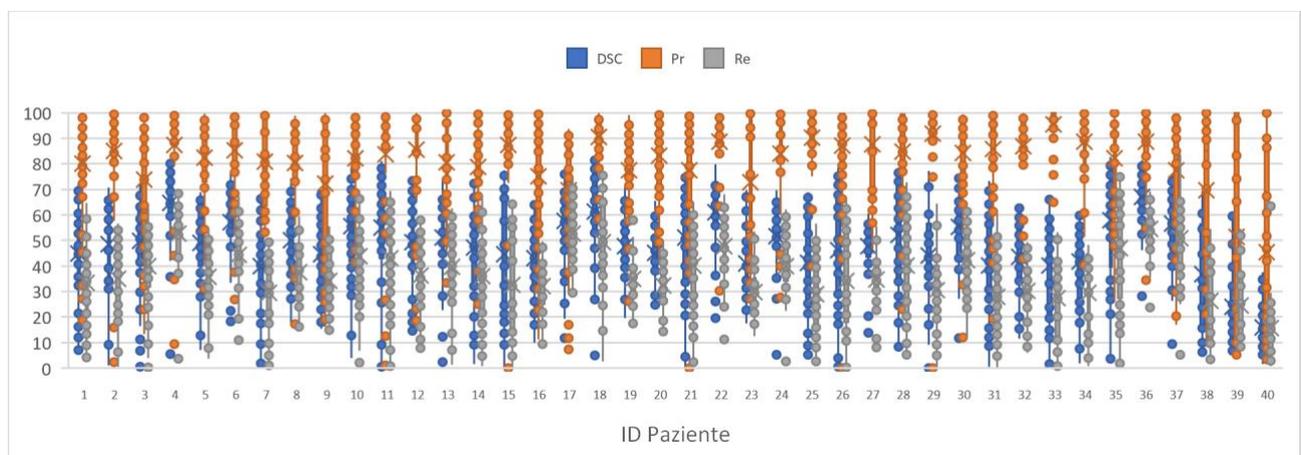


Grafico 21, boxplot prima rete, pazienti 1-40 (training set)

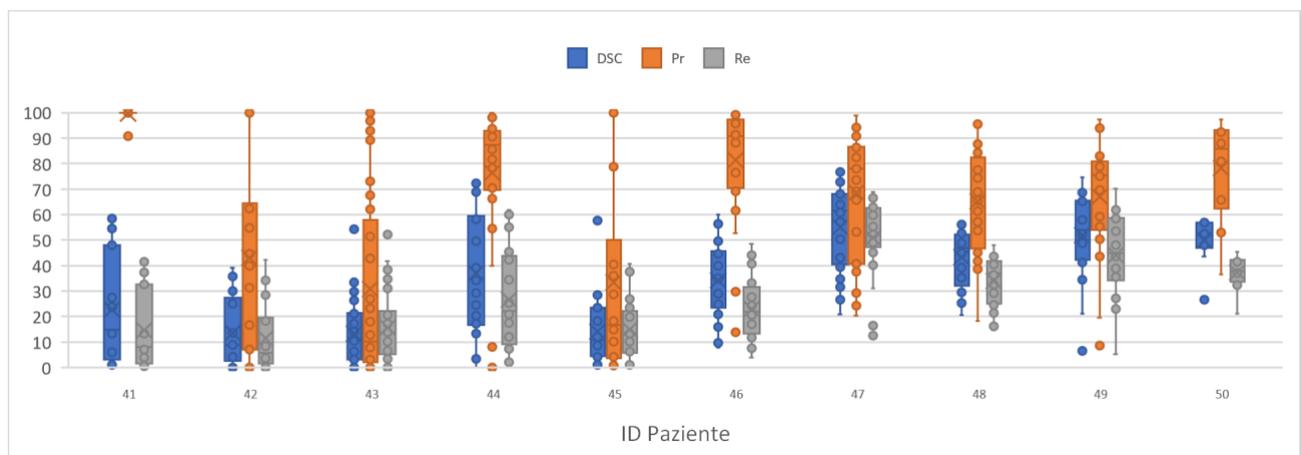


Grafico 22, boxplot prima rete, pazienti 41-50 (validation set)

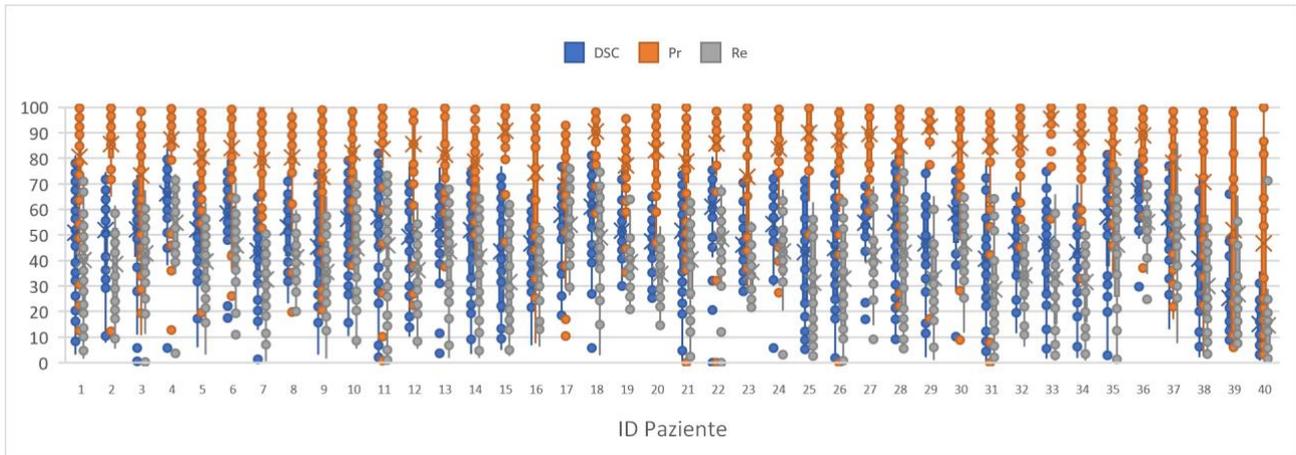


Grafico 23, boxplot seconda rete, pazienti 1-40 (training set)

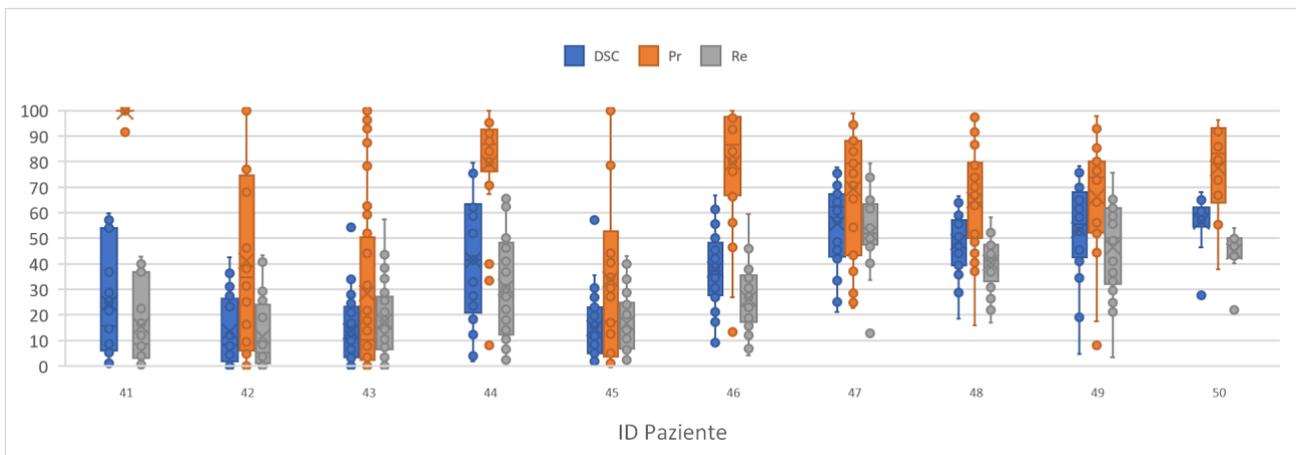


Grafico 24, boxplot seconda rete, pazienti 41-50 (validation set)

Come è possibile osservare dai boxplot, vi è un crollo della qualità di segmentazione nei pazienti non facenti parte del training set (41-50). Inoltre, anche i pazienti 39 e 40 hanno prestazioni molto più basse rispetto agli altri con entrambe le reti. Tuttavia, come era prevedibile guardando ai risultati della precision sul test set nei paragrafi precedenti, la precision riportata sui boxplot precedenti si mantiene molto elevata, e raggiunge il 100% in diverse slice. Al contrario, la recall è un parametro critico per la maggior parte dei pazienti. E' quindi emersa una problematica delle DNN adoperate: tendono a sottosegmentare il midollo osseo. Nei due grafici seguenti si è effettuato un confronto tra le prestazioni delle due reti, mostrando i valori di DSC, Pr e Re mediati su tutte le slice per ogni paziente.

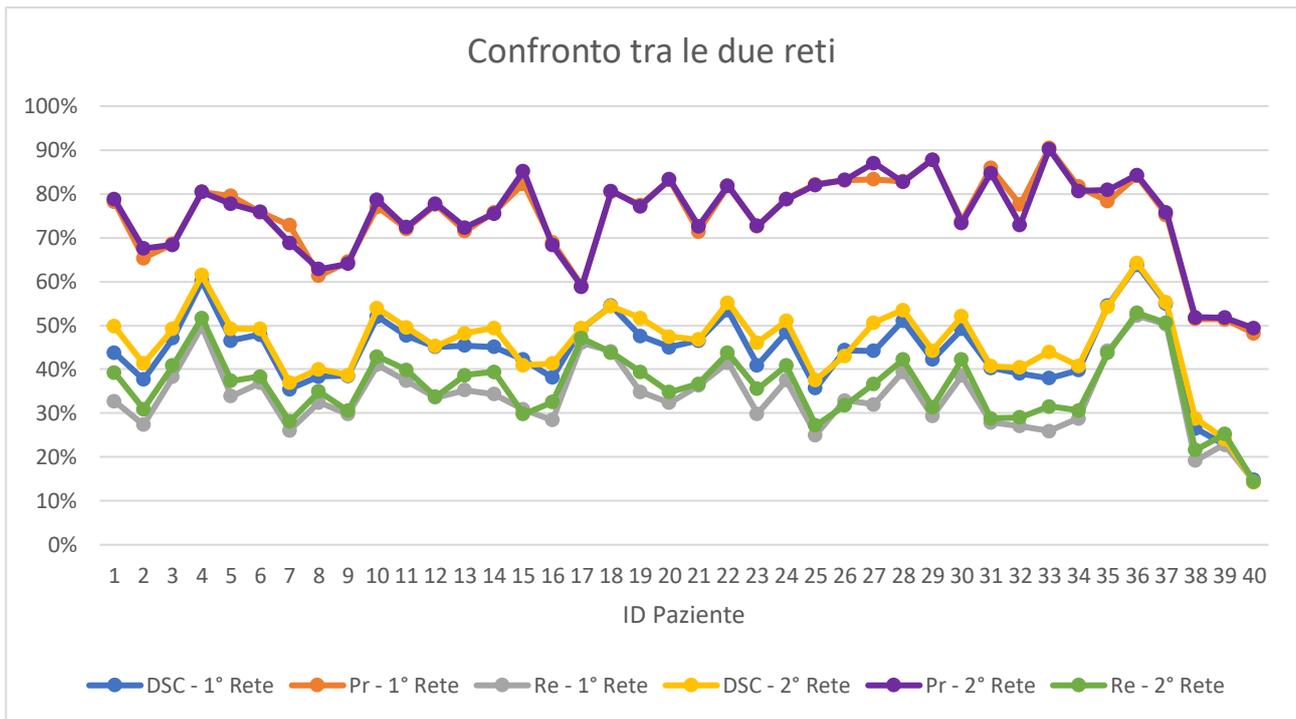


Grafico 25, Confronto tra le due reti, pazienti 1-40 (training set)

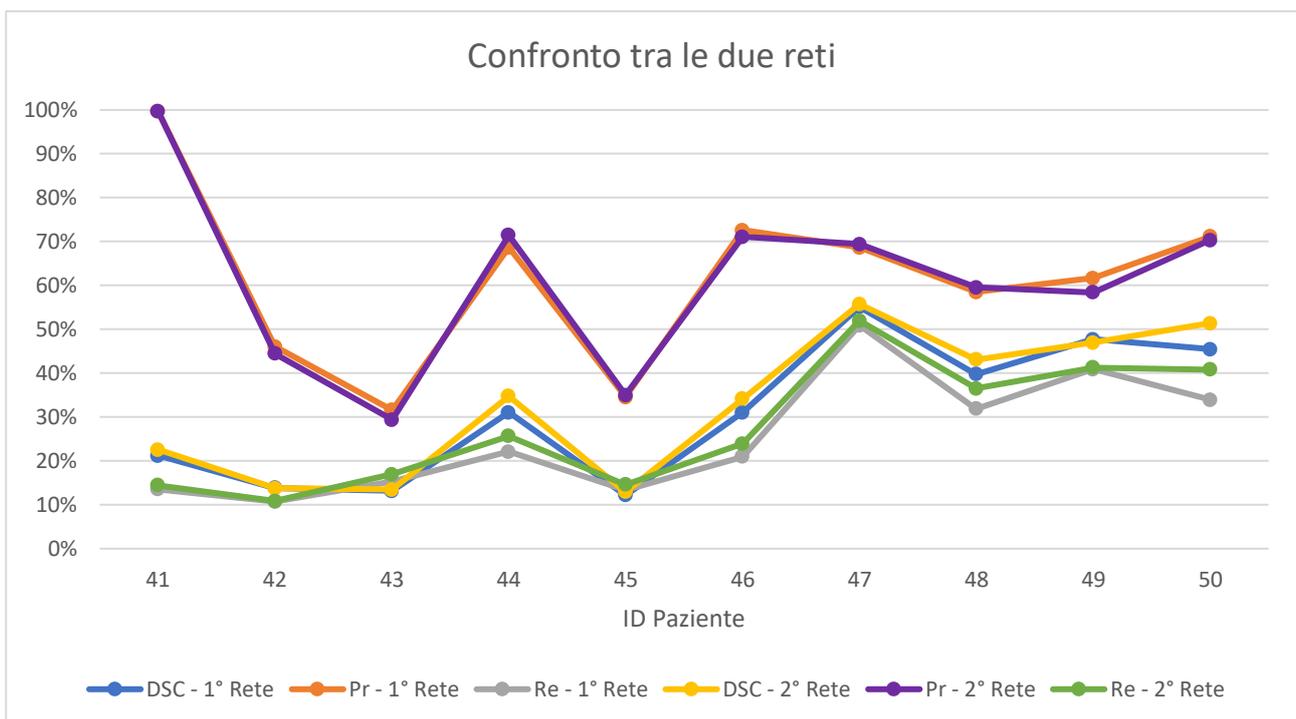


Grafico 26, Confronto tra le due reti, pazienti 41-50 (validation set)

Le due reti mostrano un andamento dei tre indici nei 50 pazienti molto simili tra loro. Tuttavia, si nota che la seconda rete ha un comportamento leggermente migliore, soprattutto in termini di Re. Anche il DSC, per diversi pazienti, è leggermente migliore nelle maschere segmentate con la seconda rete. Tuttavia il problema della sottosegmentazione permane in tutti i pazienti ed in entrambe le reti, pregiudicando pesantemente i risultati.

4. Convolutional Neural Networks (CNNs)

4.1 Introduzione alle CNN

Le Convolutional Neural Network – CNN sono dei modelli di Deep Learning, e sono ispirate ai processi biologici dei neuroni facenti parte della corteccia visiva, ovvero i neuroni coinvolti nella vista. Il principio di funzionamento alla base delle CNN è l'applicazione di filtri e convoluzioni nelle immagini allo scopo di determinare delle dipendenze spaziali e temporali nelle immagini, non visibili ad occhio nudo o attraverso semplici analisi matematiche, che vengono sfruttate dall'algoritmo per adempiere allo scopo per cui sono progettate, come la classificazione di immagini [26] [27]. Nella seguente figura (figura 45) è raffigurata una tipica struttura di CNN utilizzata per classificazione di immagini.

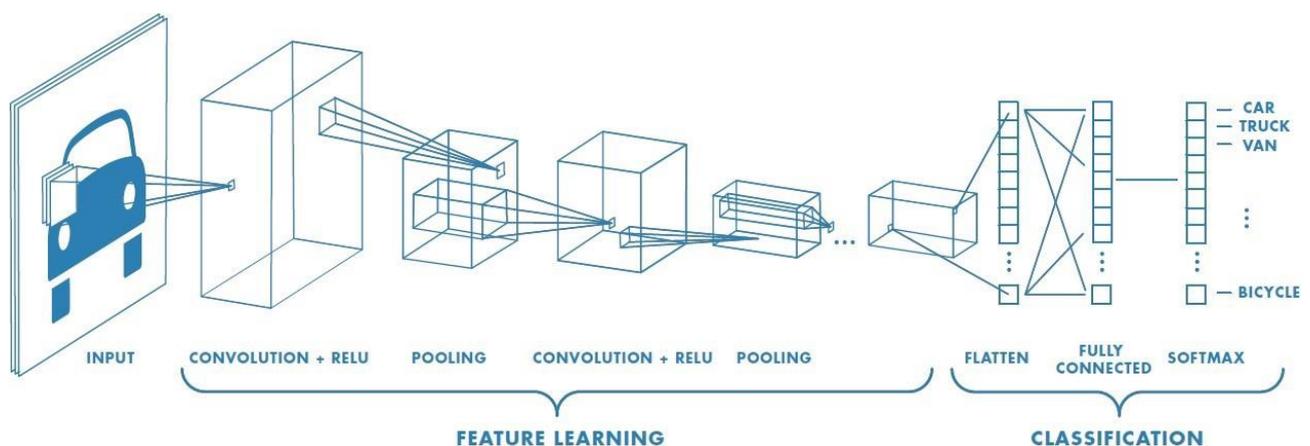


Figura 45, Tipica struttura CNN

A differenza delle DNN, in questo caso l'input è proprio l'immagine, che può essere in 3 canali come le immagini a colori RGB o, come nel caso in esame, ad un unico canale per le immagini in scala di grigi.

I blocchi solitamente utilizzati nella progettazione delle CNN sono i seguenti:

- Convolutional Layer: ha come ingresso l'immagine input se è il primo layer, oppure le attivazioni del layer precedente. Viene utilizzato per effettuare la convoluzione con una matrice di filtri, di dimensione preimpostata, allo scopo di effettuare l'estrazione delle features. E' possibile che, al momento della sovrapposizione del kernel con l'immagine, la matrice sfiori dai bordi dell'immagine; in quel caso vengono utilizzate tecniche di padding, come l'aggiunta di zeri (zero padding). E' possibile osservare questo processo nelle seguenti figure (figure 46, 47, 48).

3	1	1	2	8	4	...
1	0	7	3	2	6	...
2	3	5	1	1	3	...
1	4	1	2	6	5	...
3	2	1	3	7	2	...
9	2	6	2	5	1	...
...

Figura 46, Immagine d'esempio con filtro sovrapposto (in verde)

1	0	-1
1	0	-1
1	0	-1

Figura 47, Filtro generico applicato

-7
...
...
...
...

