

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Biomedica

Tesi di Laurea Magistrale

Individuazione del Midollo Osseo Attivo tramite algoritmi di Deep Learning in immagini TC



Relatrici:
Prof.ssa Gabriella Balestra
Prof.ssa Samanta Rosati

Candidato:
Giovanni Lo Bocchiaro

Ai miei cari

Indice

Abstract	1
1. Introduzione	3
1.1 Il midollo osseo	3
1.1.1 Tecniche di identificazione del RM	4
1.2 Reti Neurali Artificiali	7
2. Reti Neurali Convolutionali	10
2.1 Introduzione alle CNN	10
2.2 Applicazione delle CNN per classificare RM	13
2.3 Analisi del Data Set	13
2.4 Implementazione CNN con diversi layer di convoluzione	16
2.4.1 Risultati con 1 layer	19
2.4.2 Risultati con 2 layer	23
2.4.3 Risultati con 3 layer	28
2.4.4 Discussioni sui 3 modelli di CNN	33
2.5 Implementazione CNN con Training Set diviso al 50%	33
2.5.1 Risultati della CNN con Training Set e Test Set equi	35
2.6 Analisi dei dati tramite Dendrogrammi	39
2.6.1 Dendrogrammi con ROI RM e YM	41
3. U-Net	45
3.1 Introduzione a U-Net	45
3.2 Applicazione U-Net	46
3.2.1 Risultati Paziente 1	52
3.2.2 Risultati Paziente 2	55
3.2.3 Risultati Paziente 3	59
3.2.4 Risultati Paziente 4	62
3.2.5 Risultati Paziente 5	66
3.3 Applicazione U-Net con Data Set suddiviso	71
3.3.1 Primo Data Set	72
3.3.2 Secondo Data Set	105
3.3.3 Terzo Data Set	145
3.4 Applicazione U-Net con 6 pazienti	174
3.4.1 Risultati Paziente 1	177
3.4.2 Risultati Paziente 2	181
3.4.3 Risultati Paziente 3	185
3.4.4 Risultati Paziente 4	188
3.4.5 Risultati Paziente 5	191
4. Conclusioni	196
4.1 Conclusioni	196

4.2	Sviluppi Futuri	197
-----	---------------------------	-----

Abstract

Il midollo osseo ha un ruolo fondamentale nel bilanciare il livello di ematociti, questo perché una porzione di tale tessuto connettivo, chiamato midollo osseo rosso o midollo osseo attivo (RM), è responsabile della produzione delle cellule staminali ematopoietiche. Il RM è soggetto a gravi danni quando esposto alle radiazioni e.g. tossicità ematologica. Questo fenomeno può portare ad infezioni nel paziente che si sottopone a sessioni di radio-chemioterapia, con conseguente interruzione di esse e rendendo la terapia inefficace. L'obiettivo del presente progetto è di implementare un algoritmo di Deep Learning (DL) per l'individuazione del midollo osseo attivo direttamente dalle immagini di tomografia computerizzata (TC) del midollo osseo pelvico, luogo in cui risiede buona parte del RM in età adulta. Nello specifico, si è scelto di esaminare le immagini TC perché indicate dalle linee guide internazionali. Quindi, la totalità dei pazienti si sottopone a questo esame, con risparmio in termini di soldi e tempo. Tra gli algoritmi di Deep Learning a disposizione, si è optato per le Reti Neurali Convoluzionali (CNN) poiché, fin dalla loro introduzione nella prima metà degli anni 90, hanno presentato eccellenti risultati in molteplici campi, quali ad esempio il riconoscimento facciale o di cifre scritte a mano, come documentato negli studi svolti da LeCun et al.[1].

Il problema è stato affrontato seguendo due approcci, entrambi basati sulle CNNs. Nel primo caso, è stato costruito un data-set con 2200 regioni di interesse (ROI) di dimensione 5x5, ricavato dalle immagini TC di 5 pazienti. Il Data Set (suddiviso in Training Set e Validation Set) è stato usato, rispettivamente, per allenare la rete convoluzionale e validare le performance. Sono stati costruiti tre modelli, i quali differivano tra loro per il numero di Convolutional Layer. Nel caso di un singolo layer di convoluzione, i risultati nel Validation Set hanno mostrato un picco massimo intorno al 70% per Accuratezza (Acc), Sensitività (Sens), Specificità (Spec) e Precision (Pr). Nel caso di due e tre layer di convoluzione, invece, la rete tendeva al fenomeno di over-fitting poiché il Training Set presentava valori intorno al 90%, mentre nel caso del Validation Set, i parametri calcolati non superavano il 70%.

Nel secondo approccio, si è lavorato con un particolare tipo di rete convoluzionale, chiamata U-Net. La rete presenta un modello molto più complesso, in quanto consiste in un percorso di contrazione ed un percorso di espansione. Conseguentemente, rispetto al primo approccio, tale rete presenta un costo computazionale nettamente maggiore. Il Data Set era composto da 356 immagini TC di dimensioni 512x512 appartenenti a 5 pazienti. L'approccio ha presentato risultati migliori rispetto alla normale rete CNN. Nello specifico, non si sono presentati fenomeni di overfitting ed i valori calcolati di Dice Similarity Coefficient (DSC), Precision (Pr) e Recall (Re), nel Validation Set sono risultati intorno al 90% nella maggior parte dei casi. È stata svolta un'ulteriore prova, consistente nella suddivisione del Data Set originario in tre Data Set che presentavano la stessa quantità di slice contigue. Anche in questo caso si sono ottenuti valori molto promettenti di DSC, Pr e Re, intorno al 90% nella maggior parte dei casi, mantenendosi stabili rispetto al primo approccio. Tuttavia, questa prova ha evidenziato la

difficoltà della rete nel riconoscere il RM in alcune slice relative alle sezione anatomica dove è predominante la presenza dei femori, con valori di Dice inferiori al 70%. Sono stati ottenuti dei valori bassi (inferiori al 30%) in slice particolarmente critiche, poiché consistevano in aree limitate, strette e frastagliate di RM e con molta predominanza di sfondo nero. È stata svolta un'ultima prova consistente l'inserimento di un sesto paziente, aumentando la numerosità del Data Set (424 elementi). L'aggiunta di un nuovo paziente ha portato al miglioramento di alcune delle criticità sopracitate, infatti alcune slice che presentavano valori di DSC intorno al 60%, sono riuscite a raggiungere il 90%; in altri casi, invece, il valore di DSC è rimasto invariato o è variato leggermente.

1. Introduzione

1.1 Il midollo osseo

Il midollo osseo ha un ruolo importante nell'ematopoiesi, responsabile della produzione di ematociti.

Il risultato di questo processo sono, infatti, le cellule staminali ematopoietiche, differenziate in due sotto-famiglie di cellule ematiche:

1. cellule linfoidi, che includono le cellule T, cellule B e cellule NK;
2. cellule mieloidi, che includono monociti, macrofagi, neutrofilo, basofili, eosinofili, eritrociti e megacariociti.

Il midollo osseo è un tessuto connettivo, localizzato in diverse aree del corpo umano, come ad esempio le vertebre, lo sterno, le costole e il bacino. È possibile, come in Figura 1.1, distinguerne due tipologie:

- il midollo osseo rosso (RM), al quale ci si può riferire anche come midollo osseo attivo o midollo osseo ematopoietico. È dove si verifica il processo di ematopoiesi, portando alla produzione di nuove cellule del sangue e l'eliminazione di quelle vecchie. La sua quantità si presenta elevata durante i primi anni di vita, ma col tempo viene sostituito dal midollo osseo giallo;
- il midollo osseo giallo (YM), è dove gli adipociti adempiono al compito di conservazione dei lipidi per usarli, in seguito, come fonte di energia.

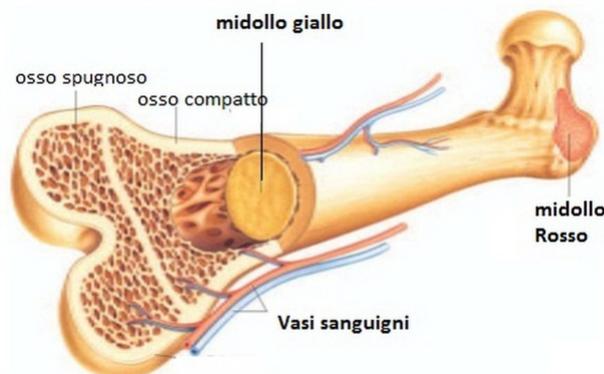


Figura 1.1: Rappresentazione dei due tipi di midollo osseo giallo e rosso.

L'individuazione dell'area di RM è importante nell'ambito dei trattamenti clinici di diverse malattie, come ad esempio il cancro al midollo osseo pelvico. Infatti, nel bacino e nella spina lombare di un adulto risiede più della metà di RM totale presente nel corpo umano[2].

É largamente risaputo che il cancro, spesso, necessita di molte sessioni programmate di radio-chemioterapia. Nel corso di questo trattamento, il midollo osseo ematopoietico potrebbe essere compromesso dall'esposizione alle radiazioni portando ad una tossicità ematologica, come ad esempio la neutropenia, una forma di leucopenia, che consiste nella riduzione dei neutrofili sotto il valore di soglia [3] [4] [5] [6] [7]. Ciò può diventare un fattore limitante durante il sopracitato trattamento, poiché potrebbe portare ad infezioni e conseguente ricovero del paziente. Ed infine, comportare uno spostamento delle sessioni di radio-chemioterapia o la loro interruzione, andando a ledere l'efficacia del trattamento.

Un'alternativa potrebbe essere quella di effettuare una radio-chemioterapia con intensità focalizzata, in modo da non rendere inefficace la terapia ed evitare di compromettere la zona di RM.[8].

1.1.1 Tecniche di identificazione del RM

É chiaro quanto sia essenziale trovare una strategia appropriata per minimizzare il più possibile i danni al paziente. Negli anni, sono stati condotti diversi studi volti a sviluppare nuove tecniche di riconoscimento delle aree di RM[9]:

- Individuazione tramite risonanza magnetica (MRI), la quale si basa sulla differenza morfologica tra RM e YM. Infatti, il primo è molto più vascolarizzato del midollo osseo giallo, mentre il secondo è composto da tessuto adiposo. Nel dettaglio, queste differenze possono essere evidenziate dal rapporto di grasso rispetto l'acqua, il quale definisce l'indice di frazione di grasso. Tuttavia, questa tecnica non misura il livello di attività del midollo osseo rosso e nel caso in cui sia presente un tessuto che ha la stessa frazione di grasso del RM, ma nessuna attività ematopoietica, ciò creerebbe un'incertezza nella misura di aree di midollo osseo attivo. Un altro svantaggio di questo tipo di esame, è l'elevato costo sia in termini di soldi che di tempo [2];
- Tomografia a singola emissione di positrone (SPECT), la quale è un approccio funzionale basato sull'isomero nucleare Tc-99m [10] che attrae i macrofagi correlati al midollo osseo rosso e permette di ottenere una mappa tridimensionale della distribuzione di RM. Tuttavia, solo una piccola frazione dell'intera dose di radiofarmaco è assorbita dal sistema reticolare endoteliale del RM, mentre la maggior parte viene assorbita da fegato e reni. Inoltre, questo approccio si presta ad una bassa generalizzazione poiché si basa su parametri specifici del paziente[11].
- (¹⁸F)Fluoro-2-deossi-2-D-glucosio (FDG)-PET, si basa sul un radioisotopo FDG, che è simile al glucosio. Le cellule metaboliche attive presenti nel RM sono sensibili al FDG e conseguentemente la loro assenza permette di differenziare il midollo osseo rosso da quello giallo. A differenza del glucosio, il FDG non è catabolizzato, ciò porta al suo decadimento in ¹⁸F e conseguente emissione di positroni. É importante notare che questo radiofarmaco ha un bassissimo tempo di emivita, dunque per poterlo sfruttare è necessario che la sua produzione avvenga in un ciclotrone vicino ad una macchina PET [12]. Un ulteriore svantaggio è che

l'esame con approccio $^{18}\text{FDG-PET}$ è considerato un esame opzionale dalle linee guida internazionali per quanto concerne il processo diagnostico e dunque non viene sottoposto a tutti i pazienti [13].

- Radiomica, consiste nell'estrazione di una serie di parametri che riflettono le caratteristiche morfologiche dell'immagine in modo da fornire più dati per ottenere migliori performance da parte degli algoritmi predittivi. Nei recenti studi [13] [14], l'approccio mediante radiomica veniva affiancato dall'utilizzo di immagini di tomografia computerizzata (TC). Quest'ultime sono ottenute da un esame più semplice e meno costoso rispetto agli esami PET e MRI. Dunque, ottenute le singole slice tramite TC, si estraggono tutte le feature matematiche calcolabili considerando i pixel. In particolare, nel precedente studio [14] sono state estratte un totale di 36 feature matematiche, le quali consistono in:

1. 4 feature statistiche del primo ordine:

- Media;
- Deviazione Standard;
- Skewness;
- Kurtosis.

2. 32 feature statistiche del secondo ordine (texture feature):

- 22 texture feature calcolate usando la Matrice Grey Level Co-occurrence (GLCM);
- 5 texture feature calcolate usando la Matrice Grey Level Dependence (GLDM);
- 5 texture feature calcolate usando la Matrice Grey-Level Run Length (GLRLM).

Al processo di estrazione delle feature ha seguito il processo di selezione volto ad ottimizzare i classificatori. L'algoritmo di ottimizzazione usato si chiama Algoritmo Genetico, il cui obiettivo è quello di identificare, tra le 36 feature, quelle più adatte allo scopo. L'algoritmo genetico si basa sul concetto biologico di selezione naturale ed evoluzione. L'insieme di soluzioni iniziali viene ricombinato nel corso delle iterazioni, fino ad ottenere una soluzione ottima. Ovviamente, essendo questo processo di ricombinazione casuale, pur partendo dalle stesse condizioni iniziali, il risultato finale può variare. Sono state ottenute 5 soluzioni per ogni classificatore utilizzato [14] e che rappresentano il sotto-gruppo di feature ottimali.

Dettagliatamente, si sono ottenuti vettori binari dove il bit "1" rappresenta una feature considerata ed il bit "0" una feature scartata. L'algoritmo genetico è stato programmato per iterare 100 volte al fine di ottenere una soluzione; è stata inoltre implementata una condizione di stop che permetteva l'interruzione delle iterazione se dopo 30 iterazioni il valore di fitness non variava. Tale valore numerico definisce l'idoneità della soluzione al problema di ottimizzazione. Essendo la fitness definita come nell'Equazione 1.1 [14], il problema di ottimizzazione si basava nel trovare un valore di minimo.

$$f = 1 - acc + 0.3 \cdot |sens - spec| \quad (1.1)$$

Dove il parametro *acc* si riferisce all'accuratezza, *sens* alla sensitività e *spec* alla specificità. Questi tre parametri vengono definiti dalle Equazione 2.3 2.5 e 2.6.

I classificatori utilizzati in questo studio sono:

1. Decision Tree (DT), consiste in un classificatore ad albero in cui ogni nodo definisce una feature, mentre gli archi rappresentano determinati valori della specifica feature a cui corrispondono. La struttura viene modellata durante l'allenamento. Le condizioni che permettono di fermare l'allenamento sono due: o training set risulta partizionato in modo tale che in ogni partizione vi sono elementi appartenenti alla stessa classe oppure non vi sono più variabili utilizzabili per partizionare. Infine, la classe del nodo deriva dalla classe maggiormente rappresentata all'interno della partizione. Questo classificatore non necessita di ottimizzazione dei parametri, quindi la soluzione da far evolvere consiste in un vettore da 36 elementi;
2. K-Nearest Neighbors (KNN), consiste in un classificatore basato sul calcolo della distanza tra gli elementi presenti all'interno del training set. In particolare, di ogni elemento si calcola la distanza con gli altri elementi all'interno del training set e si prendono le prime K posizioni in ordine crescente. In questo caso, la soluzione del GA prevede un vettore che, oltre i 36 elementi che definiscono le feature, vi siano altri 6 bit in coda che definiscono i 2^6 valori di K intorno ad un valore iniziale precedentemente calcolato [14];
3. Neural Network (NN), di cui si parlerà in maniera più approfondita nel Capitolo 1.2. In questo caso, l'algoritmo genetico si occupa di fornire il numero di hidden layer tra input ed output;
4. Support Vector Machine (SVM), è un classificatore specializzato nella classificazione binaria. Permette di rappresentare ogni elemento del Training Set in uno spazio multi-dimensionale e, successivamente, traslarlo in un nuovo spazio multi-dimensionale di dimensione maggiore rispetto al precedente. Mediante un iper-piano si ottiene la classificazione; esso massimizza la distanza tra gli elementi più vicini al piano di separazione, ottenendo una separazione che identifica le due classi a cui possono appartenere gli elementi. L'algoritmo genetico si occupa di ottimizzare la funzione kernel che permette di traslare gli elementi nel nuovo spazio multi-dimensionale ed il parametro C che regola il margine del piano di separazione e l'errore di classificazione.