Figura 48, Risultato dell'applicazione del filtro

- Activation Layer: in questo layer viene elaborato il risultato della convoluzione attraverso una funzione di attivazione. Come per le DNN, vi sono diverse funzioni di attivazione che possono essere adoperate.
- Pooling Layer: effettua il processo di sotto-campionamento dell'immagine allo scopo di ottenere una riduzione dimensionale. Il risultato è una sottomatrice della mappa delle features. Il pooling può essere condotto con diverse tecniche, come il max-pooling (viene preso il valore maggiore della porzione che si vuole sottocampionare, figura 49), l'average-pooling (viene calcolata la media dell'area da ridurre), il sum-pooling (si sommano tutti gli elementi della sottomatrice)



Figura 49, Max-pooling per ridurre la dimensionalità da 4x4 a 2x2

- Flatten Layer: utilizzato per rimuovere il riferimento spaziale dei pixel, vettorizzando la struttura dei dati. È possibile vedere l'applicazione del flatten layer nella figura 50

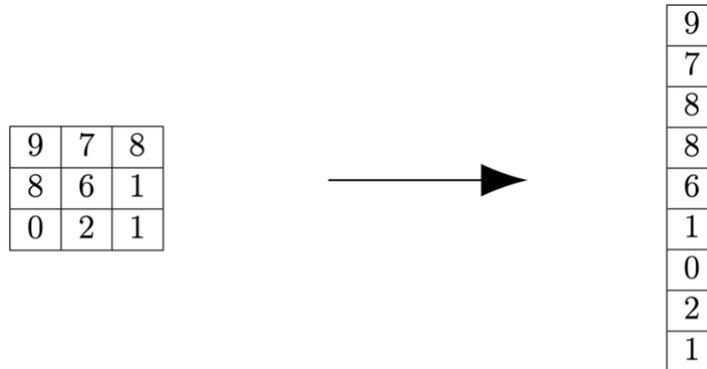


Figura 50, Applicazione Flatten Layer

- Dropout Layer: rimuove casualmente dei neuroni dalla rete durante la fase di addestramento per aumentare l'indipendenza fra i pesi.
- Batch Normalization Layer: viene generalmente posto dopo l'Activation Layer, permette di normalizzare le mappe delle attivazioni. In particolare, la normalizzazione è ottenuta sottraendo il valore medio e dividendo per la deviazione standard per ogni epoca di allenamento.
- Fully Connected Layer (o Dense Layer): ha in ingresso il risultato dell'analisi delle features. Applica i pesi per predire la classe di appartenenza dell'input.
- SoftMaxLayer: applica la funzione softmax all'input. La funzione softmax, detta anche funzione esponenziale normalizzata, è un operatore che comprime un vettore di valori reali in un vettore della stessa dimensione, ma con valori compresi in un intervallo (0,1) e la cui somma è 1. La funzione è così descritta (equazione 12):

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad \text{per } j = 1, \dots, K \quad (12)$$

Dove K è la dimensione del vettore, ed è uguale al numero di classi da predire.

- Classification Layer: genera l'output finale della rete.

4.2 Tuning dei parametri

La struttura delle CNN addestrate è la seguente:

- ImageInputLayer di dimensione 5x5x1, che accetta in ingresso le ROI 5x5.
- Convolutional Layer con kernel 3x3
- Batch Normalization Layer
- Funzione di attivazione

I tre blocchi precedenti sono stati ripetuti un numero diverso di volte, valutando l'influenza sul risultato finale del numero di convoluzioni

- Fully Connected Layer
- Softmax Layer
- Classification Layer

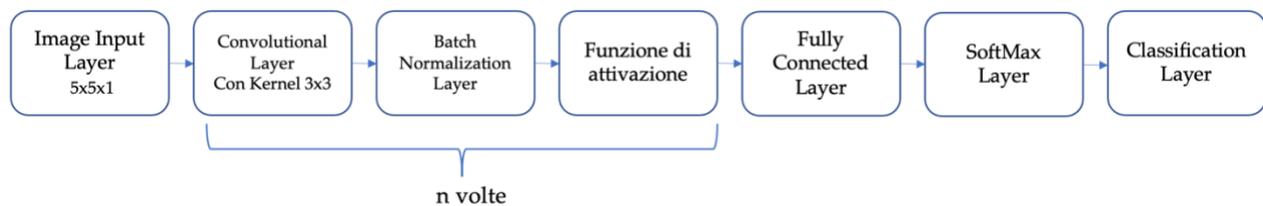


Figura 51, Struttura CNN utilizzata

Dovendo analizzare immagini molto piccole (5x5), non sono stati utilizzati layer di pooling per ridurre ulteriormente la dimensionalità. Anche per le CNN si è adoperato un addestramento supervisionato, e la loss-function scelta è nuovamente la cross-entropy (equazione 3.x). Posto in considerazione che, dalle analisi effettuate per la DNN nel capitolo precedente, il training set ottenuto tramite clustering aveva restituito risultati sensibilmente migliori, tutte le CNN sono state addestrate con questo training set. Analogamente a quanto fatto per la DNN, anche per la CNN si è effettuato un tuning dei parametri per trovare la rete ideale ai fini della segmentazione del midollo osseo attivo. Si sono svolti anche in questo caso 20 addestramenti per ogni combinazione di settaggi. Come accennato precedentemente, uno dei parametri su cui è stata posta l'attenzione è stato il numero di blocchi di convoluzione: sono state addestrate reti con 2 e 3 blocchi di convoluzione. Inoltre, si sono testate due funzioni di attivazione diverse, ovvero la *ReLU* (equazione 13) e la *tanh* (equazione 14). Infine si è provato ad aumentare iterativamente il numero di epoche di addestramento. Si è partiti da 30, successivamente si è aumentato a 40 fino ad arrivare a 50 epoche. Dopo aver constatato che i risultati migliori sono stati ottenuti con 50 epoche di addestramento e 3 blocchi di convoluzione (capitolo seguente), si sono svolte nuove prove aumentando le epoche di addestramento a 60 e 70, e con un blocco di convoluzione in più, arrivando a 4 in totale.

$$ReLU: \sigma(x) = \max(0, x) \quad (13)$$

$$tanh: \sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (14)$$

4.3 Risultati

Nei grafici seguenti sono riportati i risultati delle CNN al variare dei parametri descritti nel paragrafo precedente.

30 epoche, 2 blocchi di convoluzione:

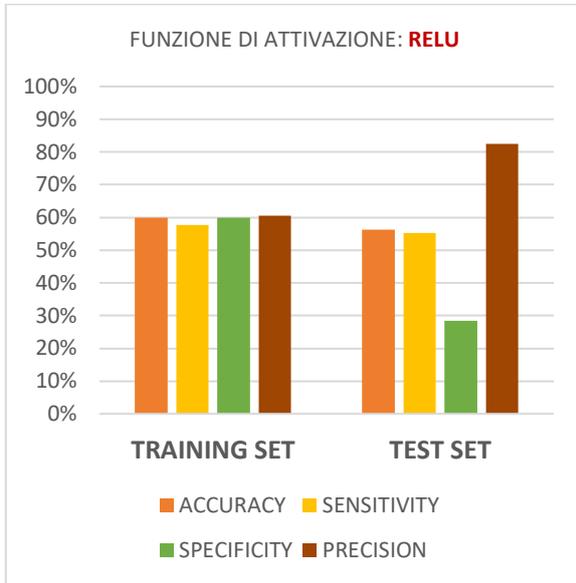


Grafico 27, Risultati CNN, 30 epoche, 2 convoluzioni, ReLu

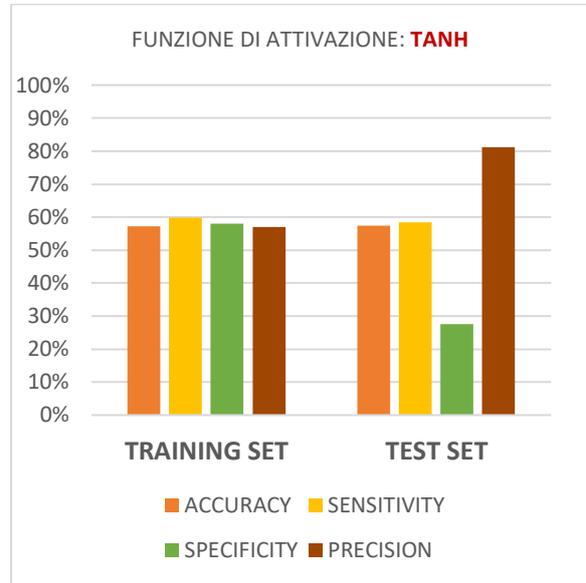


Grafico 28, Risultati CNN, 30 epoche, 2 convoluzioni, Tanh

30 epoche, 3 blocchi di convoluzione:

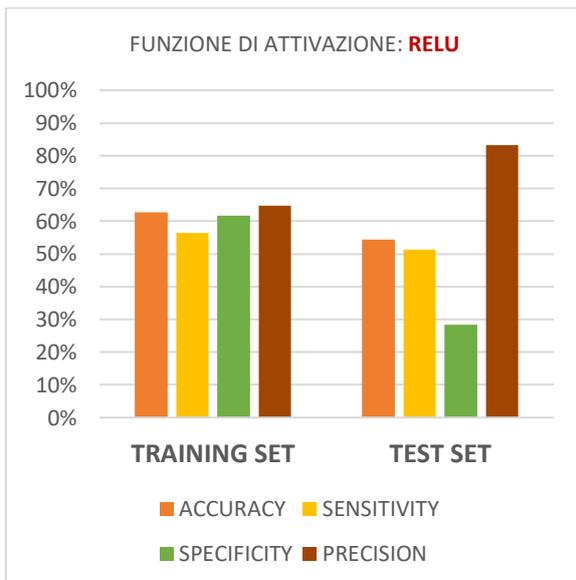


Grafico 29, Risultati CNN, 30 epoche, 3 convoluzioni, ReLu

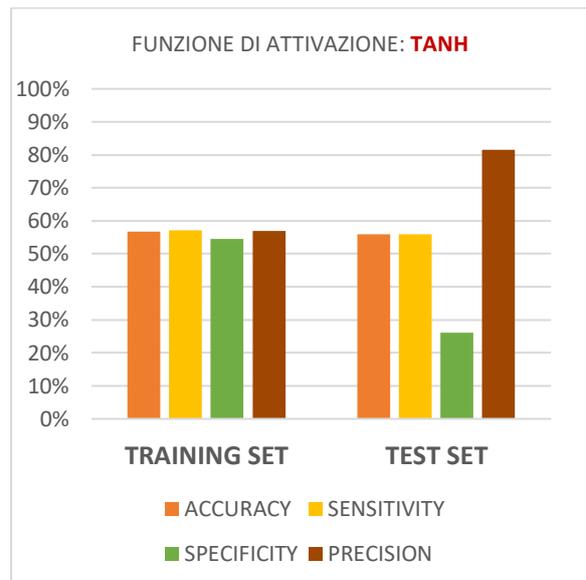


Grafico 30, Risultati CNN, 30 epoche, 3 convoluzioni, Tanh

40 epoche, 2 blocchi di convoluzione:

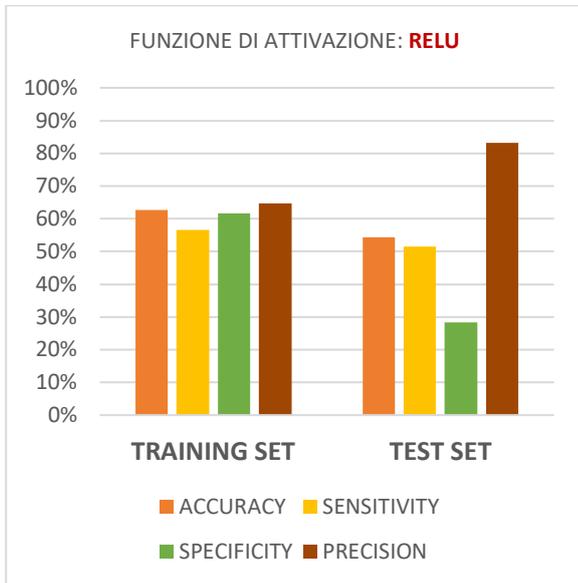


Grafico 31, Risultati CNN, 40 epoche, 2 convoluzioni, ReLu

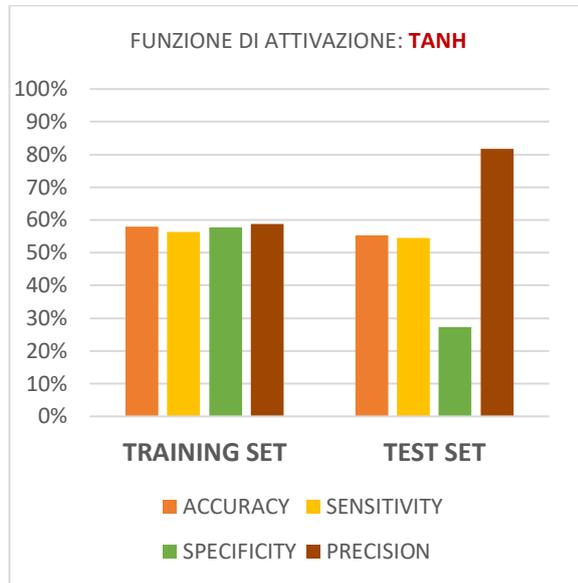


Grafico 32, Risultati CNN, 40 epoche, 2 convoluzioni, Tanh

40 epoche, 3 blocchi di convoluzione:

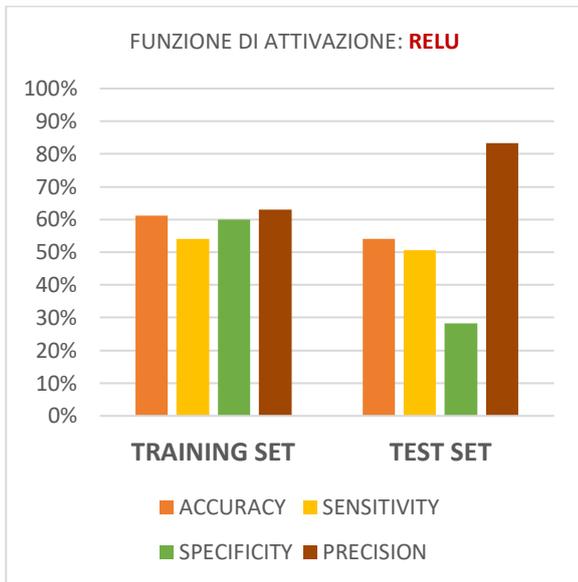


Grafico 33, Risultati CNN, 40 epoche, 3 convoluzioni, ReLu

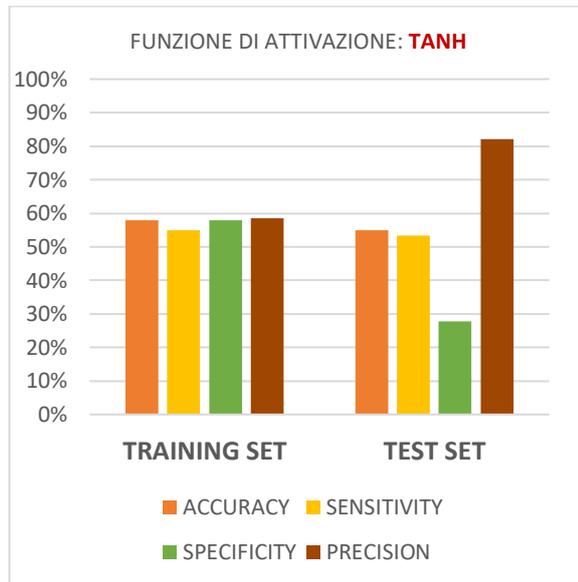


Grafico 34, Risultati CNN, 40 epoche, 3 convoluzioni, Tanh

50 epoche, 2 blocchi di convoluzione:

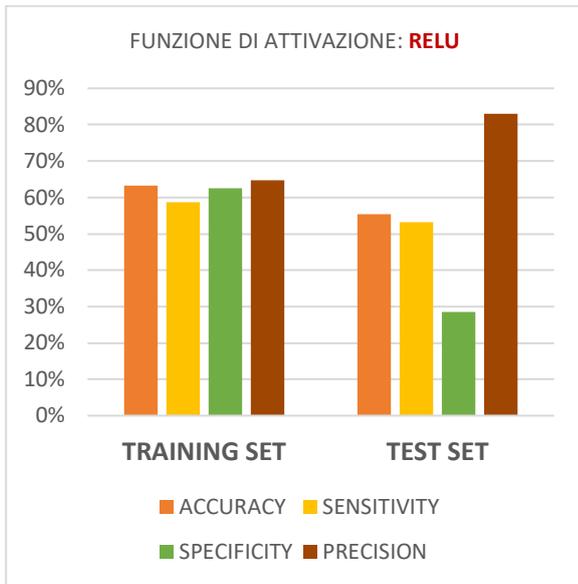


Grafico 35, Risultati CNN, 50 epoche, 2 convoluzioni, ReLu

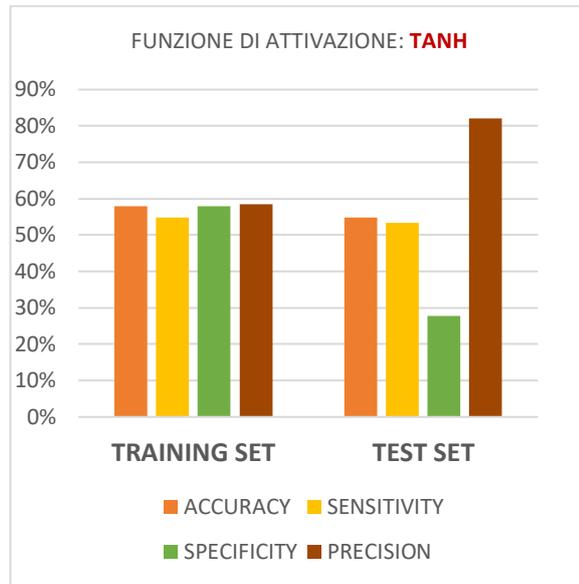


Grafico 36, Risultati CNN, 50 epoche, 2 convoluzioni, tanh

50 epoche, 3 blocchi di convoluzione:

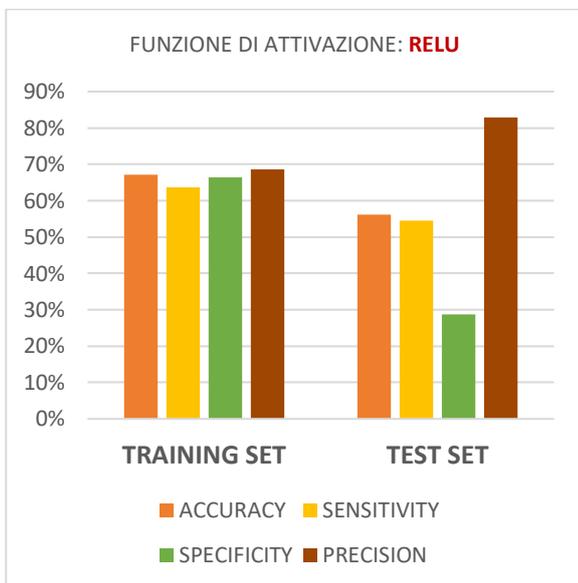


Grafico 37, Risultati CNN, 50 epoche, 3 convoluzioni, ReLu

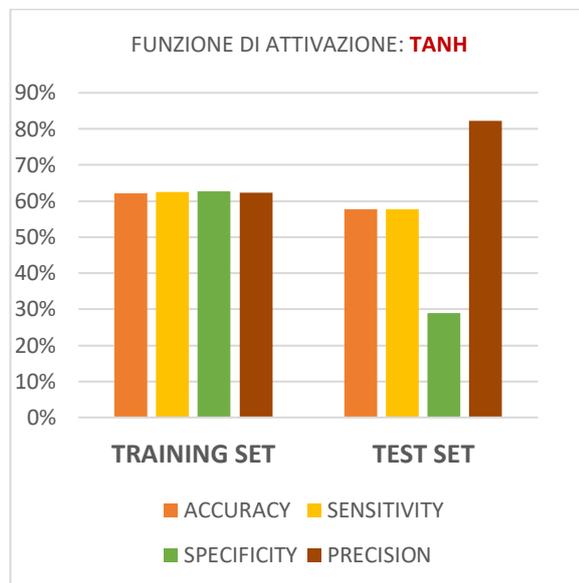


Grafico 38, Risultati CNN, 50 epoche, 3 convoluzioni, Tanh

50 epoche, 4 blocchi di convoluzione:

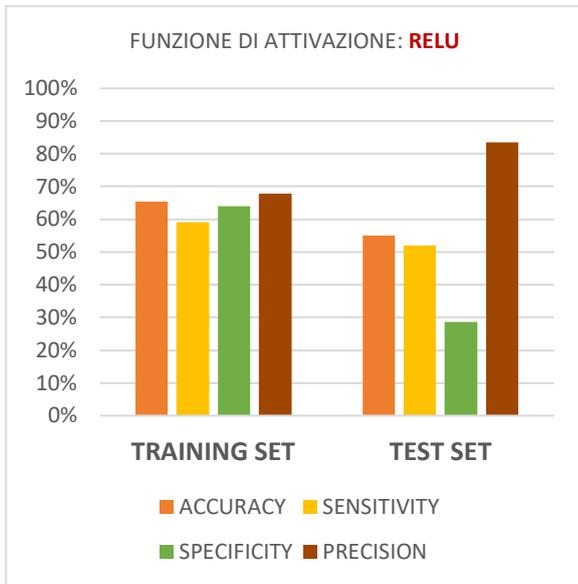


Grafico 39, Risultati CNN, 50 epoche, 4 convoluzioni, ReLu

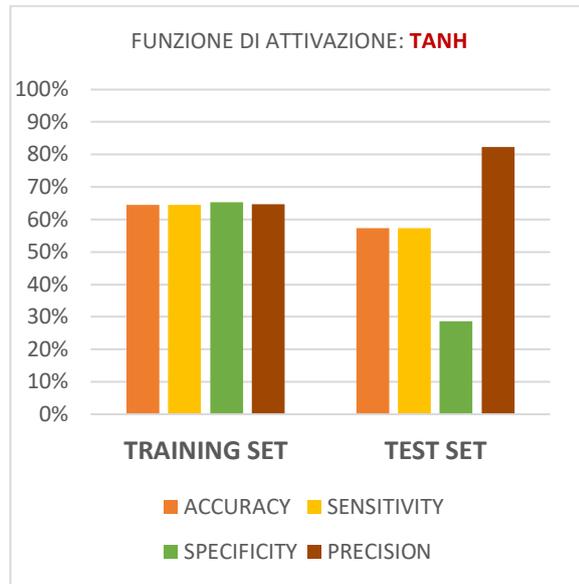


Grafico 40, Risultati CNN, 50 epoche, 4 convoluzioni, Tanh

60 epoche, 3 blocchi di convoluzione:

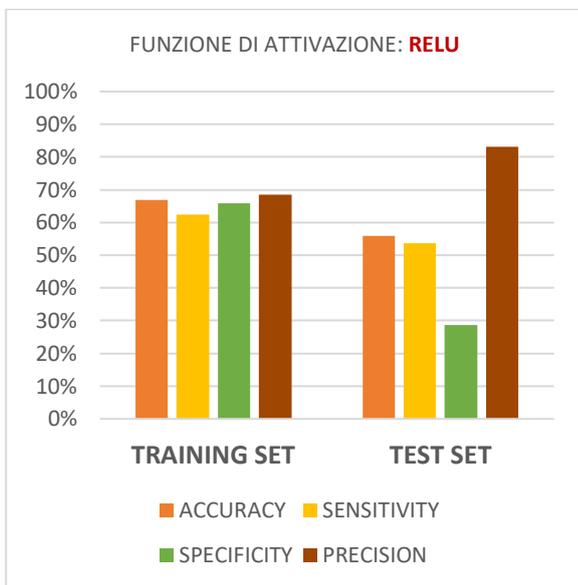


Grafico 41, Risultati CNN, 60 epoche, 3 convoluzioni, ReLu

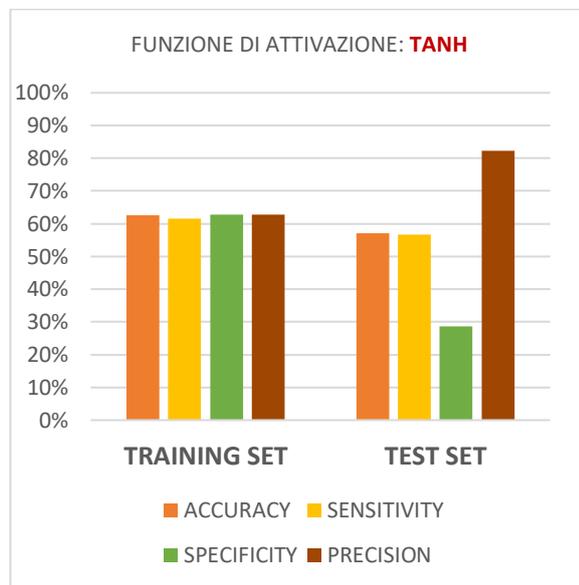


Grafico 42, Risultati CNN, 60 epoche, 3 convoluzioni, Tanh

60 epoche, 4 blocchi di convoluzione:

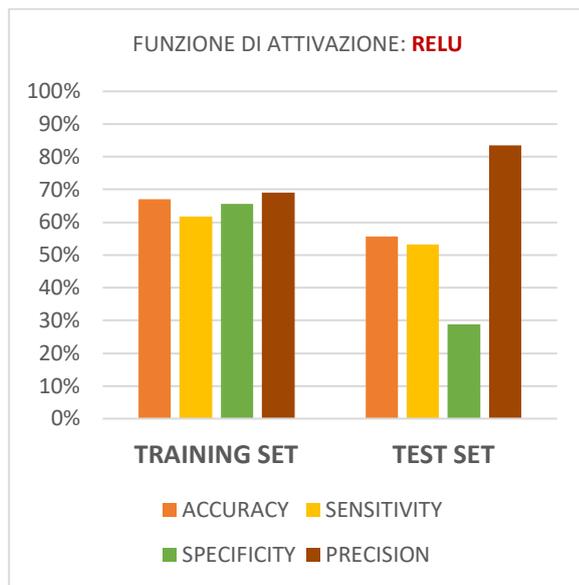


Grafico 43, Risultati CNN, 60 epoche, 4 convoluzioni, ReLu

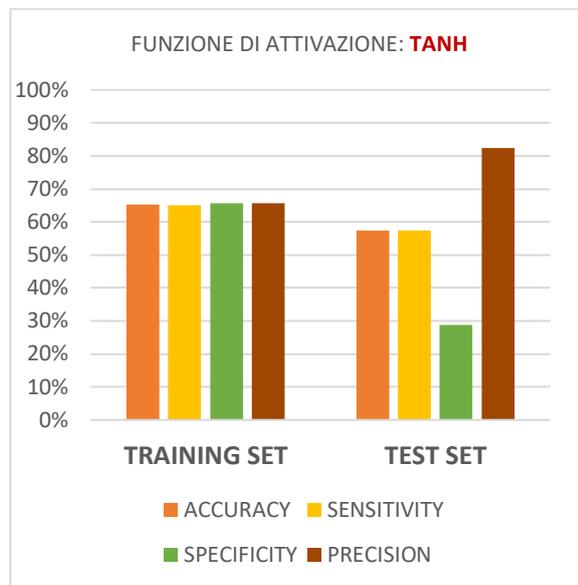


Grafico 44, Risultati CNN, 60 epoche, 4 convoluzioni, Tanh

70 epoche, 3 blocchi di convoluzione:

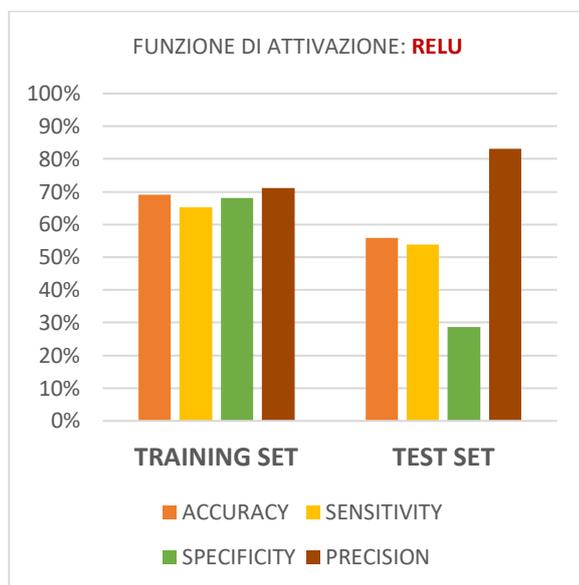


Grafico 45, Risultati CNN, 70 epoche, 3 convoluzioni, ReLu

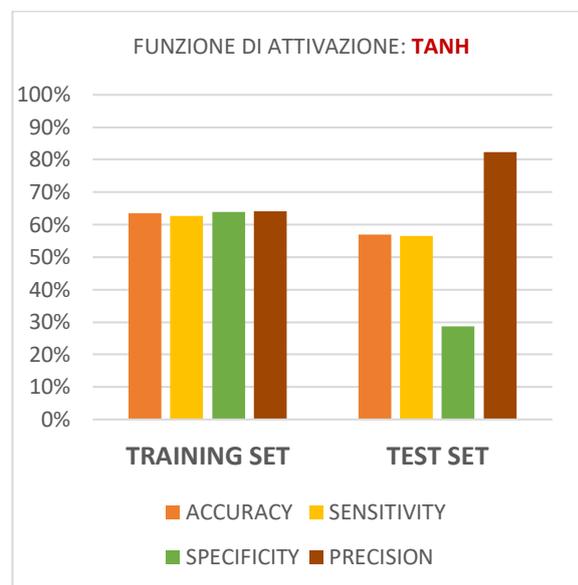


Grafico 46, Risultati CNN, 70 epoche, 3 convoluzioni, Tanh

70 epoche, 4 blocchi di convoluzione:

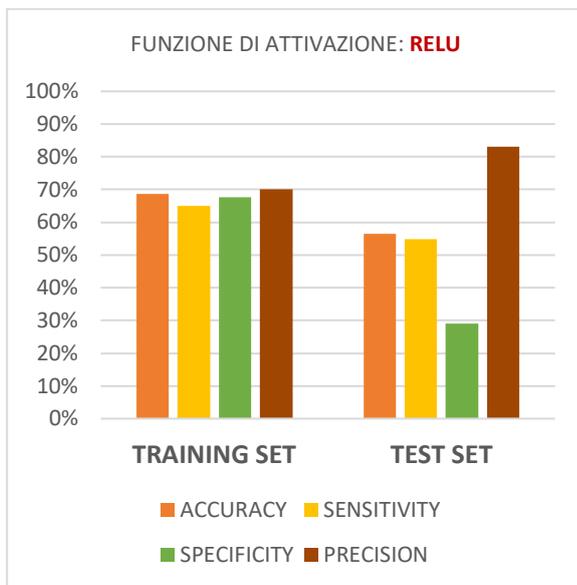


Grafico 47, Risultati CNN, 70 epoche, 4 convoluzioni, ReLu

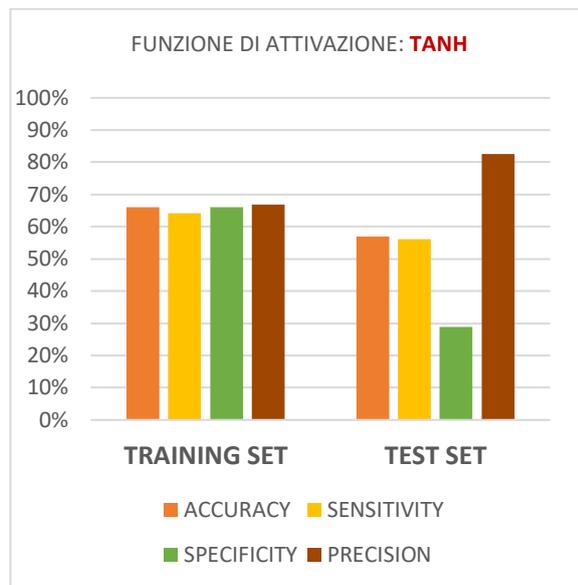


Grafico 48, Risultati CNN, 70 epoche, 4 convoluzioni, Tanh

Dai risultati appena riportati è emerso che la funzione di attivazione ReLu sembra ottenere risultati migliori nel training set rispetto all'altra funzione di attivazione, la tanh. Nel valutare le prestazioni sul test set, invece, le due funzioni hanno un comportamento molto simile e, in alcuni casi, la funzione *tanh* restituisce valori di accuracy leggermente più alti di quelli ottenuti dalle reti con funzione di attivazione *ReLu*. Inoltre, si nota in maniera evidente come i quattro indici passino dall'essere bilanciati tra loro nel training set ad essere sproporzionati nel test set, soprattutto in termini di specificity e precision. Infatti la precision si mantiene sempre molto elevata, al di sopra dell'80%, analogamente a come accaduto per le DNN, mentre la specificity subisce un crollo, fino a raggiungere valori inferiori al 30%. Il numero di blocchi di convoluzione che ha portato a risultati migliori è tre. Si era partiti analizzando reti con due e tre blocchi e, constatando che con tre i risultati erano leggermente migliori, si è provato ad aggiungerne un quarto. Tuttavia, questo non ha giovato alla rete che non ha mostrato miglioramenti nelle performance, mostrando un'accuracy uguale (intorno al 57%). Per quanto riguarda il numero di epoche di addestramento, è possibile osservare un miglioramento delle prestazioni nel passaggio da 30 a 50 epoche, con un aumento nell'accuracy dal 60% al 67% (grafici 27 e 37). Quando queste vengono aumentate a 60 e 70, in alcuni casi, si ha ovviamente un miglioramento nel training set, seguito però da un peggioramento nel test set. Ciò è dovuto al fenomeno dell'overfitting. Quest'ultimo avviene quando la rete si specializza nel riconoscere gli input facenti parte del training set, ma quando le viene sottoposto un dato non presente nel training set la rete presenta difficoltà nel classificarlo. Successivamente, si è scelta una rete per l'ottenimento delle maschere tramite CNN. Per fare ciò, si sono analizzati i risultati di tutte le 20 reti che si erano addestrate con ogni combinazione di parametri. Da questa valutazione è stata scelta una rete addestrata con 60 epoche, con 3 blocchi di convoluzione e con funzione di attivazione *tanh*, in quanto ha riportato i seguenti valori nei 4 indici nel test set: 68,26% accuracy; 78,82% sensitivity; 31,18% specificity; 79,82% precision.

4.4 Creazione maschere tramite CNN e risultati della segmentazione

Con la rete CNN descritta nel paragrafo precedente, sono state classificate le ROI di tutti i 50 pazienti per ottenere le maschere RM. Analogamente a quanto svolto nel paragrafo 3.7, sono state confrontate le maschere ottenute attraverso la predizione della rete con le maschere RM reali, valutandole attraverso gli stessi tre indici: DSC, Pr e Re. Anche in questo caso, il bordo verde rappresenta l'osso compatto, mentre il bordo viola raffigura il contorno della maschera RM reale. L'immagine in scala di grigi è ottenuta dalla moltiplicazione tra l'immagine CT originale e la maschera RM predetta dalla CNN. Di seguito è possibile osservare degli esempi.

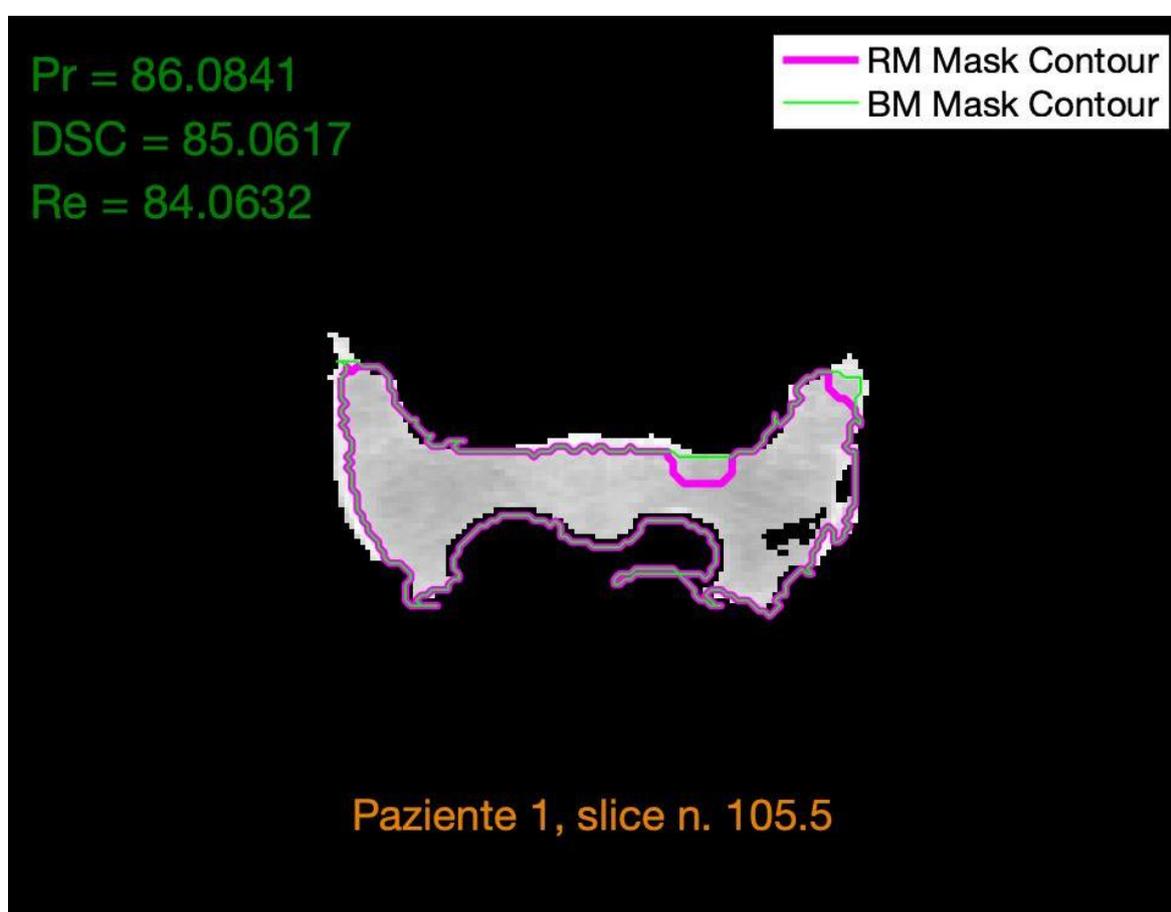


Figura 52, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

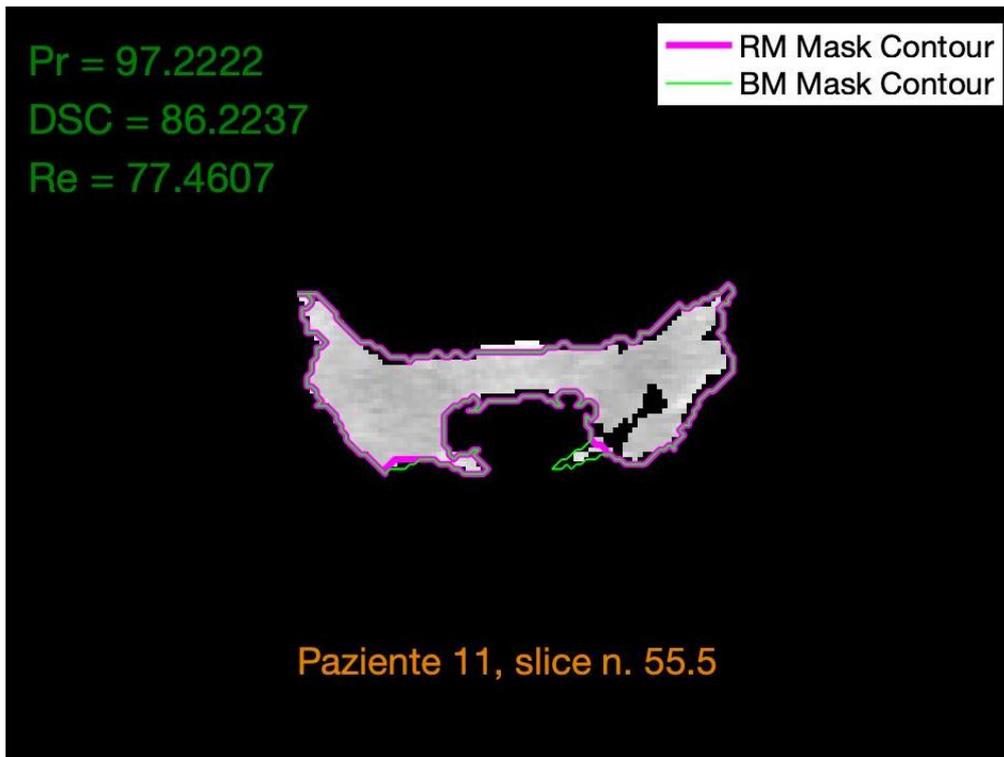


Figura 53, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

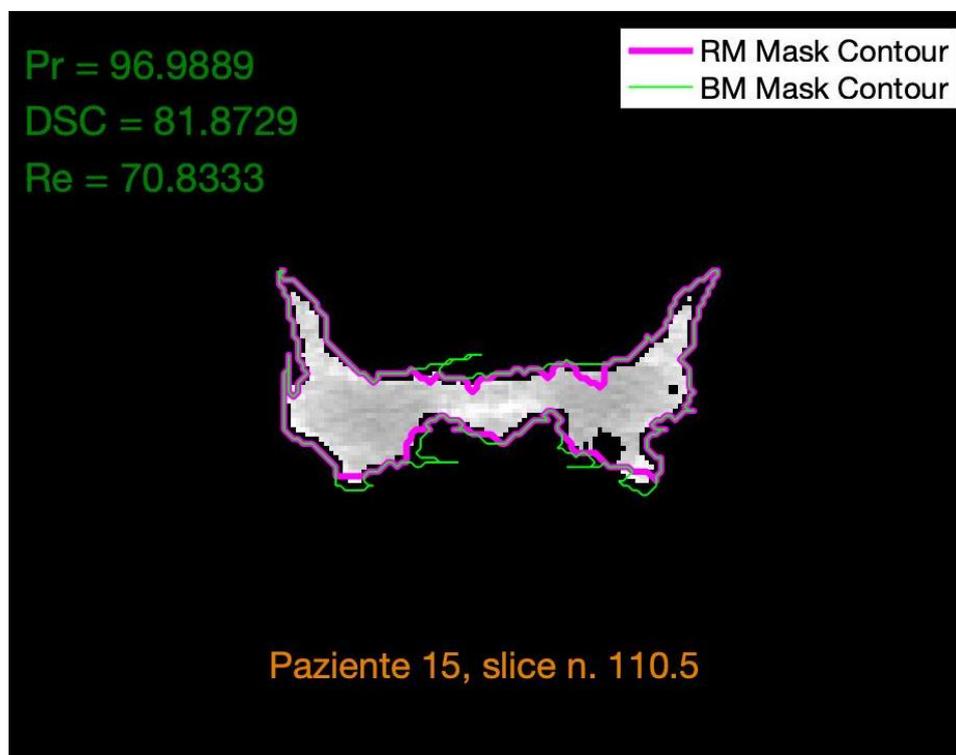


Figura 54, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

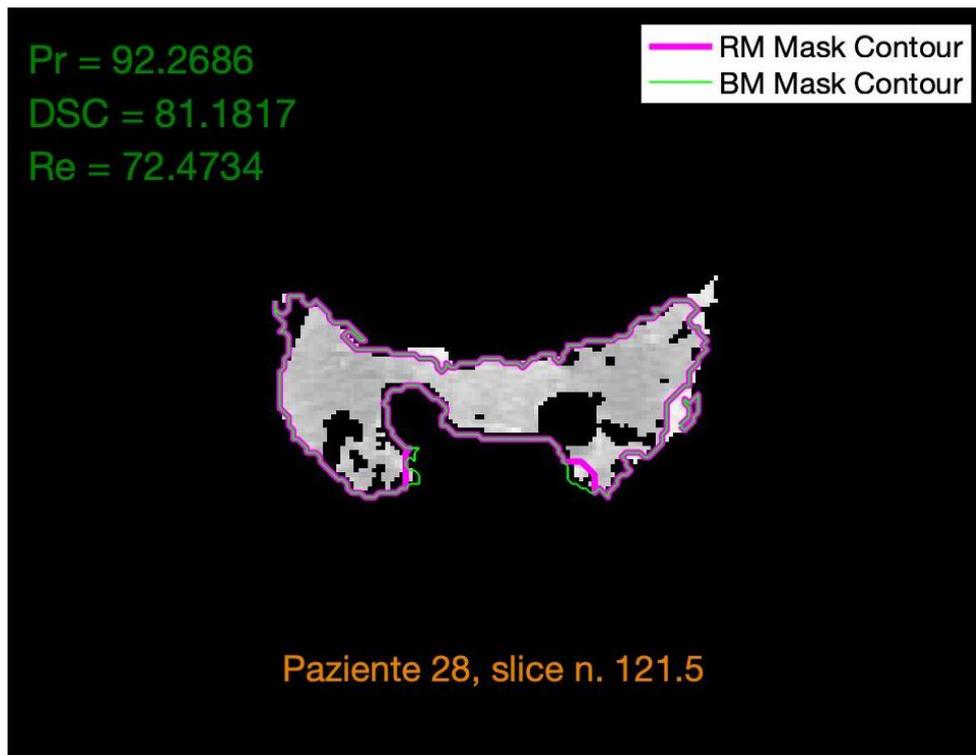


Figura 55, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

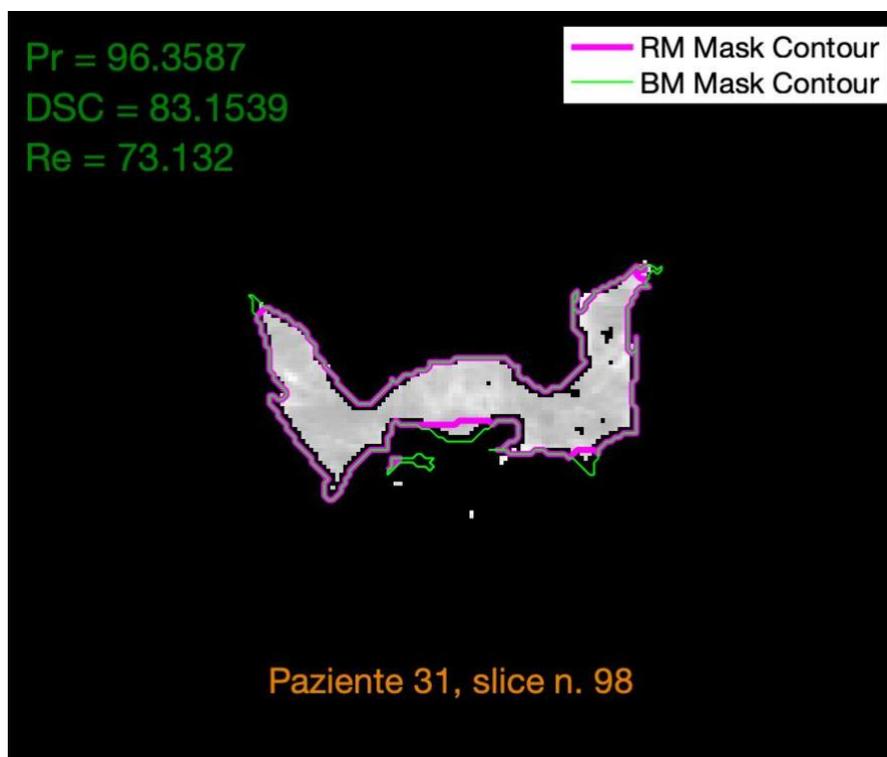


Figura 56, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

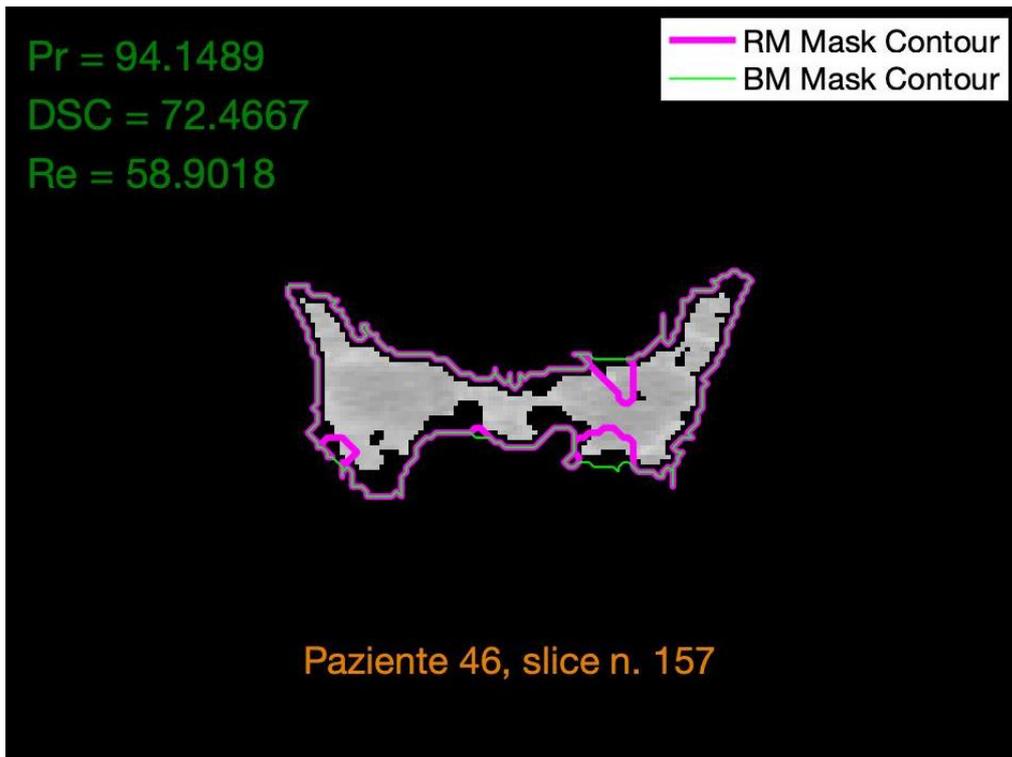


Figura 57, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

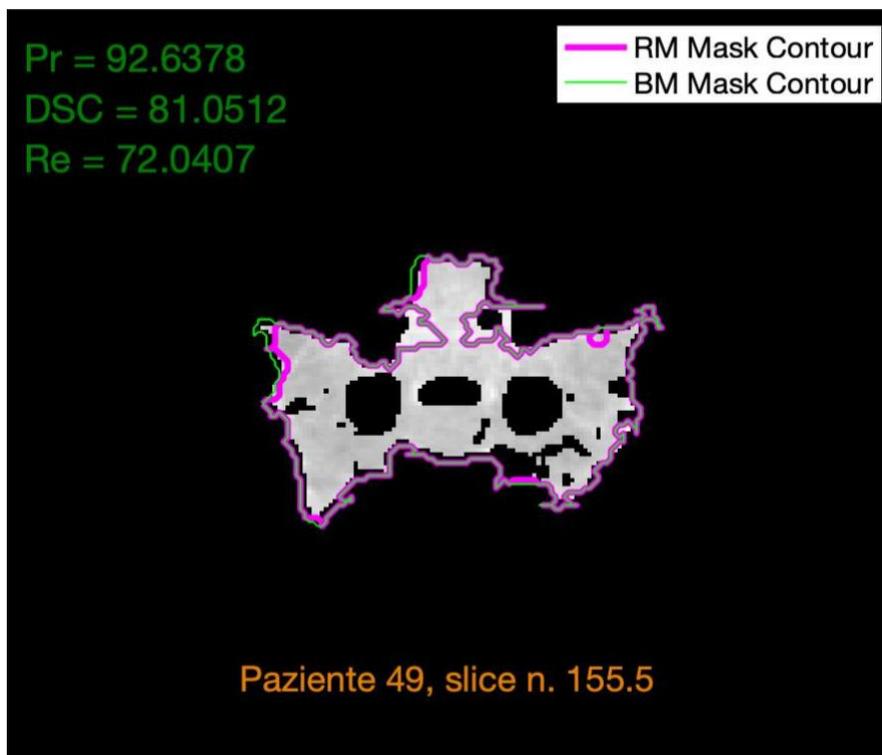


Figura 58, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

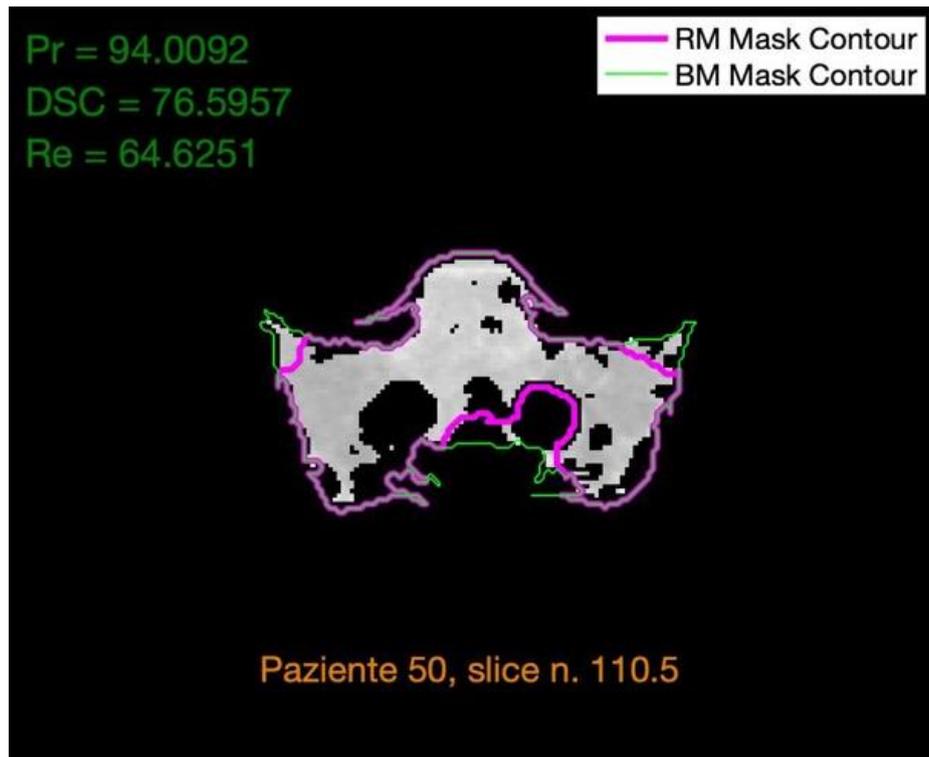


Figura 59, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

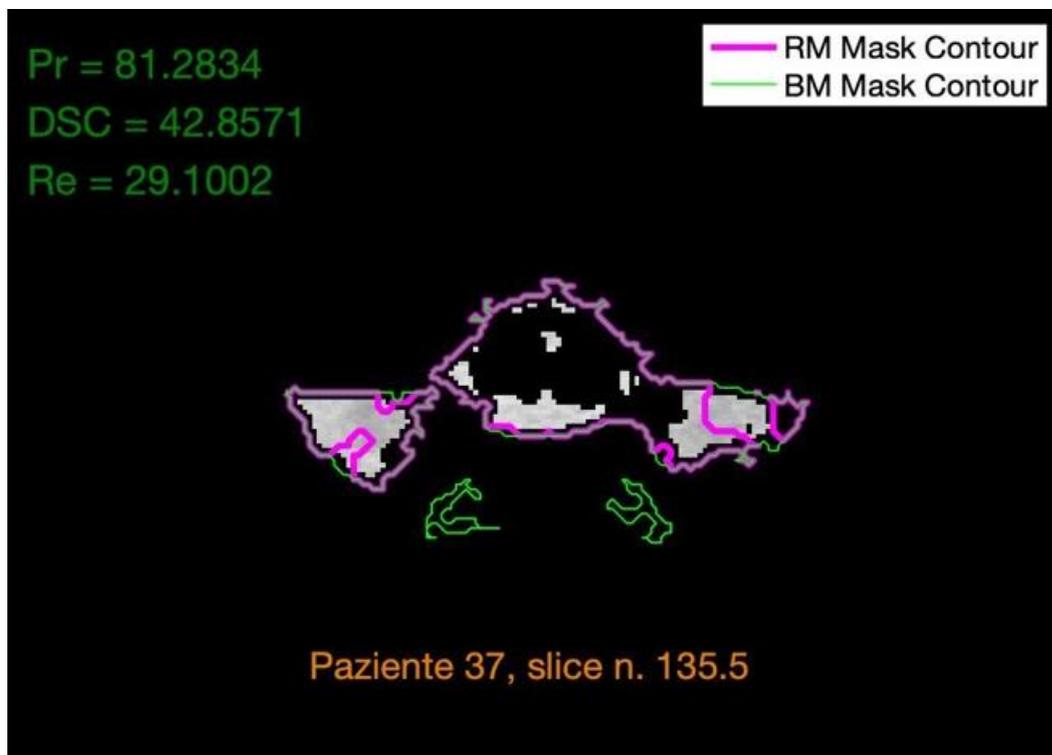


Figura 60, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

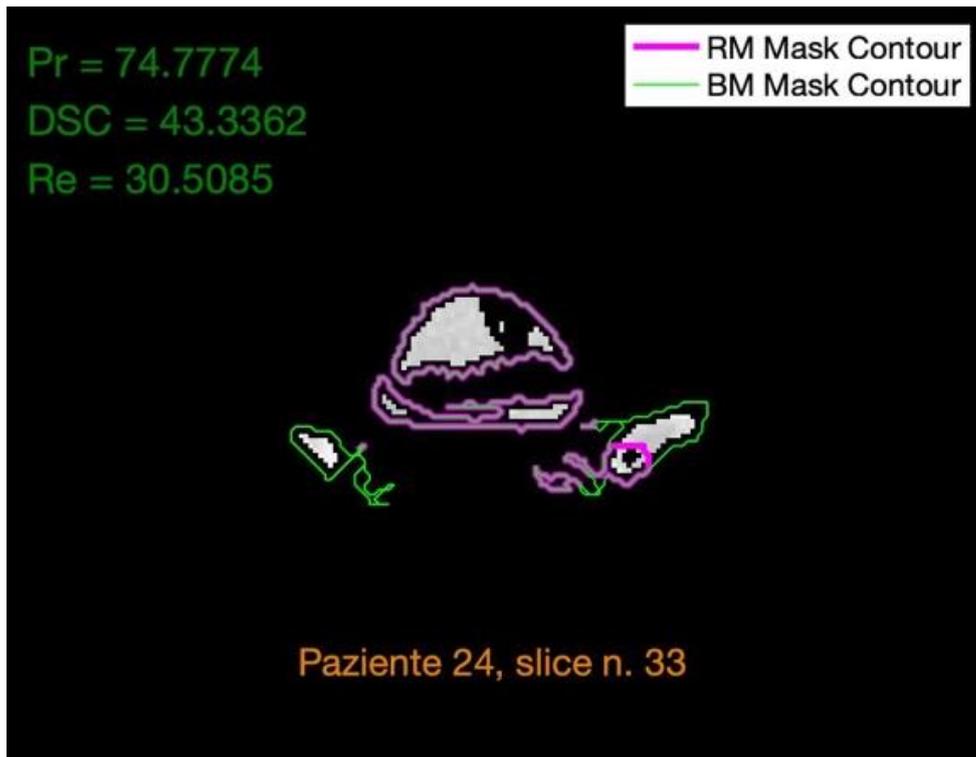


Figura 61, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

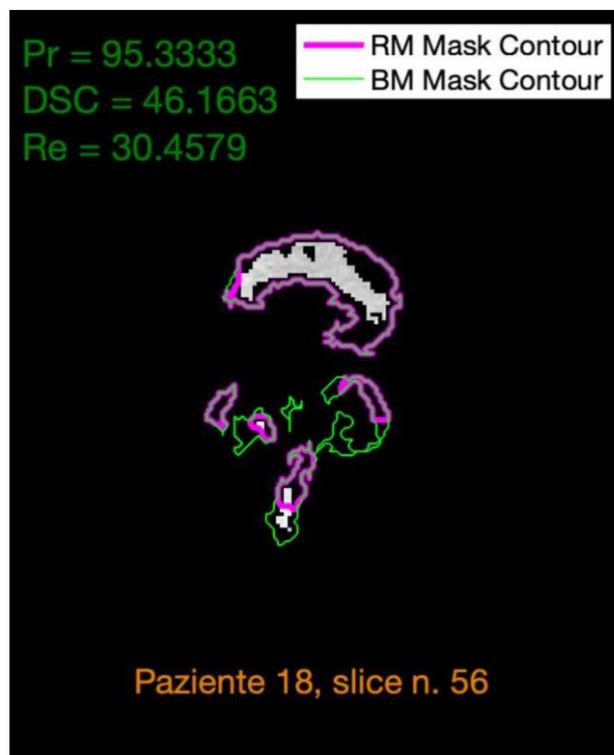


Figura 62, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)

Analizzando le immagini, emerge che la rete è in grado di fornire risultati soddisfacenti nelle slice in cui si ha una morfologia ossea regolare e più lineare, con valori di DSC e Re superiori all'80% (figure 52, 53, 54, 55, 56). Nelle slice in cui, invece, la morfologia è più frastagliata, ed il RM è suddiviso in più regioni non contingenti tra loro, la rete incontra difficoltà nella classificazione, e ciò si traduce in un vistoso peggioramento dei risultati (figure 60, 61, 62). Nella figura 61, infatti, si può notare una sovrasedgmentazione nelle due aree laterali dove non vi è contenuto midollo osseo attivo, e allo stesso tempo una sottosedgmentazione nell'area principale dove risiede il RM. Tuttavia, la rete è in grado di restituire valori buoni di DSC, superiori al 70%, anche nei pazienti non inclusi nel training set (figure 57, 58, 59), quando le slice raffigurano una porzione di tessuto regolare.