Quest'ultimo approccio alla radiomica che sfrutta immagini TC ha prodotto dei risultati molto promettenti per tutti e quattro i classificatori [14]. Ciò avvalorava l'ipotesi di sfruttare la tomografia computerizzata per algoritmi di riconoscimento del midollo osseo rosso. La TC, è un esame molto comune e a cui, dunque, vengono sottoposti tutti i pazienti di cancro prima dell'inizio di un trattamento di radio-chemioterapia. Le immagini TC vengono ottenute considerando la proporzionalità della radiopacità e l'intensità dei tessuti, in questo modo si può fornire una rappresentazione degli organi interni e dei tessuti del corpo umano[15] [16] [17]. È tuttavia doveroso tenere conto che, a parità di dose, la qualità dell'immagine PET è superiore rispetto a quella ottenuta tramite TC, considerando che ambedue sottopongono il paziente ad un'alta esposizione di radiazioni.

I promettenti risultati ottenuti lavorando con le immagini di TC ed il fatto che quest'ultima sia un esame molto meno dispendioso in termini di soldi e tempo, motivano

ulteriormente lo studio di un'implementazione di un algoritmo di Deep Learning che si basi su tali immagini per l'individuazione di zone di midollo osseo rosso.

1.2 Reti Neurali Artificiali

L'apprendimento basato su reti neurali artificiali (ANN) è una tecnica applicabile in diversi campi, tra cui quello biomedico. Il suo approccio si basa sulla simulazione delle funzioni biologiche del cervello umano tramite nodi, chiamati neuroni. [18]. Un'architettura ANN base è mostrata nella Figura 1.2 e consiste in un input layer, pochi hidden layer e un output layer.

L'Equazione 1.2 descrive il processo effettuato da un neurone in una ANN generica.

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \quad (1.2)$$

dove a_j^l definisce l'attivazione del j-esimo neurone al l-esimo layer, σ è una generica funzione di attivazione, w_{jk}^l è il peso che va dal k-esimo neurone nel layer precedente al j-esimo neurone nel l-esimo layer e b_j^l è il bias associato al j-esimo neurone del l-esimo layer.

Peso e bias sono due parametri da cui dipende la funzione di attivazione: il primo controlla la sua ripidezza, mentre il secondo permette di aggiungere un certo ritardo all'attivazione. La funzione di attivazione permette di definire il tasso di potenziale d'azione di un neurone.

Peso e bias vengono decisi tramite la minimizzazione di una funzione di costo, come ad esempio la funzione di costo quadratica definita nell'Equazione 1.3.

$$C(w, b) = \frac{1}{2n} \sum_x \|y(x) - a(x, w, b)\|^2 \quad (1.3)$$

Tale obiettivo, può essere raggiunto utilizzando la tecnica della discesa del gradiente, la quale consiste nel partire da un punto iniziale x_0 di una funzione multi-dimensionale $f(x_i)$, dove inizialmente $x_i = x_0$. Successivamente, calcolarne il gradiente in modo da ottenere la direzione di discesa che viene moltiplicata per un coefficiente che indica il passo di discesa, ed infine, aggiornare il punto in relazione a questo prodotto e ricalcolare il gradiente nel nuovo punto fino a quando non è nullo.

Il procedimento che porta all'apprendimento è chiamato algoritmo di apprendimento e consiste nella modifica ordinata dei pesi della rete in modo da ottenere l'obiettivo desiderato. Un algoritmo di apprendimento basato sulla discesa di gradiente molto utilizzato è chiamato Backpropagation. Tale algoritmo calcola la somma degli errori quadratici per misurare il tasso di apprendimento della rete. Ciò significa che l'errore tra l'attuale output della rete e l'output desiderato è calcolato durante la fase di addestramento e propagato indietro in modo da aggiustare i pesi.

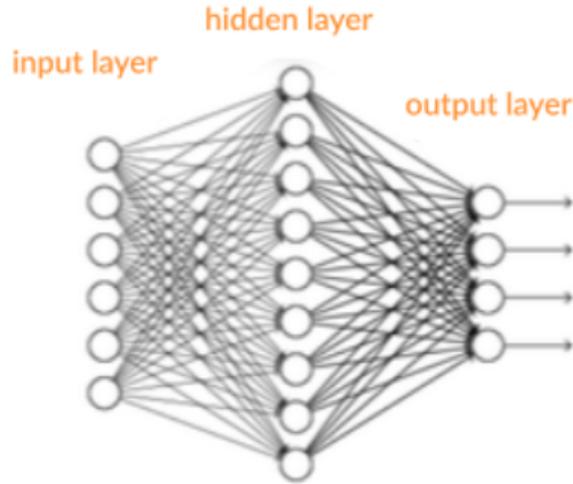


Figura 1.2: Struttura base di una Rete Neurale Artificiale (ANN).

I neuroni sono caratterizzati dalla loro funzione di attivazione. Il primo modello di neurone artificiale è il Percettrone, un tipo di classificatore binario caratterizzato da un numero n di pesi aggiustabili e da un threshold; se il prodotto scalare dei pesi con l'input supera tale valore di soglia, allora il classificatore darà in output 1, in caso contrario 0.

Questo tipo di modello veniva utilizzato intorno agli anni '40 dove, non avendo le tecnologie attuali, l'obiettivo di minimizzazione dell'errore di classificazione era ottenuto mediante un processo euristico di aggiornamento dei pesi. Questa tecnica ricorda l'algoritmo di discesa del gradiente, ma non possiede le stesse proprietà. Infatti, quest'ultimo algoritmo è basato sul calcolo della derivata e quindi non può funzionare su una funzione che ha un punto di non derivabilità e nei suoi intervalli differenziabili deriva a 0 [19].

Tutti i nodi sono connessi tra loro e si trasmettono informazioni, simulando le sinapsi. Di conseguenza, l'intelligenza artificiale (IA), attraverso l'implementazione di un modello matematico specifico per il *Machine Learning*, è capace di risolvere problemi di elevata complessità, quali ad esempio la classificazione.

Ci sono diversi tipi di algoritmi di *machine learning*, i più diffusi sono:

- Apprendimento supervisionato, dove all'ANN viene fornito un Training Set consistente un vettore input e le corrispondenti classi (ovvero gli output desiderati). Questo viene dato in input ripetutamente alla rete, dove ogni ciclo indica un'epoca. I pesi vengono modificati in relazione all'errore tra l'output della rete e le reali classi; infatti, l'obiettivo è quello di ottenere una combinazione di pesi in grado di minimizzare l'errore;
- Apprendimento non supervisionato, dove all'ANN viene fornito soltanto il Training Set consistente un vettore input, mentre le effettive classi sono sconosciute. L'obiettivo di tale approccio è quello raggruppare gli elementi del Training Set in cluster omogenei.

Gli algoritmi di *Deep Learning* (DL), sono una sotto-famiglia degli algoritmi di *machine learning*. Presentano la capacità di apprendere in maniera più approfondita e

automatica le caratteristiche dei dati in ingresso poiché la loro architettura, mostrata in Figura 1.3, consiste in input ed output layer connessi tra loro attraverso una lunga catena di hidden layer.

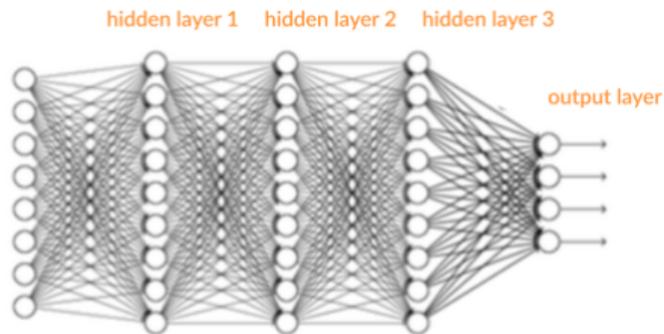


Figura 1.3: Struttura base di una Rete Neurale Deep Learning.

Gli algoritmi di DL si adattano perfettamente al campo biomedico, ad esempio per quanto concerne il riconoscimento di specifiche aree e fornendo, conseguentemente, immagini mediche segmentate in maniera automatizzata. Ciò, permette di supportare medici e radiologi nella diagnosi e nella scelta del miglior percorso terapeutico per il paziente [20]. Alcuni esempi sono l'impiego nella radioterapia [21] [22], nella rimozione di tessuti soffici da immagini PET/MRI [23] [24] e radiomica [25].