Di seguito è possibile osservare i boxplot (grafici 27 e 28) che descrivono la distribuzione di Pr, DSC e Re nei 50 pazienti.

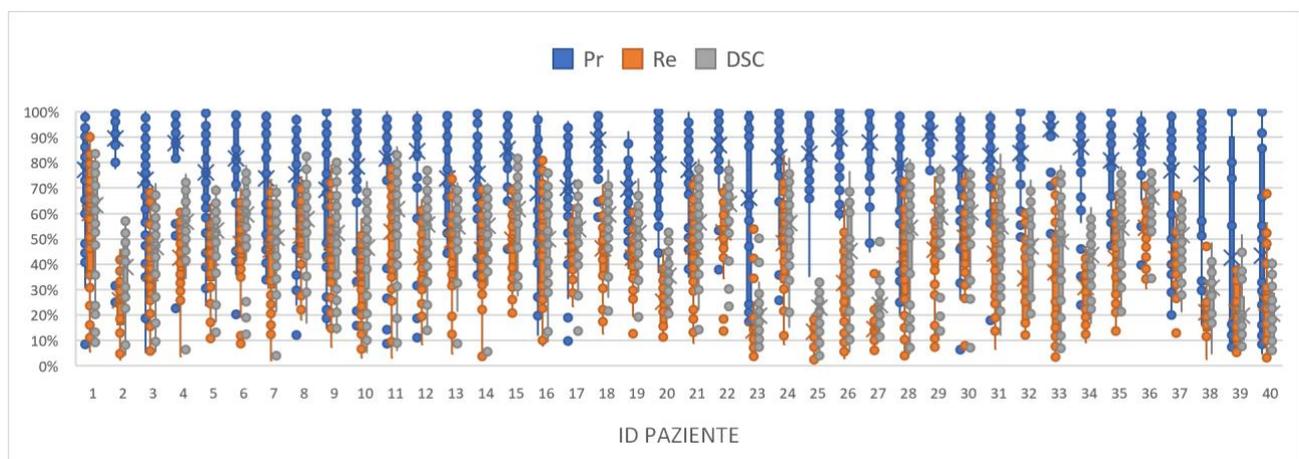


Grafico 49, Boxplot risultati CNN, pazienti 1-40 (training set)

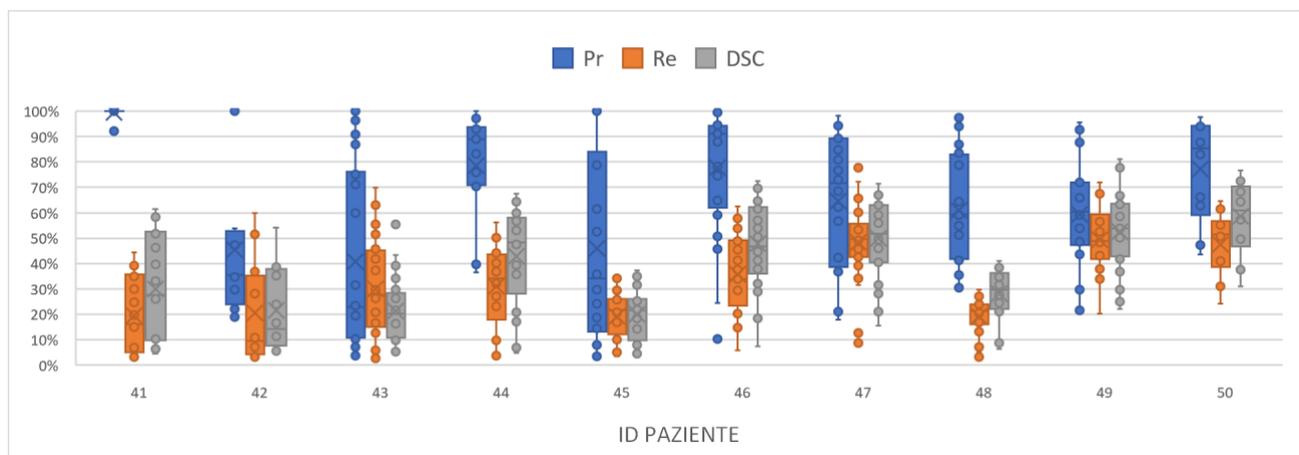


Grafico 50, Boxplot risultati CNN, pazienti 41-50 (validation set)

Per una più facile lettura, è riportato l'andamento medio nei 50 pazienti dei 3 indici.

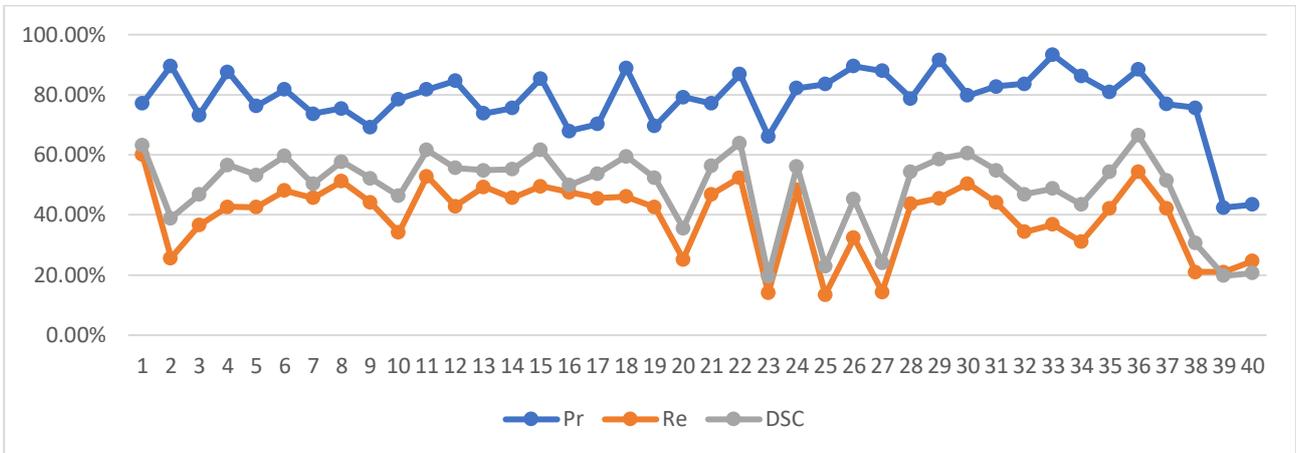


Grafico 51, Media risultati, pazienti 1-40 (training set)

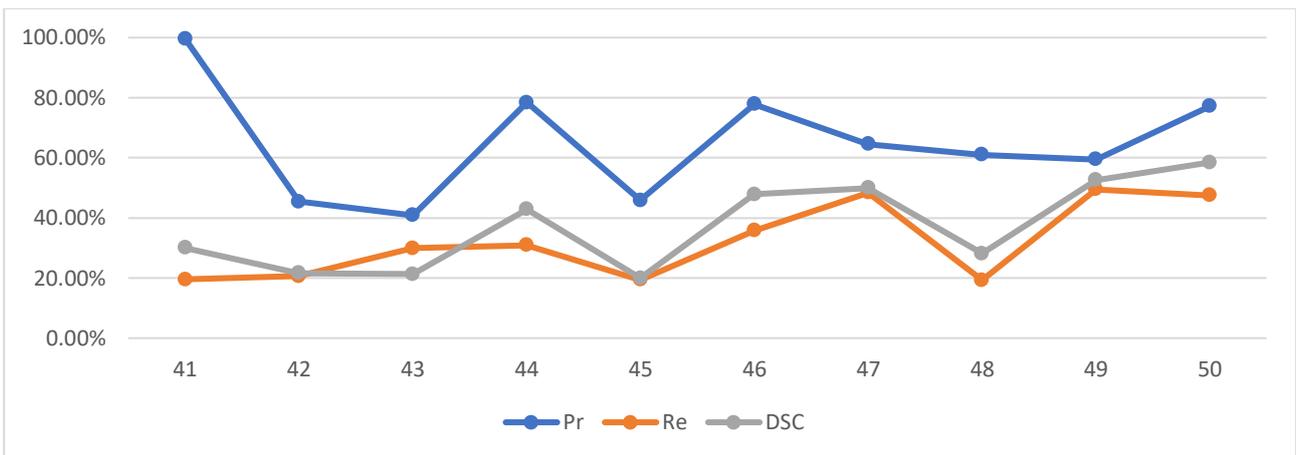


Grafico 52, Media risultati, pazienti 41-50 (validation set)

E' possibile notare come, in linea a quanto accaduto per le DNN, vi sia un netto peggioramento dei risultati nei pazienti non utilizzati come training set (41-50) (grafico 30). In aggiunta, vi sono altri pazienti in cui i risultati sono particolarmente critici, come il 23, 25 (grafico 29), 27, 39,40 (grafico 30). Anche per la CNN, si ha un valore mediamente molto alto per la Pr che si attesta intorno all'80%, salvo subire anch'essa un calo nei pazienti 39, 40, 42, 43, 45, dove raggiunge valori del 50%. La Re, invece, rimane il parametro più basso per la totalità dei pazienti analizzati.

Nel capitolo seguente viene proposto un confronto tra i risultati ottenuti tramite DNN e quelli raggiunti attraverso CNN.

5. Confronto risultati DNN / CNN

5.1 Confronto maschere create con DNN e CNN

Nel seguente capitolo vengono analizzate nel dettaglio le maschere create dai due diversi classificatori, la DNN e la CNN, per comprenderne le differenze ed i punti in comune nella segmentazione del RM. Per la DNN, vengono considerate le immagini create dalla 2° rete DNN, che come emerso nel paragrafo 3.7, ha restituito risultati leggermente migliori della 1° rete DNN. Di seguito vengono proposte le immagini, rappresentando nell'immagine la CT in scala di grigi moltiplicata per la maschera reale ottenuta tramite FDG-PET, e raffigurando in viola il bordo descritto dalla maschera predetta dal classificatore DNN, mentre in blu sarà possibile osservare il bordo della maschera segmentato dalla CNN. In verde vi è il bordo della maschera reale. Vengono stampati nella parte alta a sinistra della figura, nello stesso colore della maschera predetta, i valori di DSC (equazione 9), Pr (equazione 10) e Re (equazione 11) ottenuti da entrambi i classificatore rispetto all'immagine reale. Anche in questo caso, è stato implementato un algoritmo che ha permesso di tagliare l'immagine totale per centrarla nella maschera reale. Nello specifico, sono stati trovati i pixel che delimitavano il bordo superiore, inferiore, destro e sinistro, e sono state tagliate le immagini prendendo 60 pixel oltre ognuno di questi bordi. Così facendo si sono ottenute immagini di grandezza $120 + \text{altezza della maschera} \times 120 + \text{lunghezza della maschera}$ (figura 63).

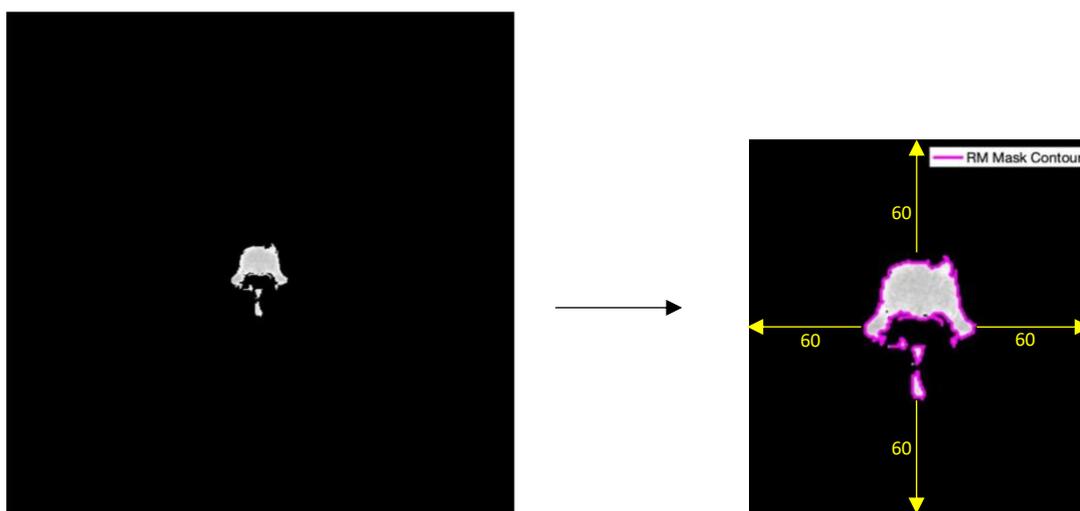


Figura 63, Taglio dell'immagine centrata nella maschera reale

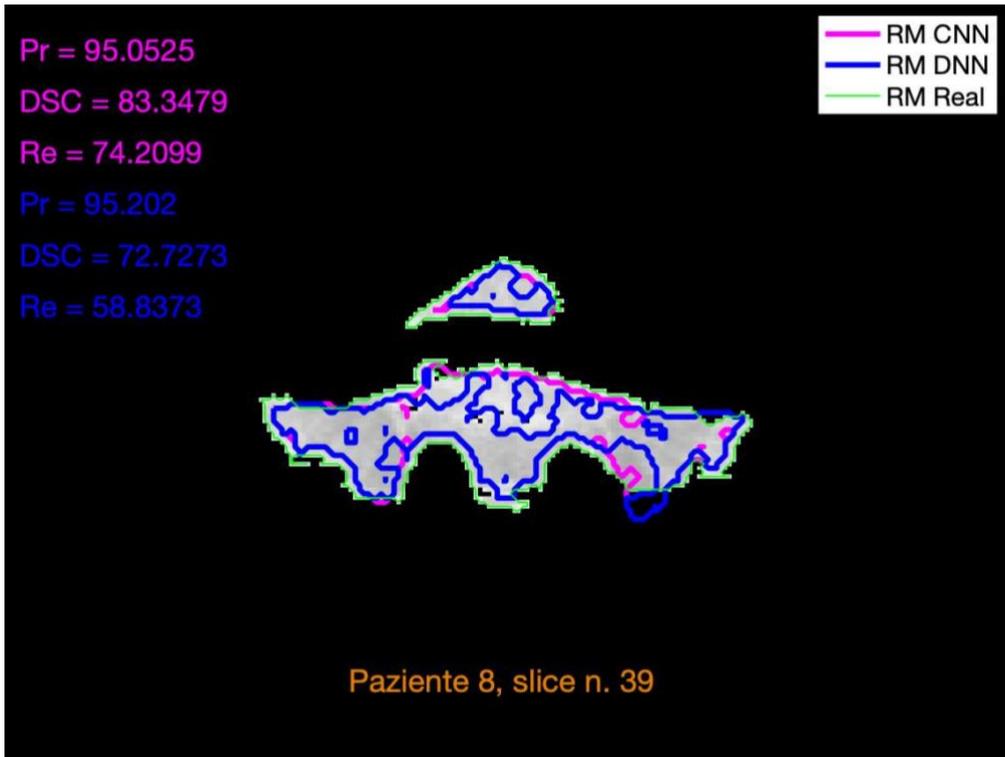


Figura 64, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

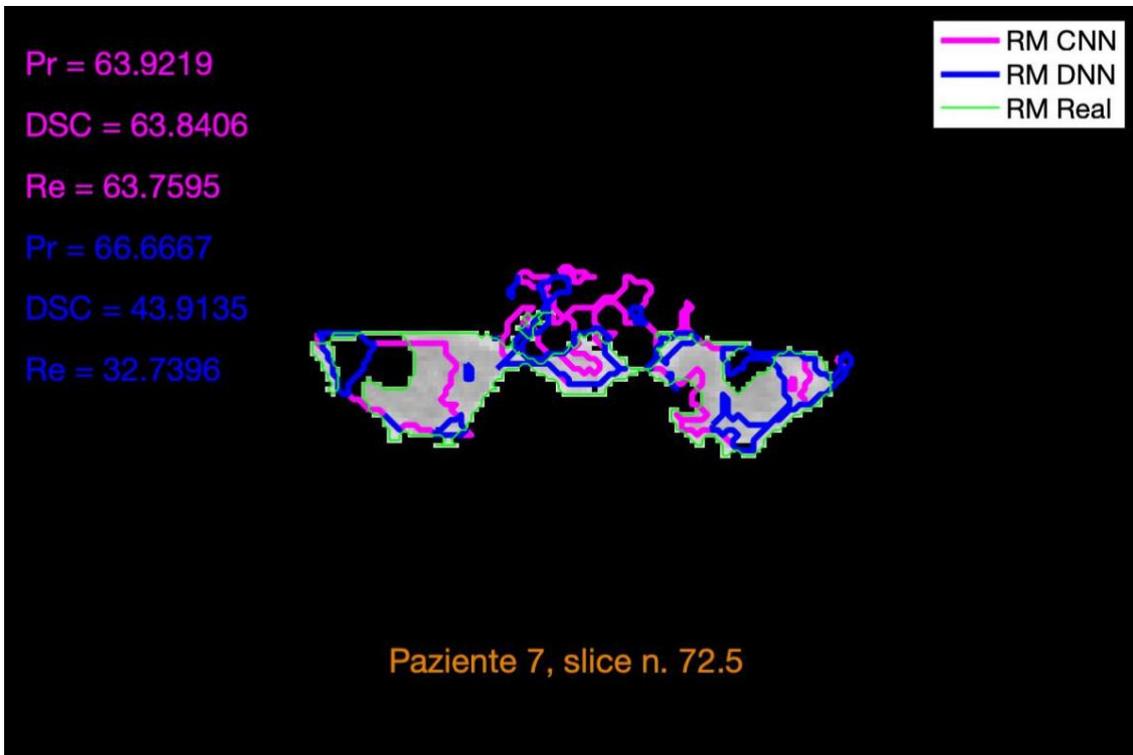


Figura 65, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

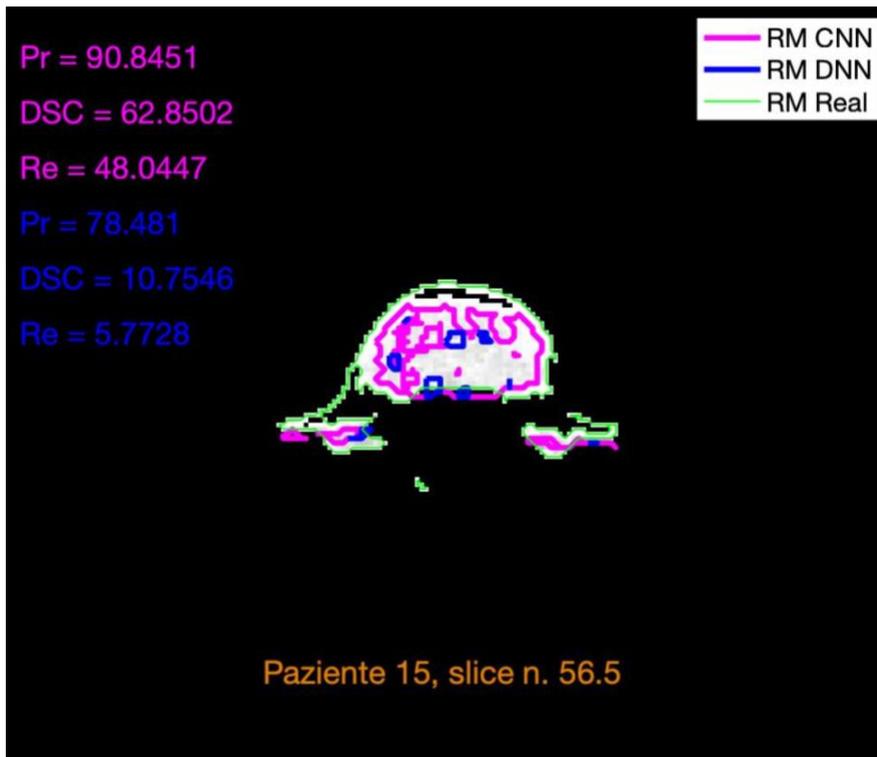


Figura 66, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

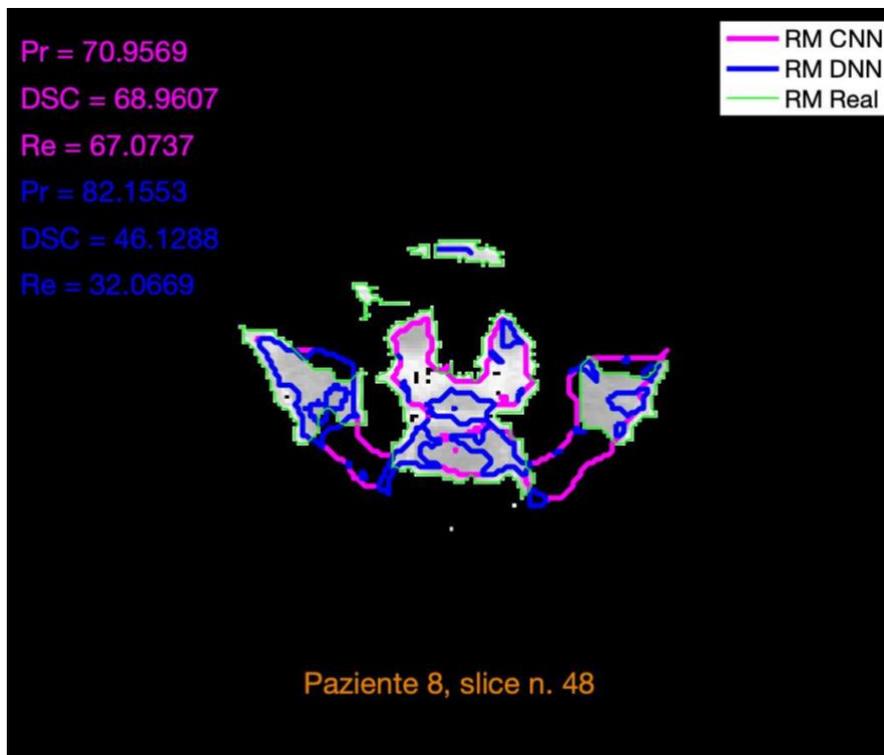


Figura 67, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

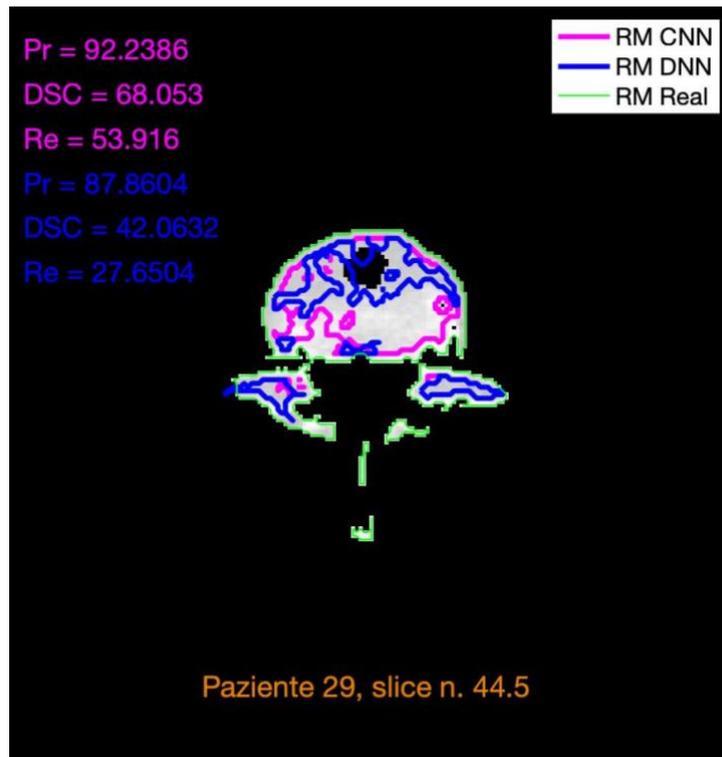


Figura 68, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

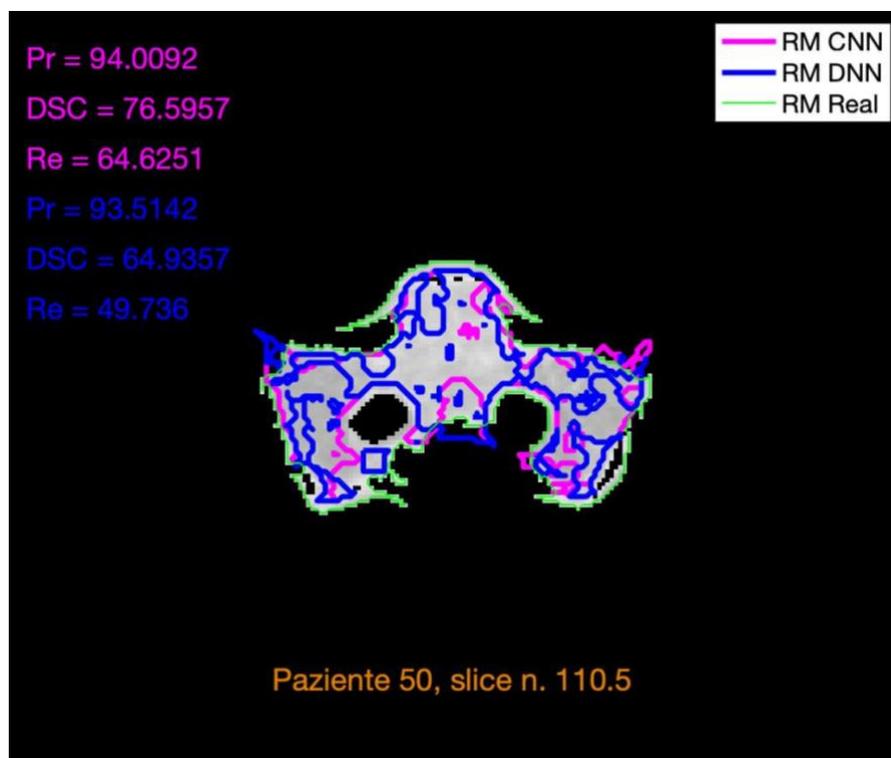


Figura 69, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

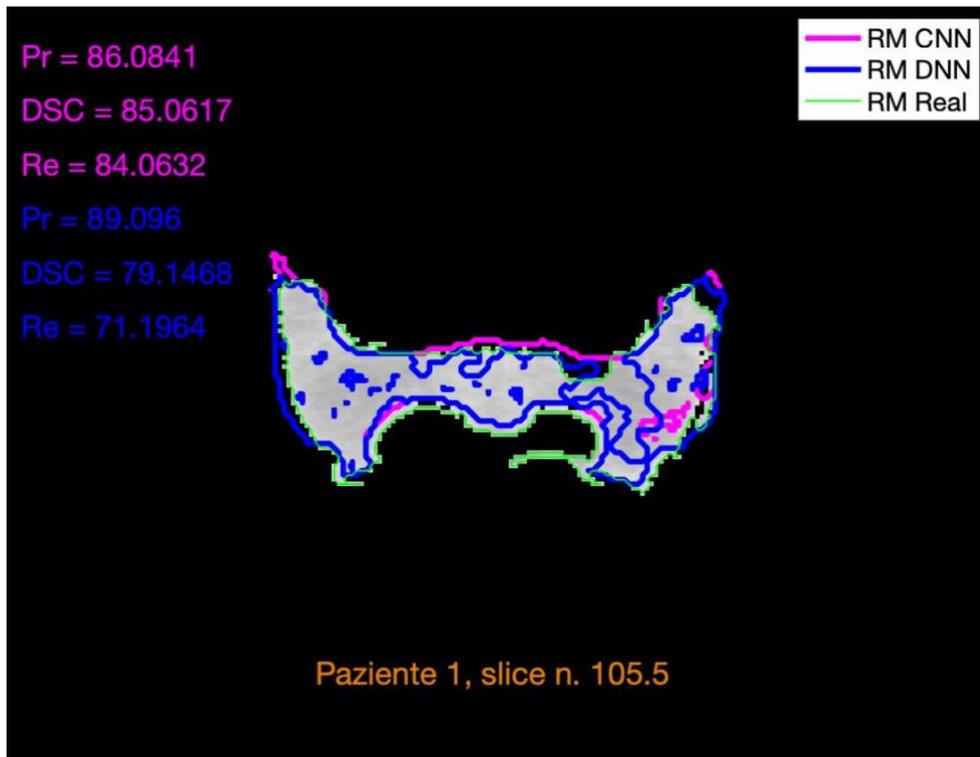


Figura 70, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

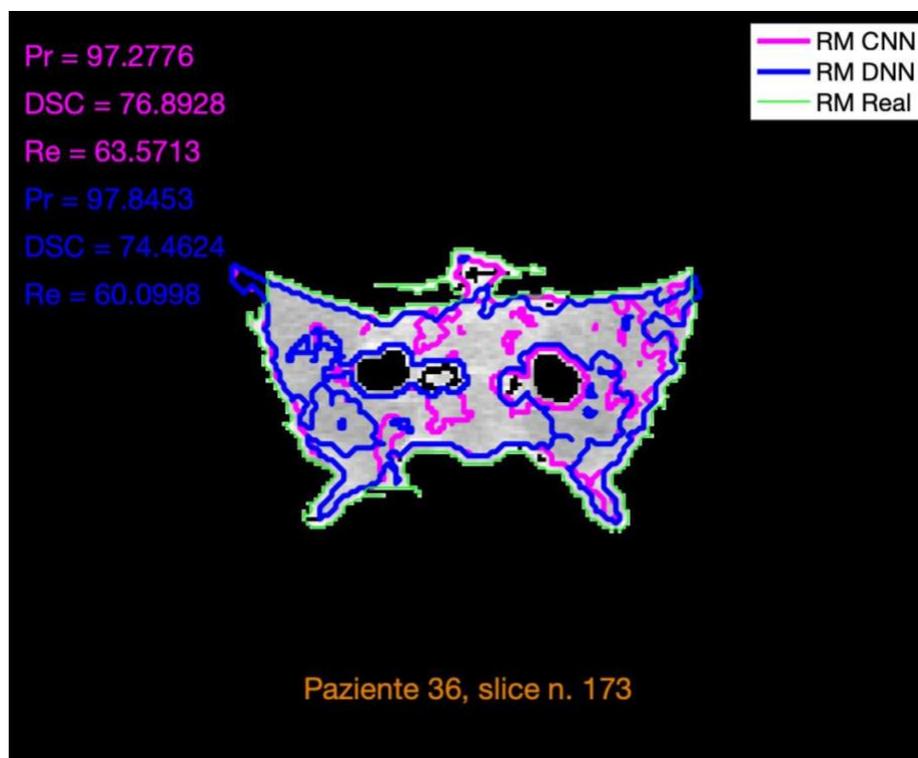


Figura 71, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

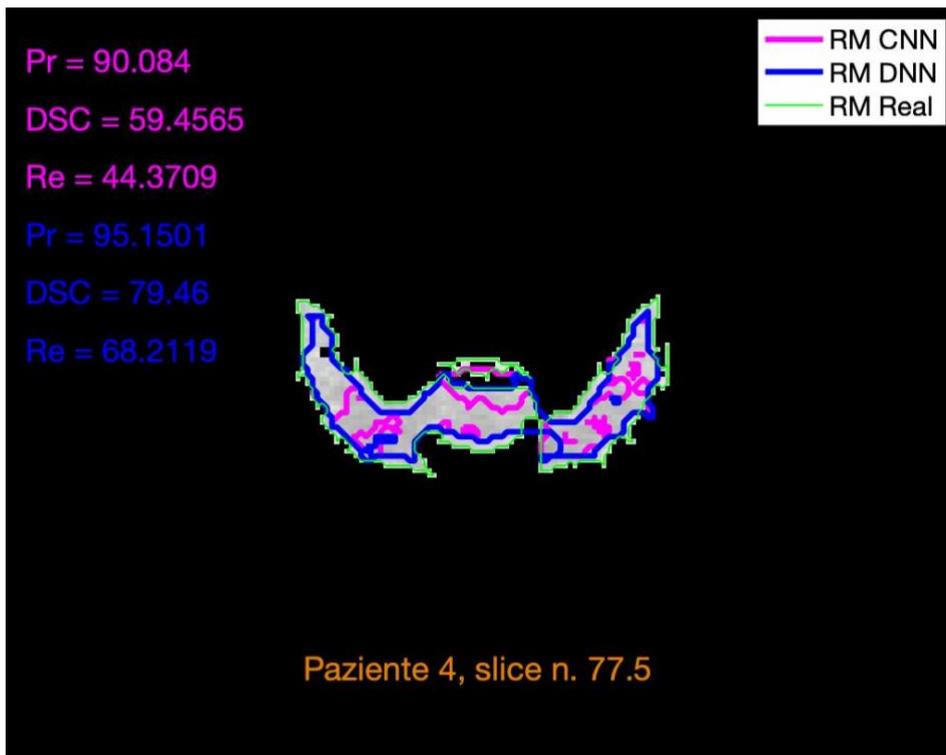


Figura 72, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

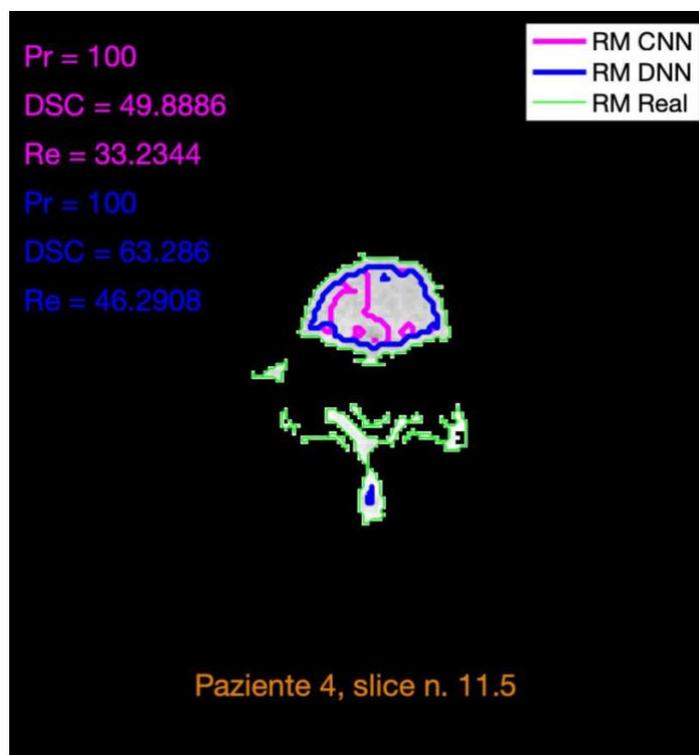


Figura 73, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

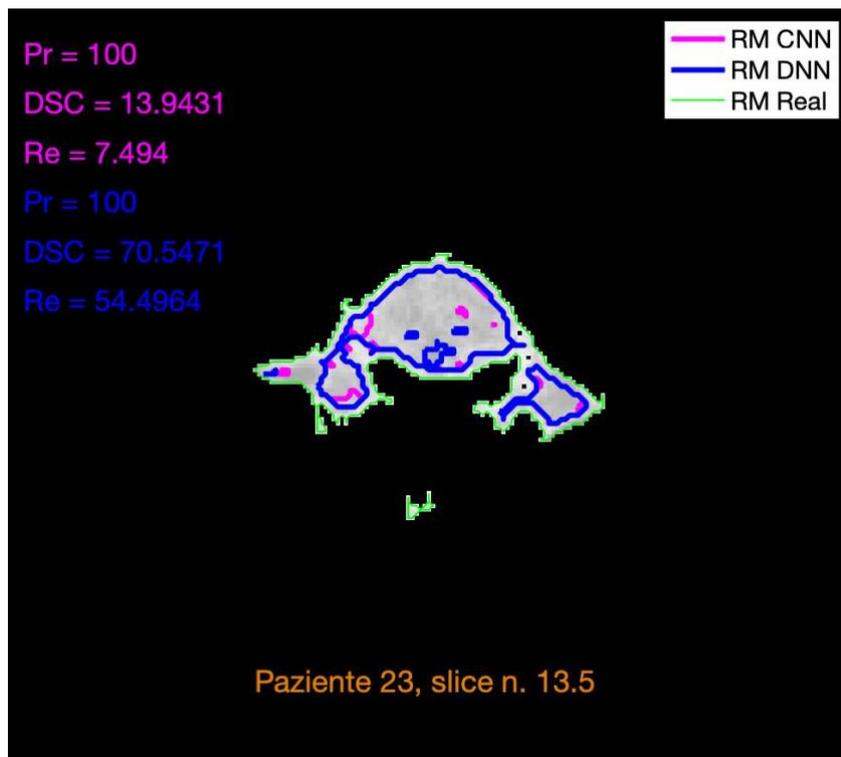


Figura 74, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

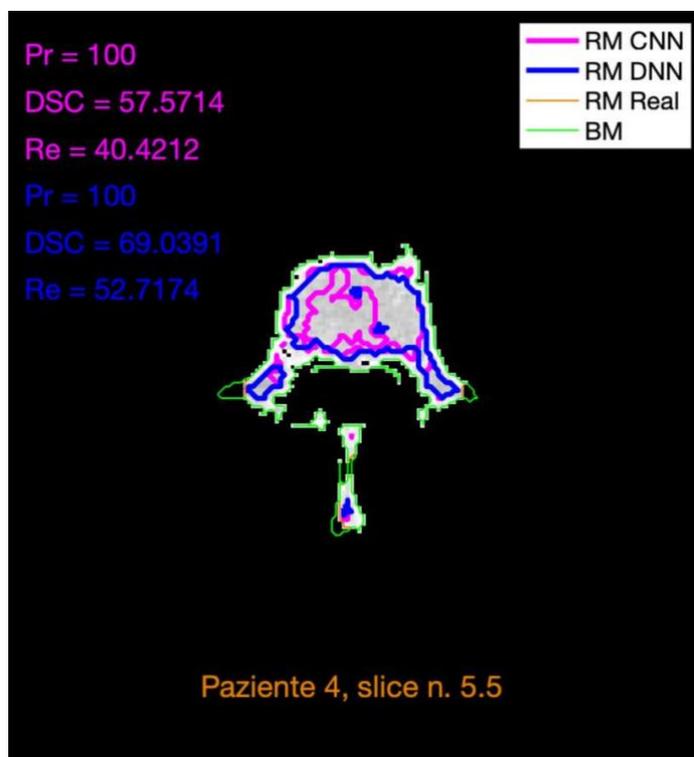


Figura 75, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu)

Nelle precedenti immagini sono visibili differenze sostanziali nella segmentazione adoperata da DNN e CNN. Ad esempio, nella figura 64, le due reti restituiscono valori di Pr simili tra loro, ma la CNN ha 11% in più di DSC e 16% in più di Re, che si traduce visivamente in una maschera più omogenea, mentre la maschera predetta dalla DNN presenta una sottosegmentazione che genera dei buchi. Nella figura 65 la DNN presenta un valore di Pr poco più alto rispetto alla CNN, ma quest'ultima ha però il 20% in più di DSC e più del 30% in più di Re: anche in questo caso la DNN tende a sottosegmentare il RM, anche se entrambe le maschere, in particolar modo la CNN, sovrarsegmentano nella parte centrale. Nella figura 66 è evidente come la DNN non riesca in alcuni casi a riconoscere il RM, fornendo valori di DSC del 10% e di Re di appena 5%. Nella stessa immagine, la CNN riesce invece a segmentare in maniera corretta la quasi totalità della parte centrale del tessuto midollare, mentre non classifica correttamente le aree minori. Un comportamento simile può essere osservato anche nella figura 67, dove la DNN genera una maschera disomogenea, mentre la CNN crea una maschera continua nella zone centrali; tuttavia, anche in questo caso vi è una sottosegmentazione nelle aree isolate più piccole, oltre ad una sovrarsegmentazione nelle aree dell'osso compatto che comporta l'ottenimento di una maschera continua anche nelle aree dove non vi è RM. Nella figura 68 si ripete quanto detto, con la CNN (con oltre 20% in più di DSC) che classifica correttamente gran parte dell'area centrale midollare. In questo caso, la DNN sembra classificare però meglio le aree minori, se si guarda l'area situata in basso a sinistra rispetto all'area principale. Nella coppia di figure 69 e 70 si continua a notare una tendenza da parte della DNN a sottosegmentare le aree RM, mentre la CNN fornisce maschere più omogenee all'interno delle aree circoscritte dalla maschere reale RM.