2. Reti Neurali Convoluzionali

2.1 Introduzione alle CNN

Le reti neurali convoluzionali (CNN) sono un modello che si basa sulla tecnica del Deep Learning. L'obiettivo di questa tesi è quello di sviluppare un modello di CNN che, dopo l'allenamento attraverso un Data Set di immagini TC, opportunamente pre-processate, sia in grado di riconoscere l'area di RM in maniera automatica.

Le CNN, negli ultimi anni, hanno ottenuto delle brillanti performance nel risolvere numerosi problemi, come ad esempio la classificazione di immagini [26] [27]. Più nello specifico, le CNN sono in grado di apprendere sfruttando un Training Set molto numeroso, mantenendo contemporaneamente un costo computazionale minimo, questo anche grazie all'evoluzione e sviluppo di GPU sempre più performanti. Nonostante questa vantaggiosa premessa, ad oggi, è ancora difficile comprendere a pieno il comportamento complesso di tali reti.

In riferimento alla Figura 2.1, una tipica struttura CNN è composta da:

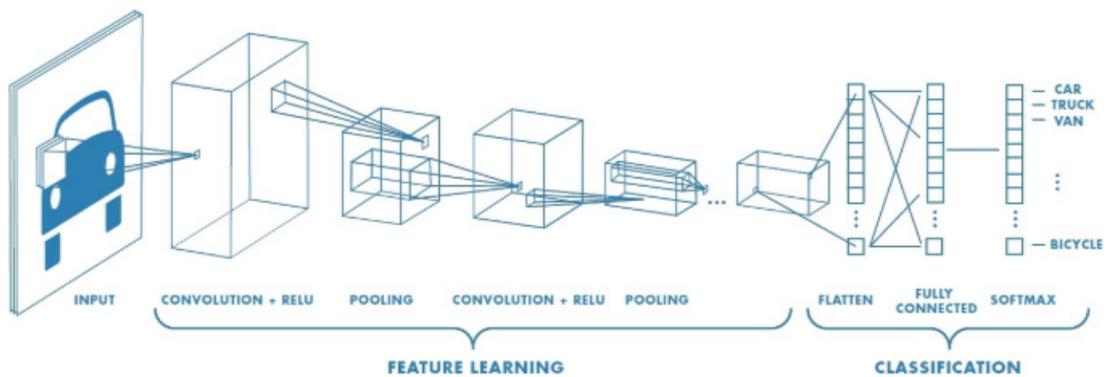


Figura 2.1: Tipica struttura CNN [28]

1. Convolutional Layer, prende in ingresso l'immagine input (oppure le attivazioni del layer precedente) ed effettua la convoluzione con una matrice di filtri avente dimensione preimpostata; in tal modo è possibile estrarre le feature dall'input. Un altro parametro da impostare è il passo della matrice di filtri, ovvero di quanto il filtro viene spostato. Può capitare che la matrice di filtri non combaci perfettamente con l'immagine, in quel caso è possibile utilizzare delle tecniche di *Padding* come l'aggiunta di zeri o, in alternativa, ignorare i pixel in eccesso. Il risultato di questo layer è la mappa delle feature. Una rappresentazione grafica di questo processo è mostrata nelle Figure 2.2 2.3 e 2.4:

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	...
...

Porzione di una generica immagine Input con il verde filtro sovrapposto

Figura 2.2: Il primo passo è la sovrapposizione del filtro.

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

Generico filtro 3x3

Figura 2.3: Il secondo passo è l'applicazione del filtro.

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

Generico output dopo l'applicazione del filtro

Figura 2.4: Il risultato di questo processo è la feature delle mappe.

2. Activation Layer, la mappa delle feature in ingresso viene elaborata con una funzione di attivazione non lineare. Generalmente viene utilizzata la funzione di unità lineare rettificata (ReLU) ed è definita come $ReLU(x) = \max(0, x)$;
3. Pooling Layer, il quale si occupa del processo di sotto-campionamento che porta ad una riduzione della dimensionalità di ogni mappa di feature. Essenzialmente si ottiene una sotto-matrice più piccola della mappa delle feature, quest'ultima viene elaborata seguendo tre approcci:
 - Max-Pooling, che consiste nel calcolare il valore più grande all'interno di ogni mappa di feature. Tale tecnica è rappresentata nella Figura 2.5;
 - Average-Pooling, che consiste nel calcolare il valore medio all'interno di ogni mappa di feature;
 - Sum-Pooling, che consiste nel calcolare la somma dei valori all'interno di ogni mappa di feature.

Poiché il Pooling Layer introduce una maggiore complessità della rete, lo stesso effetto di sotto-campionamento può essere ottenuto applicando un Convolutional Layer con un passo molto alto;



Figura 2.5: Applicazione del Max-Pooling considerando una dimensione 2x2.

4. Flatten, il quale si occupa di preparare i dati alla predizione. Nello specifico, vettorizza la struttura dei dati facendo perdere, così, il riferimento spaziale. Il risultato di questo processo è un vettore di dati che viene inviato all'output layer, un esempio è presentato nella Tabella 2.6;

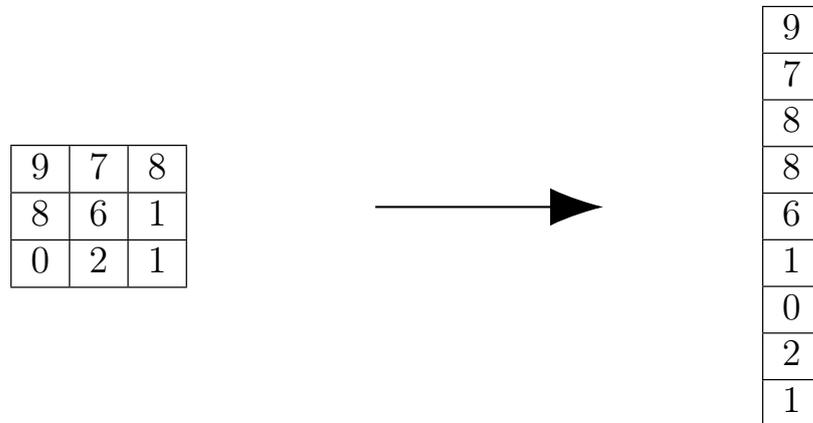


Figura 2.6: Applicazione del Flatten.

5. Dense Layer, implementa l'operazione $output = activation((input, kernel) + bias)$, dove il kernel rappresenta la matrice dei pesi. Indicando la funzione di attivazione e la dimensione del layer, fornisce un output. Nel caso della CNN viene posto come ultimo layer quindi avrà 1 sola unità in modo da fornire un solo output.

Altri elementi comuni nelle strutture CNN sono:

- Dropout Layer, consiste nel rimuovere casualmente neuroni durante la fase di allenamento, di modo da avere maggiore indipendenza fra i pesi;
- Batch Normalization Layer, 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 batch di allenamento.

2.2 Applicazione delle CNN per classificare RM

Gli algoritmi di Deep Learning, in particolare le CNN, hanno evidenziato delle prestazioni promettenti, nello specifico si sono ottenuti degli ottimi risultati per quanto concerne il problema di identificazione nelle immagini [1].

Difatti, le CNN essendo intrinsecamente elementi non lineari, si adattano a questo specifico problema; ciò porta alla diretta conclusione che tali reti possono essere implementate per l'individuazione del midollo osseo attivo. Difatti, l'intensità dei pixel e il contorno hanno un peso predominante nella classificazione del midollo osseo ematopoietico.

Tuttavia, nell'utilizzo delle reti neurali convoluzionali, c'è un enorme svantaggio da tenere in considerazione: per poter ottenere risultati rilevanti, si necessita di un Training Set numeroso. Questo è un requisito complicato da soddisfare nel caso di applicazione di CNN per classificare midollo osseo attivo, e in generale nel campo biomedico, poiché è difficile reperire elementi del Data Set a causa di problematiche relative all'etica e la privacy del paziente.

Come menzionato sopra, inoltre, non si è ancora compreso pienamente il processo dietro l'elaborazione effettuata dalle CNN e questo comporta uno svantaggio in fase di costruzione del modello. Per questo una soluzione alternativa potrebbe essere quella di estrarre e visualizzare le *feature* ottenute dalle matrici sopra menzionate nel Capitolo 2.1. In questo modo, si potrebbe comprendere meglio il processo dietro le CNN ed individuare i parametri da ottimizzare. Purtroppo, solo il primo layer di convoluzione fornisce una proiezione spaziale dei pixel, lasciando i restanti layer difficili da interpretare [1].

2.3 Analisi del Data Set

L'obiettivo della tesi è di allenare un modello di rete convoluzionale al riconoscimento del midollo osseo attivo partendo da immagini TC di pazienti affetti da tumori della pelvi. Ciò, per programmare delle sessioni di trattamenti clinici che permettano di evitare le zone di RM.