Nelle figure 71 entrambe le reti ottengono valori di Pr, DSC e Re relativamente alti (rispettivamente intorno al 97%, 75% e 62%).

Osservando la figura 72, invece, si nota come la rete DNN ottenga risultati sensibilmente migliori rispetto alla CNN. Quest'ultima infatti classifica in maniera non ottimale le ROI, ottenendo un DSC del 59%, mentre nella CNN arriva al 79%; lo stesso si verifica per la Re, che ha il 24% in più per la DNN (dal 44% al 68%). La Pr rimane alta per entrambi i classificatori. In questo caso, è la CNN a sottosegmentare e a restituire una maschera disomogenea, mentre la DNN effettua una classificazione prevalentemente corretta. Similmente a ciò, anche nella figura 73 vi è un miglior comportamento da parte della DNN. Quest'ultima segmenta correttamente il tessuto midollare principale (con il 63% di DSC), mentre la CNN ottiene una maschera discontinua all'interno dello stesso (50% DSC). Tuttavia, entrambe le reti non classificano correttamente la quasi totalità dei pixel contenuti nelle aree RM minori, sottostanti l'area principale. Avendo le reti predetto come RM solo ROI circoscritte dal bordo della maschera RM reale, entrambe hanno ottenuto una Pr del 100%. Infine, nella figura 74, si nota un crollo nelle prestazioni della CNN, che restituisce anche in questo caso una Pr del 100%, ma un DSC del 14% ed una Re del 7%. Dalla stessa immagine, la DNN è invece in grado di effettuare una buona classificazione, con una Pr del 100%, un DSC del 70% ed una Re del 55%.

Come è chiaro da quanto emerso, le due reti sono in grado di restituire risultati buoni sulle stesse figure, ma spesso una rete incontra molte più difficoltà dell'altra nella classificazione, e viceversa. Nel prossimo capitolo è possibile analizzare il comportamento delle due reti su tutti i 50 pazienti.

5.2 Confronto dei risultati segmentazione DNN - CNN

Sono riportati i grafici di confronto tra le due reti dei 3 indici. Sono rappresentati i valori mediati tra le slice per ognuno dei 50 pazienti di DSC, Pr e Re, sia per la DNN che per la CNN.

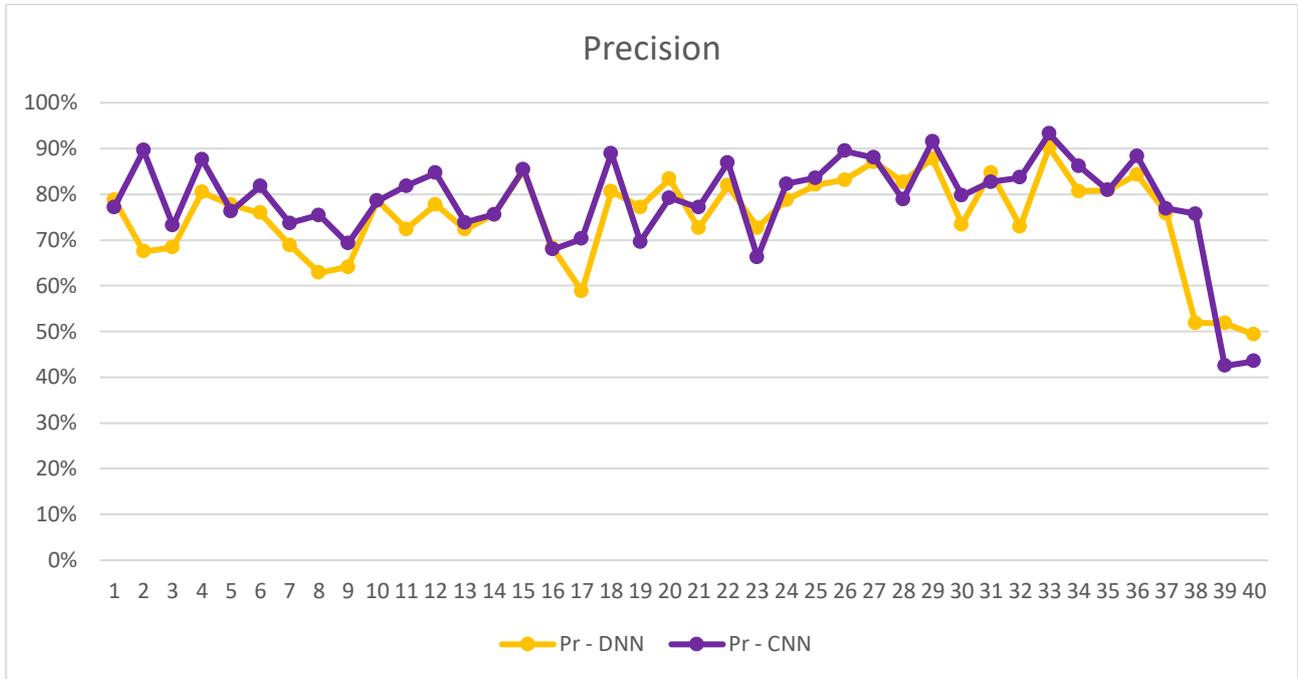


Grafico 53, Confronto DNN-CNN, Precision, Pazienti 1-40 (training set)

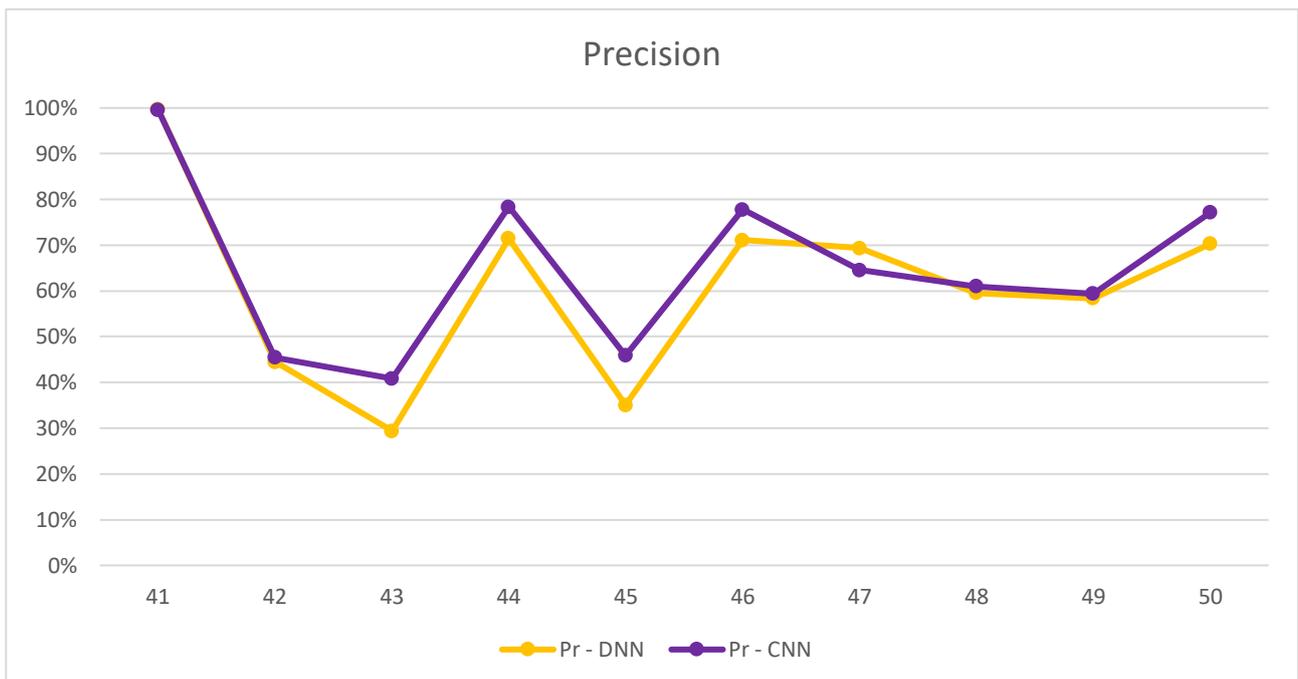


Grafico 54, Confronto DNN-CNN, Precision, Pazienti 41-50 (validation set)

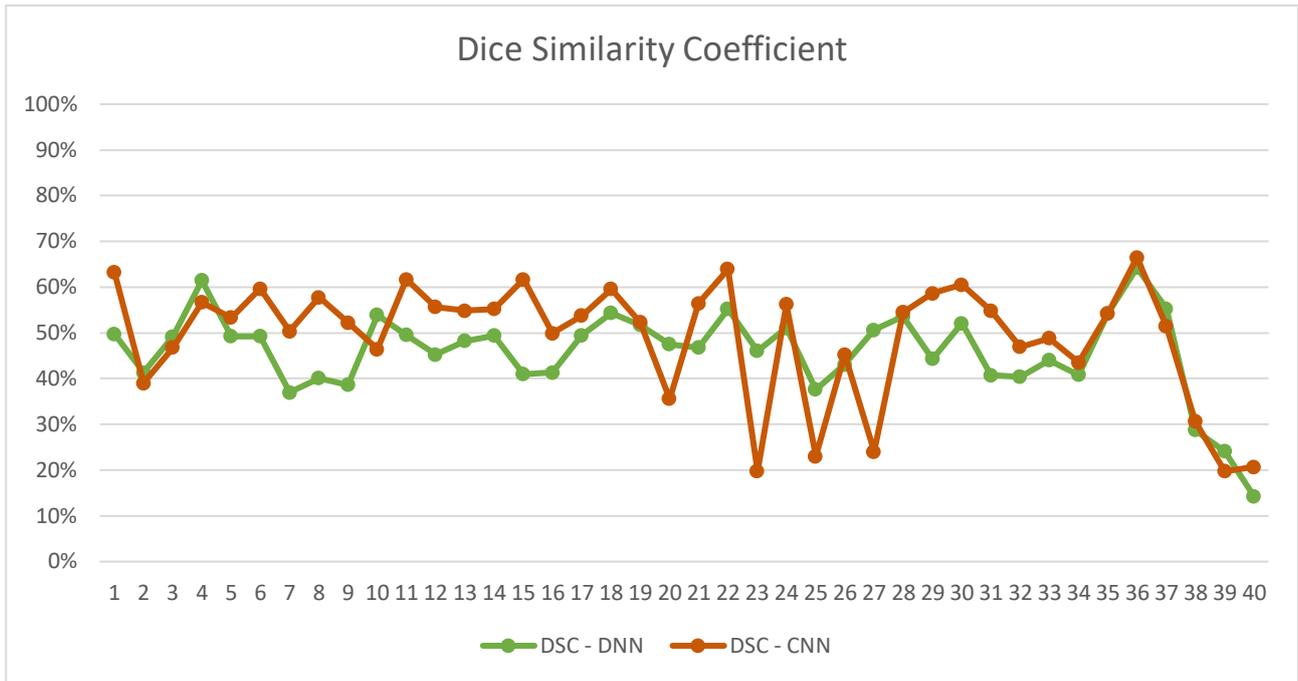


Grafico 55, Confronto DNN-CNN, Dice Similarity Coefficient, Pazienti 1-40 (training set)

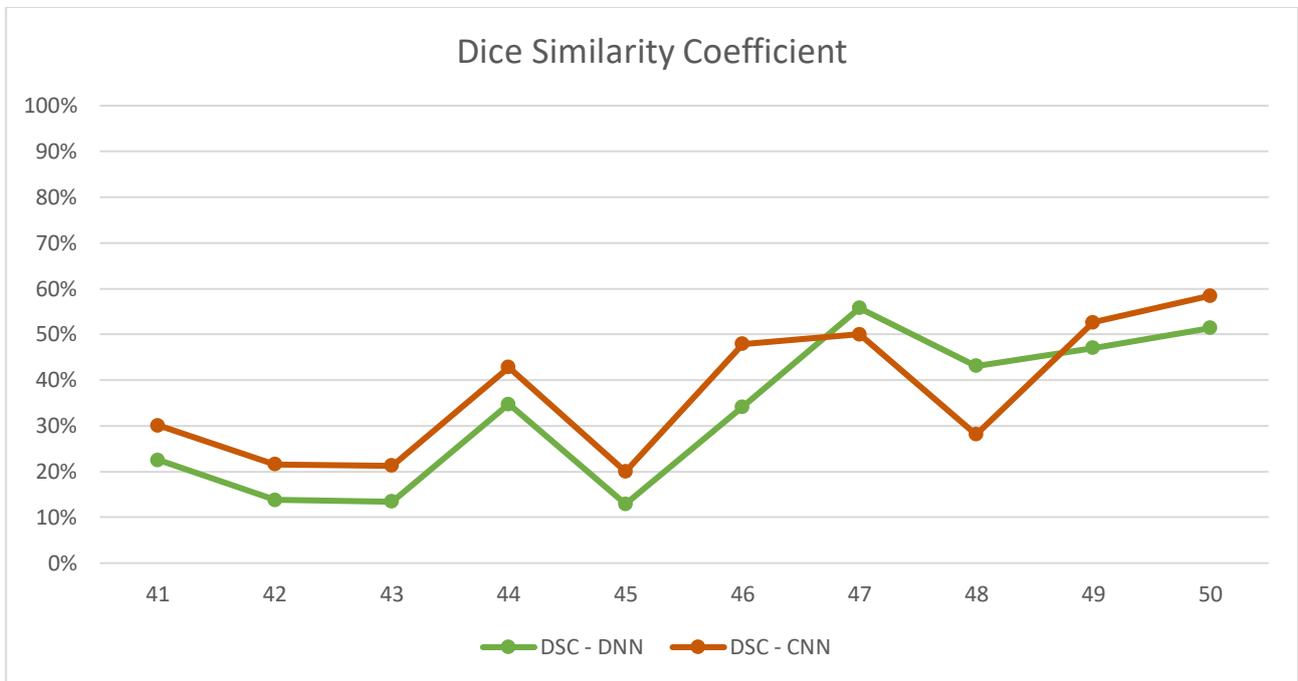


Grafico 56, Confronto DNN-CNN, Dice Similarity Coefficient, Pazienti 41-50 (validation set)

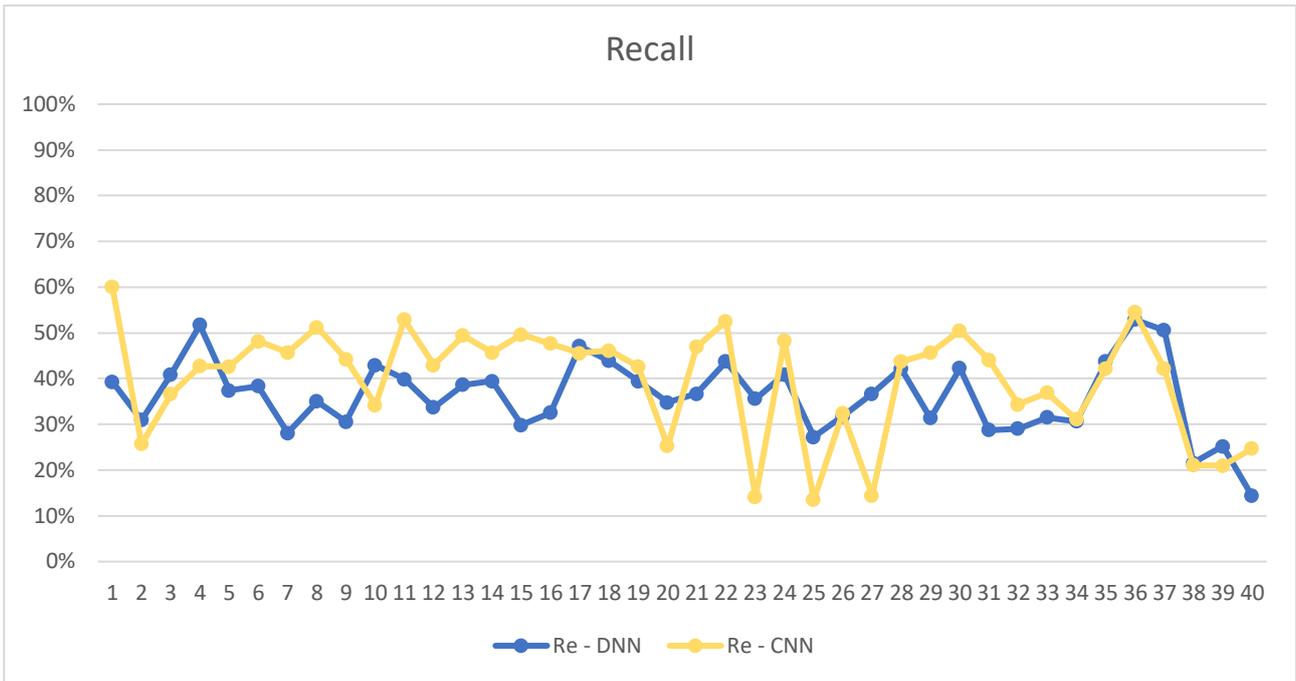


Grafico 56, Confronto DNN-CNN, Recall, Pazienti 1-40 (training set)

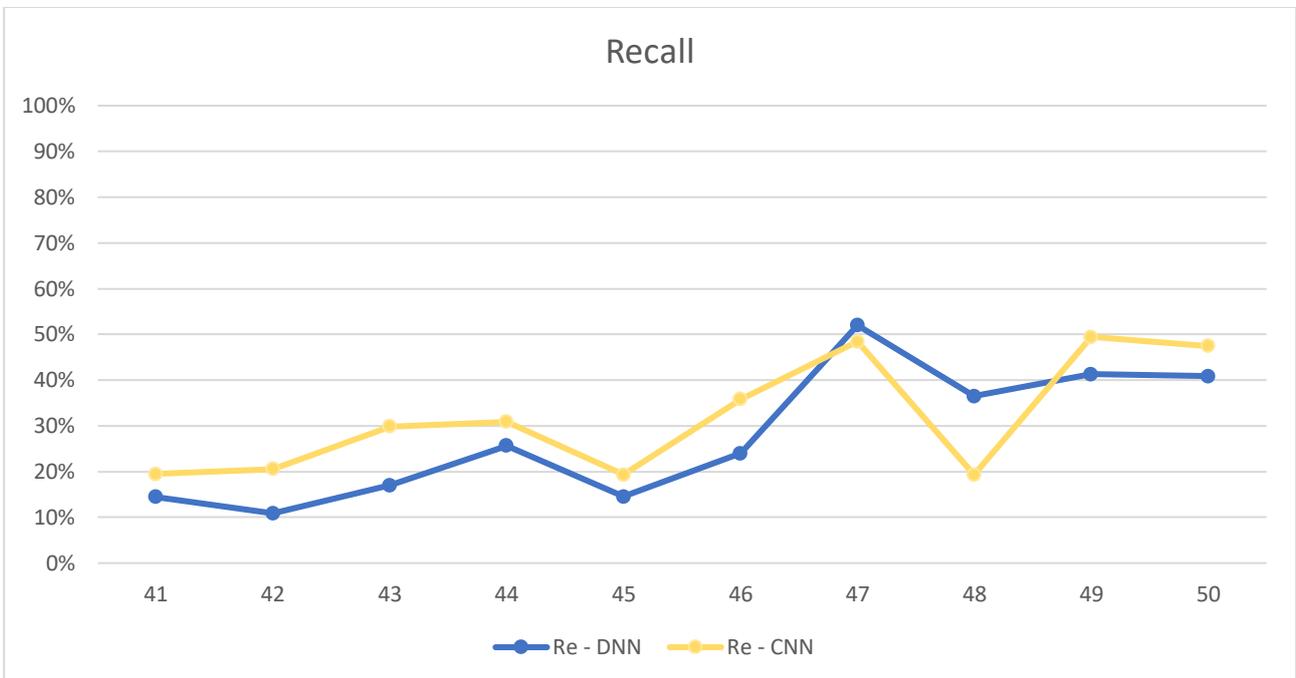


Grafico 57, Confronto DNN-CNN, Recall, Pazienti 41-50 (validation set)

I grafici raffiguranti l'andamento della Pr (grafici 53 e 54) mostrano come questo indice si mantenga prevalentemente alto, al di sopra dell'80%: ciò indica che entrambe le reti tendono a classificare come RM quelle ROI che effettivamente fanno parte di quel tessuto. Vi sono delle eccezioni, soprattutto in alcuni pazienti facenti parte del validation set. Infatti, nei pazienti 38, 39, 40, 42, 43 e 45 la Pr media raggiunge o scende al di sotto del 50%. Confrontando le due reti, è evidente come la CNN abbia mediamente valori più elevati rispetto alla DNN. Solo nei pazienti 20, 28, 39 e 43 la DNN ha una Pr più elevata della CNN.