In questo caso si sono utilizzate immagini appartenenti a pazienti affetti da carcinomi a cellule transizionali e/o carcinomi a cellule squamose in cura presso il Dipartimento Oncologico di Torino. L'esame di tomografia computerizzata è stato sottoposto a tutti i pazienti usando uno scanner TC Philips "BigBore" con spaziatura per i pixel di 0.93 mm e 3 mm per ogni slice. È importante precisare che tutti i pazienti erano in posizione supina con supporti per caviglie e ginocchia. Le slice ottenute sono state successivamente manualmente divise da un radioterapista oncologico in tre sotto-regioni [13]:

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

In questo studio è stata considerata solo la sotto-regione LSBM. Il Data Set utilizzato, è lo stesso ottenuto nei precedenti lavori [14]. A partire da questo, prendendo in considerazione le slice TC di 5 pazienti sono state estrapolate, in maniera casuale, 2200 ROI di dimensione 5x5 parzialmente sovrapposte. Tutte le immagini TC hanno attraversato una fase di pre-processing consistente la rimozione dell'osso compatto, poiché in esso non risiede alcuna porzione di midollo osseo; in aggiunta, sono stati rimossi anche i tessuti non rilevanti ed il tavolino medico. Questo processo ha permesso di ottenere le maschere binarie del midollo osseo e conseguentemente le maschere binarie contenenti RM[14]. Ovviamente, la differenza tra le due maschere ha permesso di ricavare anche le maschere binarie YM, le quali al momento non sono state utilizzate, ma sono state sfruttate nel Capitolo 3.3. Il risultato di questo processo è visibile nelle Figure 2.9a e 2.9b

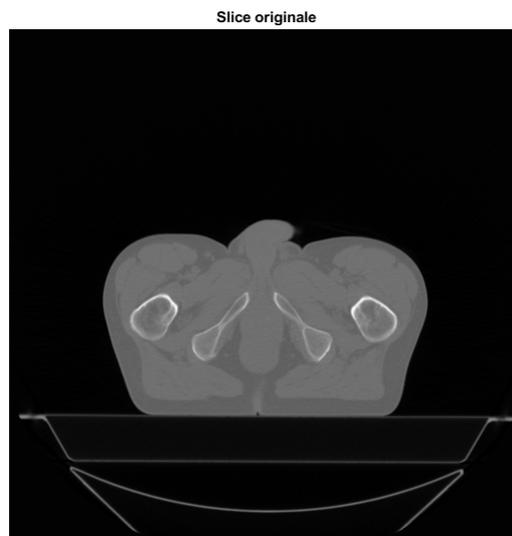
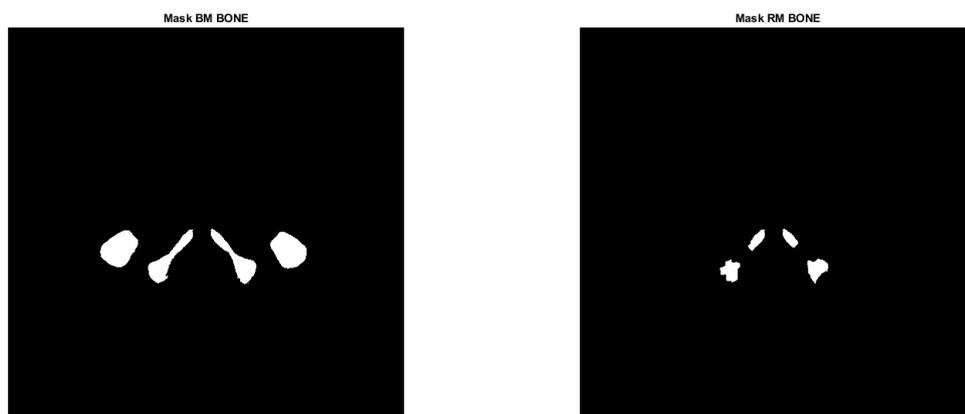
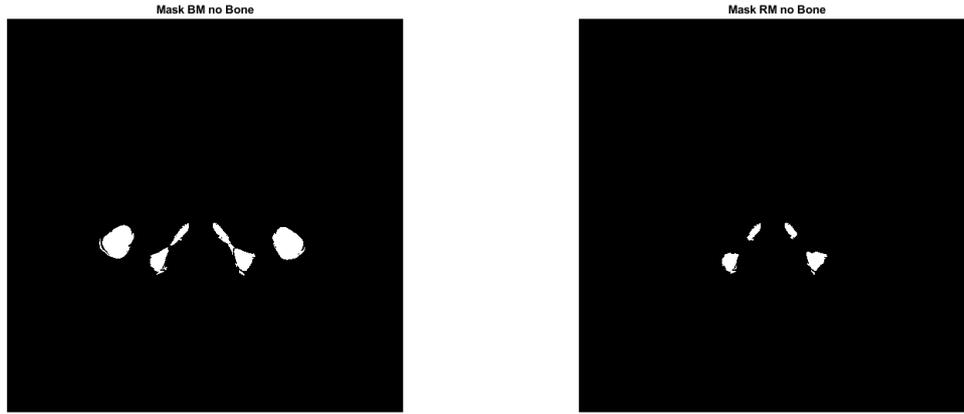


Figura 2.7: Slice originale



(a) Maschera binaria di BM con osso compatto (b) Maschera binaria di RM con osso compatto

Figura 2.8



(a) Maschera binaria di BM senza osso compatto (b) Maschera binaria di RM senza osso compatto

Figura 2.9

Nel Data Set utilizzato, per ogni paziente sono state prese cinque slice, ad eccezione del paziente #2 dal quale ne sono state estratte solamente due. Per ogni slice sono state prese 50 ROI identificate come RM (con etichetta "1") e 50 ROI identificate come YM (con etichetta "0"). La Tabella 2.1 indica le slice scelte per ogni paziente.

Paziente	Slice #1	Slice #2	Slice #3	Slice #4	Slice #5
1	-57.5	-72.5	-81.5	-24.5	-105.5
2	-125	-54	\	\	\
3	-160	-166	-127	-88	-154
4	-29.5	-32.5	-35.5	-89.5	-154
5	-30	-87	-72	-78	-81

Tabella 2.1: Tabella che riassume le slice prese per ogni paziente presente nel Data Set

Nel Data Set sono stati aggiunti alcuni campi che permettono una migliore rintracciabilità dell'origine delle ROI. Precisamente, sono stati inseriti gli indici del paziente, l'indice della ROI, la slice corrispondente, le coordinate x e y del pixel centrale nell'immagine originale e la classe di appartenenza. Un esempio è indicato nella tabella 2.2.

Paziente	indice ROI	Slice	Coor X	Coor Y	Classe
1	37	-57.5	236	340	0
1	415	-57.5	252	348	0
1	844	-57.5	271	339	0
1	845	-57.5	272	339	0
1	454	-57.5	253	351	0

Tabella 2.2: Esempio del Data Set in cui si può leggere l'indice del paziente, l'indice della ROI, la slice, le coordinate x e y del pixel centrale e la classe.

Prima di estrapolare e salvare le ROI dalle immagini, di modo da poterle utilizzare come input della CNN, le 22 slice prese in considerazione hanno subito una fase di

pre-processing: l'intera immagine è stata normalizzata tramite min-max scaling, come definito nell'Equazione 2.1

$$p' = \frac{p - \min(V)}{\max(V) - \min(V)} \quad (2.1)$$

dove p' definisce il valore normalizzato di intensità del pixel, p il valore originale di intensità del pixel, $\max(V)$ indica il valore massimo di intensità tra i pixel e $\min(V)$ individua il valore minimo di intensità tra i pixel. È doveroso osservare che tale normalizzazione è stata fatta considerando solo i pixel della porzione di midollo osseo, infatti l'immagine è stata mascherata sfruttando maschere come in Figura 2.9a e quindi trascurando tutto lo sfondo nero che poteva compromettere i valori normalizzati. L'algoritmo di min-max scaling permette di normalizzare i valori di intensità dei pixel nel range $[0,1]$, per permettere alle funzioni di attivazione dei layer di lavorare in maniera più performante. Si è deciso di procedere seguendo due approcci:

1. Applicare l'algoritmo individuando l'intensità massima e minima della singola slice;
2. Applicare l'algoritmo individuando l'intensità massima e minima tra tutte le slice del paziente;

Ciò ha portato ad una lieve variazione dei valori standardizzati, poiché considerando l'intero paziente, le sezioni anatomiche più visibili in una slice rispetto ad un'altra influenzano il valore di normalizzazione. Un esempio di tale variazione è osservabile nelle due Tabelle 2.3 2.4 che mostra la stessa matrice 5x5 di una porzione della stessa slice considerando i due approcci di normalizzazione:

0.185	0.228	0.264	0.293	0.389
0.231	0.273	0.228	0.261	0.435
0.311	0.266	0.217	0.368	0.488
0.505	0.496	0.545	0.671	0.632
0.695	0.710	0.717	0.740	0.732

Tabella 2.3: Valori di pixel normalizzati considerando l'intero paziente

0.159	0.098	0.043	0.031	0.081
0.212	0.124	0.160	0.085	0.135
0.286	0.147	0.232	0.180	0.127
0.361	0.366	0.467	0.411	0.401
0.535	0.571	0.648	0.638	0.655

Tabella 2.4: Valori di pixel normalizzati considerando la singola slice

2.4 Implementazione CNN con diversi layer di convoluzione

Le 2200 ROI di dimensione 5x5 indicate nel Data Set sono state estratte a partire dalle slice normalizzate, formando così due nuovi Data Set: il primo con le ROI aventi normalizzazione relativa al paziente e il secondo con ROI aventi normalizzazione relativa alla singola slice. Inoltre, in fase di estrazione è stato tenuto conto della classe di appartenenza delle ROI, ricordando che quelle che definivano una zona RM erano identificate con la classe "1", mentre quelle che definivano una zona di YM erano identificate con la classe "0". In questa implementazione tutte le ROI prese sono all'interno del midollo osseo, quindi non si hanno ROI che sono completamente sfondo.

Le CNN sono state implementate usando il linguaggio di programmazione Python in ambiente di sviluppo Spyder versione 3.6 ed in particolare usando il pacchetto API

Keras. Poiché scritte in questo linguaggio estremamente popolare e libero, è possibile eseguire il codice su una qualunque macchina, di conseguenza questo rende possibile l'esecuzione di questa implementazione anche su un generico servizio di cloud computing. Questa operazione è stata necessaria per poter allenare U-Net, dato il suo alto costo a livello computazionale, come viene descritto nella Sezione 3.2.

Una volta caricato uno dei due Data Set contenenti le ROI con le due differenti normalizzazioni e le classi corrispondenti ad esse, si è proceduto estraendo un 15% casuale del Data Set che è stato estratto per costruire il Validation Set. Il restante 85% è stato diviso rispettivamente in 70% per il Training Set e 15% per il Test Set da utilizzare in tempo reale durante il ciclo di apprendimento della rete. Infatti, si è deciso di utilizzare, durante la fase di "fit" del modello, la proprietà shuffle per mescolare costantemente gli elementi del Training Set e del Test Set. Questa scelta è stata fatta per tentare di trovare un compromesso con la bassa quantità di elementi del Data Set. È doveroso aggiungere che i tre modelli, che verranno successivamente presentati nel Capitolo 2.4, condividono lo stesso Validation Set; pertanto, tra gli allenamenti, soltanto il Training Set è stato variato a seguito del comando shuffle.

All'ingresso delle CNN quindi si hanno le ROI di dimensione 5x5; per analizzare al meglio le performance delle CNN sono state create tre architetture:

1. Architettura CNN con un solo layer di convoluzione avente 30 filtri di dimensione 3x3 e funzione di attivazione ReLu;
2. Architettura CNN con due layer di convoluzione in sequenza aventi, rispettivamente, 20 e 30 filtri di dimensione 3x3 e funzione di attivazione ReLu;
3. Architettura CNN con tre layer di convoluzione in sequenza aventi, rispettivamente, 20, 30 e 40 filtri di dimensione 3x3 e funzione di attivazione ReLu;

In questo caso, non è stato inserito il layer di Pooling descritto nel capitolo 2.1, poiché avendo delle ROI di dimensione 5x5 e dunque molto piccole, effettuare un sub-sampling sarebbe risultato controproducente. Prima di fornire i dati al layer di flatten, si è scelto di inserire un layer di Dropout. Infatti, come già menzionato, il Data Set non è molto numeroso, quindi si rischia che la rete non riesca a generalizzare. Per evitare questo, il layer di Dropout permette in fase di allenamento di annullare in modo casuale con una probabilità impostabile, un certo numero di connessioni e conseguentemente, di far ricevere ai layer successivi meno informazioni. Per questa specifica applicazione il layer Dropout è stato impostato con un rate del 50%, ovvero la metà delle connessioni veniva disattiva ad ogni iterazione. Infine, come già menzionato nella Sezione 2.1, si hanno il layer Flatten ed il layer Dense, un particolare tipo di layer che presenta un solo neurone, di modo da ricevere solo un dato alla volta. In questo caso, il layer sfrutta come funzione di attivazione la funzione sigmoide, che viene descritta nell'Equazione 2.2:

$$\sigma(x) = \frac{e^x}{1 + e^x} \quad (2.2)$$

Questa funzione di attivazione è stata scelta poiché presenta un output nel range [0,1], è utilizzabile visto il basso numero di layer ed inoltre basandosi su una funzione esponenziale, ha un costo computazionale molto basso.

Come algoritmo di ottimizzazione è stato usato Adam, il quale ha il vantaggio di essere la combinazione di due varianti della discesa stocastica del gradiente:

- Discesa del gradiente stocastica adattiva, in questa variante, invece di avere un singolo tasso di apprendimento, ne viene salvato uno per ogni peso;
- Propagazione del valor medio quadratico, che permette ai tassi di apprendimento di adattarsi al valore del gradiente specifico del peso. Dunque, questo approccio ottiene dei buoni risultati per problemi non stazionari.

La rete in tutte e 3 le sue strutture è stata allenata per 10 volte ognuna per 200 epoche in modo da analizzarne la stabilità. Inoltre come indicatori delle performance si sono calcolati loss e accuratezza. La prima è stata calcolata tramite la binary cross-entropy poiché si tratta di una classificazione binaria, invece, la seconda è definita dall'Equazione 2.3:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

- TP indica il numero di ROI predette come RM e che effettivamente lo erano;
- TN indica il numero di ROI predette come YM e che effettivamente erano lo erano;
- FP indica il numero di ROI predette come RM ma che in realtà erano YM;
- FN indica il numero di ROI predette come YM ma che in realtà erano RM.

Sia per Training Set che per Validation Set sono stati calcolati altri tre parametri quali:

1. Precisione, definita dall'Equazione 2.4

$$PPV = \frac{TP}{TP + FP} \quad (2.4)$$

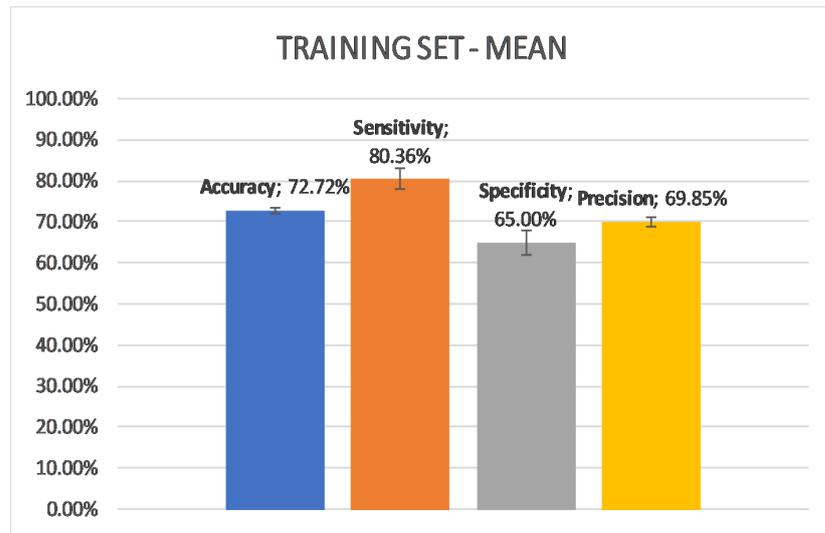
2. Sensitività, definita dall'Equazione 2.5

$$TPR = \frac{TP}{TP + FN} \quad (2.5)$$

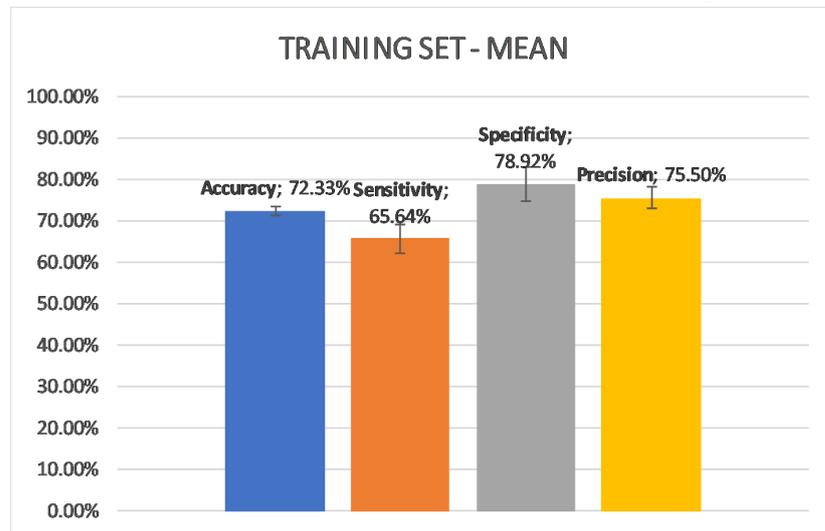
3. Specificità, definita dall'Equazione 2.6

$$TNR = \frac{TN}{TN + FP} \quad (2.6)$$

2.4.1 Risultati con 1 layer



(a) Grafico relativo alle performance del Training Set con 1 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Training Set con 1 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	76.71%	32.81%
	1 (RM)	23.29%	69.08%

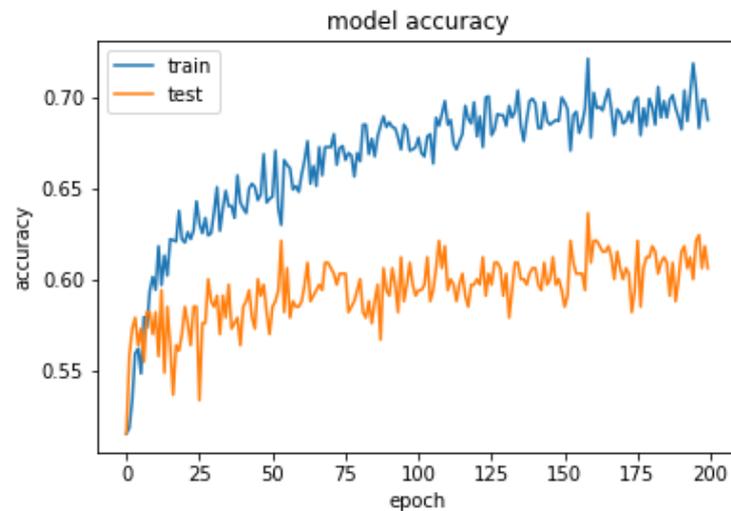
Tabella 2.5: Confusion Matrix relativa al Training Set con 1 Layer di Convoluzione considerando la normalizzazione slice per slice.

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	69.99%	24.58%
	1 (RM)	30.01%	75.42%

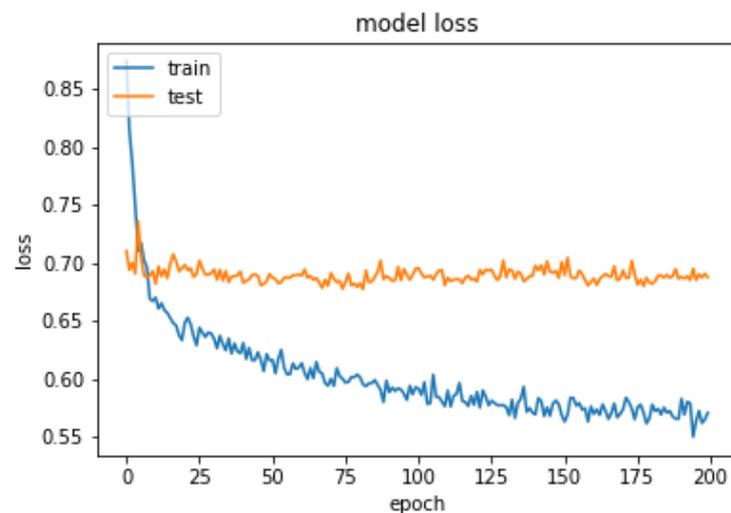
Tabella 2.6: Confusion Matrix relativa al Training Set con 1 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Mettendo a confronto i due metodi di normalizzazione sfruttati per allenare la rete nel caso con un solo layer di convoluzione composto da 30 filtri e con una matrice delle feature di grandezza 3x3, si osservano due comportamenti diversi. Nel caso della normalizzazione slice per slice il valore di sensibilità tocca il 80% (Figura 2.10a), mentre

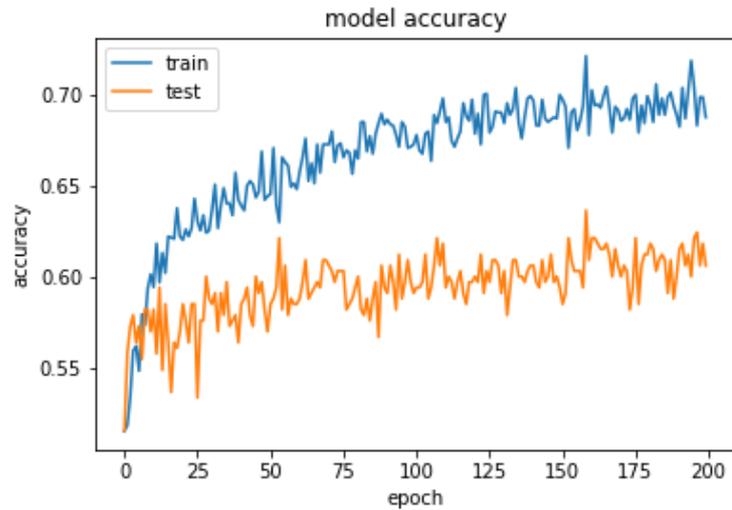
nel caso della normalizzazione considerando l'intero paziente è la specificità ad arrivare quasi a tale percentuale (Figura 2.10b). Analogamente, si ha che la percentuale di specificità nel primo caso raggiunge il 65% (Tabella 2.10a) e la sensitività ha lo stesso comportamento nel caso di normalizzazione dell'intero paziente (Tabella 2.10b). Si deduce che, nel caso di normalizzazione considerando la singola slice, la rete tende a riconoscere più le zone YM e nel dubbio mette nella classe 0, infatti si ha un tasso di falsi negativi più alto, come visibile in Tabella 2.5. Si può fare il ragionamento inverso nel caso di normalizzazione considerando l'intero paziente, dove si nota una percentuale più alta di falsi positivi a dispetto dei falsi negativi (Tabella 2.6). Sostanzialmente, in questo caso la rete tende a riconoscere di più le zone RM. Nonostante queste considerazioni, le due reti si assestano allo stesso livello di accuratezza a circa 72% (Tabelle 2.10a 2.10b).



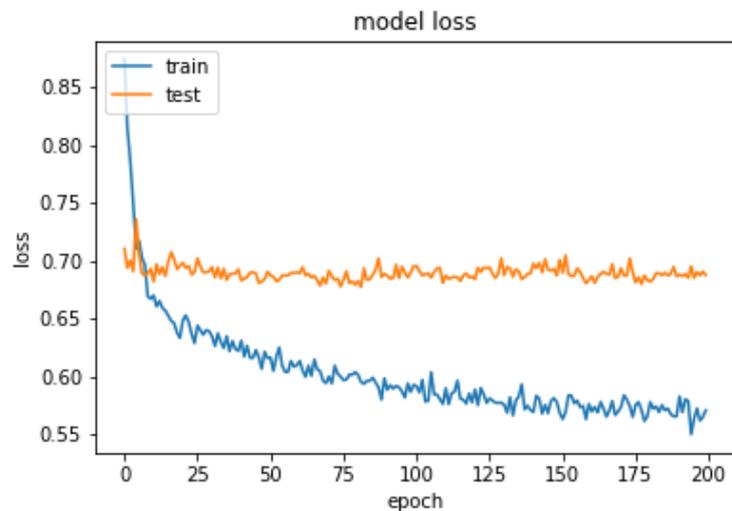
(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 1 layer di convoluzione nel caso di normalizzazione slice per slice.