Nei grafici 55 e 56, che illustrano l'andamento del DSC, si nota (come emerso nei capitoli 3 e 4) che vi sono dei pazienti critici, in maggior parte nel validation set, in cui questo indice subisce un drastico calo. Effettuando un confronto tra CNN e DNN, anche in questo caso emerge come la CNN abbia valori mediamente più elevati rispetto alla DNN. Tuttavia, vi sono dei pazienti in cui la CNN ha valori di gran lunga inferiori alla DNN, come i pazienti 23 e 27, dove la CNN presenta un DSC medio più basso del 26% rispetto alla DNN.

L'andamento del terzo parametro, la Re, è descritto nei grafici 56 e 57. Esso è l'indice più critico dei tre, attestandosi mediamente tra il 40% ed il 50%, raggiungendo valori di poco superiori al 10% in alcuni pazienti particolarmente critici. Il fatto che questo parametro assuma così bassi valori si traduce in una sottosegmentazione da parte delle reti del RM. Anche questo indice conferma come la CNN abbia risultati mediamente migliori della DNN, ma anche in questo caso con la presenza di eccezioni in cui la DNN restituisce valori maggiori, come i pazienti 2, 3, 4, 10, 23, 25, 27, 37, 39, 49.

Conclusione

In questo elaborato è stato proposto un confronto tra due tipologie di algoritmi di Deep Learning applicati nella segmentazione di midollo rosso attivo a partire da immagini CT: Le DNN e le CNN. Queste reti hanno classificato delle ROI di dimensione 5x5 per ricostruire l'immagine predetta, che è poi stata confrontata con la maschera reale ottenuta tramite FDG-PET.

Per entrambe le tipologie, si è prima proceduto con un tuning dei parametri, ovvero si sono svolte diverse prove al fine di trovare l'assetto in grado di fornire i risultati migliori. Per la DNN, si è deciso di utilizzare un'architettura con 36 neuroni nel layer di input e 7 layer nascosti con rispettivamente 31, 26, 21, 16, 11, 6 e 3 neuroni. Tutti questi neuroni sono caratterizzati dalla funzione di attivazione *tansig*. Nel layer di output, invece, un neurone con funzione di attivazione *purelin*. Per la CNN si è scelta una struttura con 3 layer di convoluzione di dimensione 3x3, con funzione di attivazione *tanh*.

Una volta ottenute le maschere predette dai classificatori, si è proceduto al confronto con le suddette maschere reali, utilizzando tre indici, il Dice Similarity Coefficient (DSC, equazione 9), la Precision (Pr, equazione 11) e la Recall (Re, equazione 10).

Analizzando visivamente le immagini create dalle reti, e guardando ai risultati numerici dei 3 indici prima citati, si è notato che entrambe le reti tendono a segmentare correttamente il midollo all'interno delle aree circoscritte dalle maschere reali. Questo si è tradotto numericamente in una Pr molto elevata (superiore all'85%) per entrambe le reti nella quasi totalità dei pazienti. Al contrario, la Re si è rilevata un parametro critico (30-50%) per molte immagini: ciò ha portato a una sottosegmentazione del RM. Questo si è verificato soprattutto nelle immagini che presentavano delle zone di midollo più piccole rispetto alla zona principale, nella quale la classificazione è stata effettuata prevalentemente in maniera corretta. Inoltre, è emerso che nel dataset sono presenti un importante numero di immagini nelle quale la quantità di midollo è molto ridotta, e ciò, in accordo a quanto affermato precedentemente, ha comportato un crollo nelle prestazioni delle rete, abbassando la media dei tre indici. La DSC, che si è attestata intorno al 60%, ha risentito del problema della sottosegmentazione. In diverse immagini dove la morfologia midollare era più regolare, le reti sono state in grado di restituire ottimi risultati, con valori dei 3 indici superiori all'80%.

Confrontando i risultati raggiunti dalle due reti, si è notato che la CNN ha ottenuto mediamente risultati migliori rispetto alla DNN, anche del 20% per i 3 indici in alcuni pazienti (considerando i valori mediati su tutte le immagini acquisite dallo stesso paziente). Infatti, in alcune immagini, mentre la DNN aveva valori di DSC del 10%, la CNN riusciva a classificare in maniera corretta il RM, ottenendo oltre il 70% di DSC. Tuttavia, in alcuni pazienti critici, è stata la DNN a restituire valori maggiori di Pr, DSC e Re.

Note bibliografiche

- [1] Keys HM, Bundy BN, Stehman FB, Muderspach LI, Chafe WE, Suggs CL 3rd, Walker JL, Gersell D. «*Cisplatin, radiation, and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma.*», 1999
- [2] Rose PG, Bundy BN, Watkins EB, et al. «*Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer*», 1999
- [3] Peters WA 3rd, Liu PY, Barrett RJ 2nd, Stock RJ, Monk BJ, Berek JS, Souhami L, Grigsby P, Gordon W Jr, Alberts DS. «*Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix.*», 2000
- [4] Green JA, Kirwan JM, Tierney JF, et al. «*Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix.*», 2001
- [5] Torres MA, Jhingran A, Thames HD Jr, Levenback CF, Bodurka DC, Ramondetta LM, Eifel PJ. «*Comparison of treatment tolerance and outcomes in patients with cervical cancer treated with concurrent chemoradiotherapy in a prospective randomized trial or with standard treatment*», 2008
- [6] Mell LK, Tiryaki H, Ahn KH, Mundt AJ, Roeske JC, Aydogan B. «*Dosimetric comparison of bone marrow-sparing-intensity-modulated radiotherapy versus conventional techniques for treatment cervical cancer.*», 2008
- [7] Judy S. Blebeal et al. «*Structural and Functional Imaging of Normal Bone Marrow and Evaluation of Its Age-Related Changes.* » *Seminars in Nuclear Medicine*, 2007, Vol.37(3), pp.185-194, 2007
- [8] Kugel H, Jung C, Schulte O, et al: «*Age- and sex-specific differences in the ¹H-spectrum of vertebral bone marrow.* » *J Magn Reson Imaging* 13:263-268, 2008
- [9] A. Andreyvchenko et al. «*The feasibility of semi-automatically generated red bone marrow segmentations based on MR-only for patients with gynecologic cancer.* » *Radiotherapy and Oncology*, April 2017, Vol.123(1), pp.164-168
- [10] Desai AG, Thakur ML. «*Radiopharmaceuticals for spleen and bone marrow studies.* » *Semin Nucl Med.* 1985;15:229–38
- [11] Kogan et al. «*Applications of PET-MRI in Musculoskeletal Disease.* » *Journal of Magnetic Resonance Imaging*, July 2018, Vol.48(1), pp.27-47
- [12] H. Zaidi, A. Abass, et al. «*Novel Quantitative PET Techniques for Clinical Decision Support in Oncology.* » *Seminars in Nuclear Medicine*, November 2018, Vol.48(6), pp.548-564
- [13] P. Franco et al. «*Incorporating ¹⁸-FDG-PET-defined pelvic active bone marrow in the automatic treatment planning process of anal cancer patients undergoing chemo-radiation.* » *BMC Cancer*, 01 November 2017, Vol.17(1), pp.1-11
- [14] T. Perk, et al. «*A method for quantitative total marrow imaging (QTMI) with PET/CT*» *Biomedical Physics & Engineering Express*, Vol.2(5), p.055006 (13pp), 2016

- [15] Gillies RJ, Kinahan PE, Hricak H. Radiomics «*Images are more than pictures, they are data.*» *Radiology*; 278:563–577. doi: 10.1148/radiol.2015151169, 2016
- [16] Rizzo S, Petrella F, Buscarino V, et al. «*CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer.*» *Eur Radiol*; 26:32–42. doi: 10.1007/s00330-015-3814-0, 2016
- [17] Larue Ruben T. H. M., van Timmeren Janna E., de Jong Evelyn E. C., Feliciani Giacomo, Leijenaar Ralph T. H., Schreurs Wendy M. J., Sosef Meindert N., Raat Frank H. P. J., van der Zande Frans H. R., Das Marco, van Elmpt Wouter, Lambin Philippe. «*Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study.*» *Acta Oncologica*; 56(11):1544–1553, 2017
- [18] Balagurunathan Y, Kumar V, Gu Y, et al. «*Test-retest reproducibility analysis of lung CT image features.*» *J Digit Imaging* ; 27:805–823. doi: 10.1007/s10278-014-9716-x, 2014
- [19] Charu C. Aggarwal. «*Neural Networks and Deep Learning.*», 2018
- [20] Lee, Gobert, Fujita, Hiroshi. «*Deep Learning in Medical Image Analysis: Challenges and Applications.*», 2020
- [21] P. Meyer, V. Noblet, C. Mazzara and A. Lallement. «*Survey on deep learning for radiotherapy.*», 2018
- [22] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu et al. «*A deep learning- based radiomics model for prediction of survival in glioblastoma multiforme.*», 2017
- [23] A. Mehranian, H. Arabi and H. Zaidi. «*Vision 20/20: magnetic resonance imaging- guided attenuation correction in PET/MRI: challenges, solutions, and opportunities.*», 2016
- [24] F. Liu, H. Jang, R. Kijowski, T. Bradshaw and A.B. McMillan. «*Deep learning MR imaging- based attenuation correction for PET/MR imaging.*», 2018
- [25] L. Oakden-Rayner, G. Carneiro, T. Bessen, J.C. Nascimento, A.P. Bradley and L.J. Palmer. «*Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework.*», 2017
- [26] Krizhevsky, A., Sutskever, I., Hinton. «*Imagenet classification with deep convolutional neural networks.*», 2012
- [27] Girshick, R., Donahue, J., Darrell, T., Malik, J. «*Rich feature hierarchies for accurate object detection and semantic segmentation.*», 2014

Indice Figure

Figura 1, Fonte: Illustration Poster Print by Gwen Shockey/Science Source, Tessuto midollare.....	1
Figura 2, Slice originale	10
Figura 3, Maschera BM con osso compatto	10
Figura 4, Maschera RM con osso compatto	10
Figura 5, Maschera BM senza osso compatto	10
Figura 6, Maschera RM senza osso compatto	10
Figura 7, Modello di Neurone di McCulloch	13
Figura 8, Esempio di struttura NN.....	14
Figura 9, Esempio di struttura DNN.....	14
Figura 10, Possibili funzioni di attivazione dei neuroni.....	15
Figura 11, Struttura triangolare DNN.....	16
Figura 12, Confusion Matrix.....	17
Figura 13, Principio di funzionamento curva ROC.....	18
Figura 14, Struttura #4 scelta per la segmentazione.....	30
Figura 15, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	36
Figura 16, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	36
Figura 17, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	37
Figura 18, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	37
Figura 19, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	38
Figura 20, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	38
Figura 21, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 22, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 23, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 24, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 25, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 26, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 27, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 28, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 29, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39

Figura 30, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	39
Figura 31, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	40
Figura 32, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	40
Figura 33, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	41
Figura 34, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	41
Figura 35, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	42
Figura 36, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	42
Figura 37, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	43
Figura 38, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	43
Figura 39, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	44
Figura 40, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	44
Figura 41, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	45
Figura 42, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	45
Figura 43, Segmentazione tramite 1° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	46
Figura 44, Segmentazione tramite 2° DNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	46
Figura 45, Tipica struttura CNN.....	50
Figura 46, Immagine d'esempio con filtro sovrapposto (in verde)	51
Figura 47, Filtro generico applicato	51
Figura 48, Risultato dell'applicazione del filtro.....	51
Figura 49, Max-pooling per ridurre la dimensionalità da 4x4 a 2x2	51
Figura 50, Applicazione Flatten Layer	52
Figura 51, Struttura CNN utilizzata	53
Figura 52, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 53, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 54, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 55, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 56, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60

Figura 57, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 58, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 59, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 60, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 61, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 62, Segmentazione tramite CNN con confronto con maschera reale (bordo viola) ed osso compatto (bordo verde)	60
Figura 63, Taglio dell'immagine centrata nella maschera reale.....	68
Figura 64, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	69
Figura 65, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	69
Figura 66, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	70
Figura 67, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	70
Figura 68, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	71
Figura 69, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	71
Figura 70, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	72
Figura 71, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	72
Figura 72, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	73
Figura 73, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	73
Figura 74, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	74
Figura 75, Maschera reale in scala di grigi (con bordo verde) confrontata con maschera CNN (bordo viola) e DNN (bordo blu).....	74

Indice Grafici

Grafico 1, Risultati rete triangolare, accuracy	19
Grafico 2, Risultati rete triangolare, sensitivity	19
Grafico 3, Risultati rete triangolare, specificity	20
Grafico 4, Risultati rete triangolare, precision.....	20
Grafico 5, Risultati 8 strutture, dataset random, funzione di attivazione tansing, accuracy	22
Grafico 6, Risultati 8 strutture, dataset random, funzione di attivazione purelin, accuracy	22
Grafico 7, Risultati 8 strutture, dataset clustering, funzione di attivazione tansing, accuracy	23
Grafico 8, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, accuracy	23
Grafico 9, Risultati 8 strutture, dataset random, funzione di attivazione tansing, sensitivity	24
Grafico 10, Risultati 8 strutture, dataset random, funzione di attivazione purelin, sensitivity	24
Grafico 11, Risultati 8 strutture, dataset clustering, funzione di attivazione tansing, sensitivity	25
Grafico 12, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, sensitivity	25
Grafico 13, Risultati 8 strutture, dataset random, funzione di attivazione tansing, specificity	26
Grafico 14, Risultati 8 strutture, dataset random, funzione di attivazione purelin, specificity	26
Grafico 15, Risultati 8 strutture, dataset clustering, funzione di attivazione tansing, specificity	27
Grafico 16, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, specificity	27
Grafico 17, Risultati 8 strutture, dataset random, funzione di attivazione tansing, precision	28
Grafico 18, Risultati 8 strutture, dataset random, funzione di attivazione purelin, precision	28
Grafico 19, Risultati 8 strutture, dataset clustering, funzione di attivazione tansing, precision	29
Grafico 20, Risultati 8 strutture, dataset clustering, funzione di attivazione purelin, precision.....	29
Grafico 21, boxplot prima rete, pazienti 1-40 (training set).....	47
Grafico 22, boxplot prima rete, pazienti 41-50 (validation set)	47
Grafico 23, boxplot seconda rete, pazienti 1-40 (training set).....	48
Grafico 24, boxplot seconda rete, pazienti 41-50 (validation set)	48
Grafico 25, Confronto tra le due reti, pazienti 1-40 (training set).....	49
Grafico 26, Confronto tra le due reti, pazienti 41-50 (validation set)	49
Grafico 27, Risultati CNN, 30 epoche, 2 convoluzioni, ReLu.....	54
Grafico 28, Risultati CNN, 30 epoche, 2 convoluzioni, Tanh.....	54
Grafico 29, Risultati CNN, 30 epoche, 3 convoluzioni, ReLu.....	54
Grafico 30, Risultati CNN, 30 epoche, 3 convoluzioni, Tanh	54
Grafico 31, Risultati CNN, 40 epoche, 2 convoluzioni, ReLu	55
Grafico 32, Risultati CNN, 40 epoche, 2 convoluzioni, Tanh.....	55
Grafico 33, Risultati CNN, 40 epoche, 3 convoluzioni, ReLu	55
Grafico 34, Risultati CNN, 40 epoche, 3 convoluzioni, Tanh.....	55
Grafico 35, Risultati CNN, 50 epoche, 2 convoluzioni, ReLu.....	56
Grafico 36, Risultati CNN, 50 epoche, 2 convoluzioni, Tanh.....	56
Grafico 37, Risultati CNN, 50 epoche, 3 convoluzioni, ReLu	56
Grafico 38, Risultati CNN, 50 epoche, 3 convoluzioni, Tanh.....	56
Grafico 39, Risultati CNN, 50 epoche, 4 convoluzioni, ReLu	57
Grafico 40, Risultati CNN, 50 epoche, 4 convoluzioni, Tanh.....	57
Grafico 41, Risultati CNN, 60 epoche, 3 convoluzioni, ReLu	57
Grafico 42, Risultati CNN, 60 epoche, 3 convoluzioni, Tanh.....	57
Grafico 43, Risultati CNN, 60 epoche, 4 convoluzioni, ReLu	58
Grafico 44, Risultati CNN, 60 epoche, 4 convoluzioni, Tanh	58

Grafico 45, Risultati CNN, 70 epoche, 3 convoluzioni, ReLu.....	58
Grafico 46, Risultati CNN, 70 epoche, 3 convoluzioni, Tanh	58
Grafico 47, Risultati CNN, 70 epoche, 4 convoluzioni, ReLu.....	59
Grafico 48, Risultati CNN, 70 epoche, 4 convoluzioni, Tanh	59
Grafico 49, Boxplot risultati CNN, pazienti 1-40 (training set)	66
Grafico 50, Boxplot risultati CNN, pazienti 41-50 (validation set).....	66
Grafico 51, Media risultati, pazienti 1-40 (training set).....	67
Grafico 52, Media risultati, pazienti 41-50 (validation set)	67
Grafico 53, Confronto DNN-CNN, Precision, Pazienti 1-40 (training set).....	76
Grafico 54, Confronto DNN-CNN, Precision, Pazienti 41-50 (validation set)	76
Grafico 55, Confronto DNN-CNN, Dice Similarity Coefficient, Pazienti 1-40 (training set)	77
Grafico 56, Confronto DNN-CNN, Dice Similarity Coefficient, Pazienti 41-50 (validation set).....	77
Grafico 57, Confronto DNN-CNN, Recall, Pazienti 1-40 (training set)	78
Grafico 58, Confronto DNN-CNN, Recall, Pazienti 41-50 (validation set).....	78

Indice equazioni

Equazione 1, Fat Fraction.....	3
Equazione 2, Min-max scaling.....	11
Equazione 3, Modello matematico neurone.....	13
Equazione 4, Cross-Entropy.....	15
Equazione 5, Accuracy.....	17
Equazione 6, Specificity.....	17
Equazione 7, Sensitivity.....	17
Equazione 8, Precision.....	17
Equazione 9, Dice Similarity Coefficient (DSC)	31
Equazione 10, Recall (Re).....	31
Equazione 11, Precision (Pr).....	31
Equazione 12, Funzione Softmax.....	52
Equazione 12, Funzione di attivazione ReLu.....	53
Equazione 12, Funzione di attivazione tanh.....	53