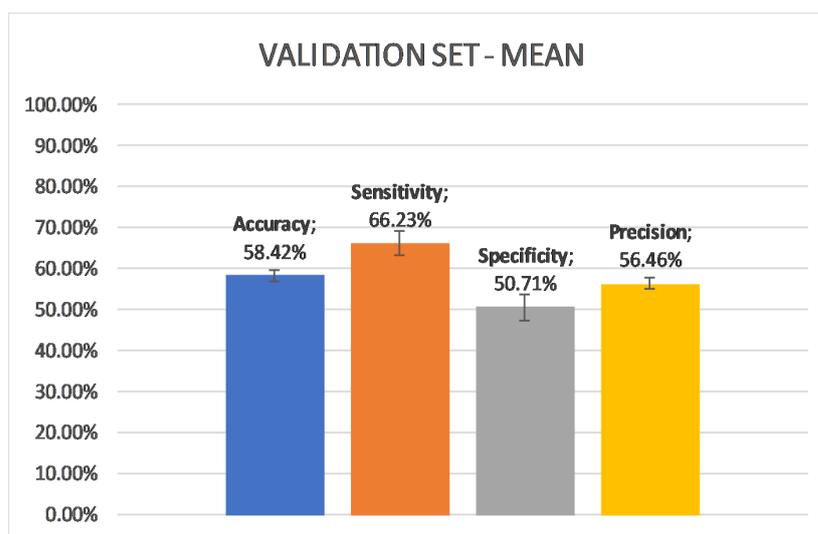
(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 1 layer di convoluzione nel caso di normalizzazione slice per slice.



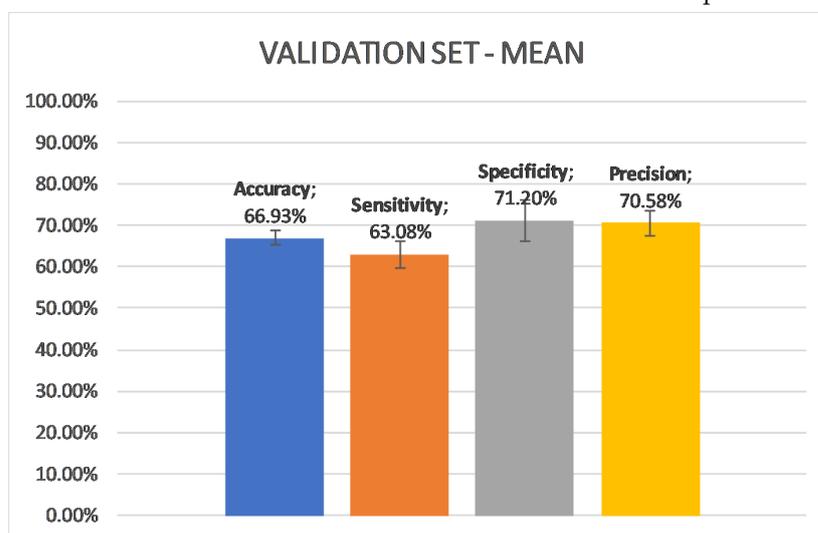
(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 1 layer di convoluzione nel caso di normalizzazione paziente per paziente.



(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 1 layer di convoluzione nel caso di normalizzazione paziente per paziente.



(a) Grafico relativo alle performance del Validation Set con 1 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Validation Set con 1 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	60.90%	43.55%
	1 (RM)	39.09%	56.44%

Tabella 2.7: Confusion Matrix relativa al Validation Set con 1 Layer di Convoluzione considerando la normalizzazione slice per slice.

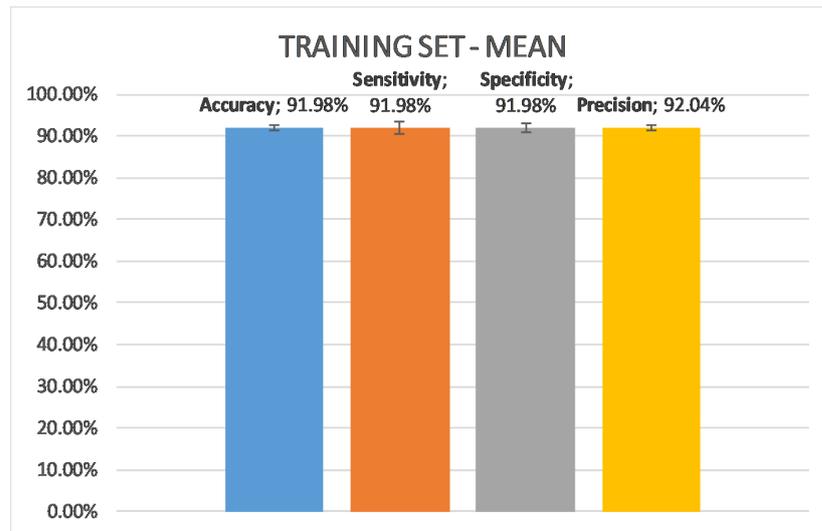
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	63.90%	29.59%
	1 (RM)	36.10%	70.41%

Tabella 2.8: Confusion Matrix relativa al Validation Set con 1 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

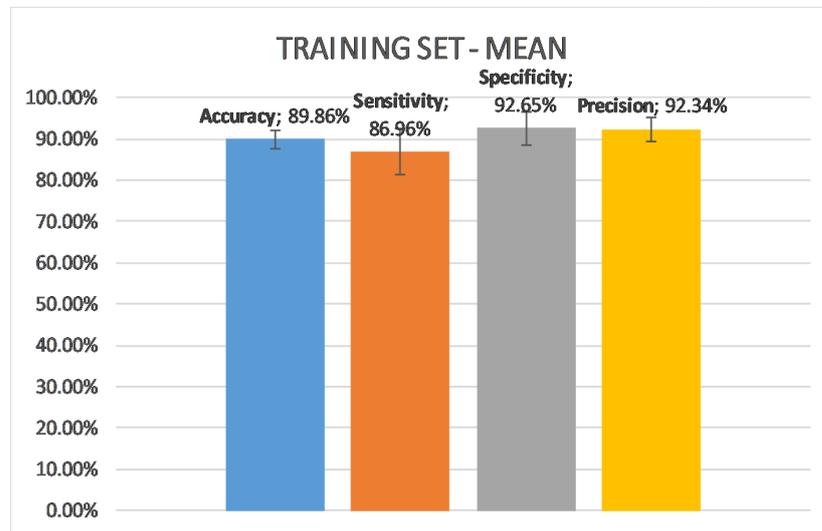
Confrontando il comportamento del Validation Set nei due approcci, si osserva un comportamento analogo al caso del Training Set, come mostrato anche dalle confusion matrix nelle Tabelle 2.7 e 2.8. In questo caso però i valori sono decisamente inferiori rispetto al Training Set (Tabelle 2.13a e 2.13b). Infatti, nel caso di normalizzazione considerando la singola slice, le performance si assestano sotto il 60% ad esclusione della sensitività che arriva a 66.23%. Mentre, nel caso di normalizzazione considerando

l'intero paziente, le performance si assestano sotto il 70%, ad eccezione della specificità che arriva a 71.20%. In definitiva, i risultati nel Validation Set, dimostrano che questo modello di CNN con un solo layer di convoluzione non riesce a identificare in maniera ottimale le zone di RM. Un ulteriore studio dei grafici di accuratezza e loss per qualsiasi approccio di normalizzazione avvalora questa idea poiché l'accuratezza dopo circa 80 epoche, su 200, si arresta (Figure 2.11a e 2.12a), mentre la loss calcolata nel test durante la fase di allenamento non decresce mai (Figure 2.11b e 2.12b).

2.4.2 Risultati con 2 layer



(a) Grafico relativo alle performance del Training Set con 2 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Training Set con 2 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

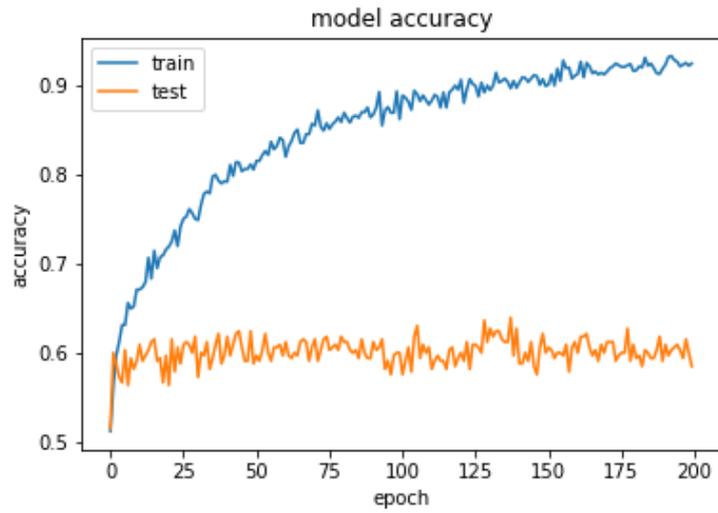
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	91.94%	7.97%
	1 (RM)	8.06%	92.03%

Tabella 2.9: Confusion Matrix relativa al Training Set con 2 Layer di Convoluzione considerando la normalizzazione slice per slice.

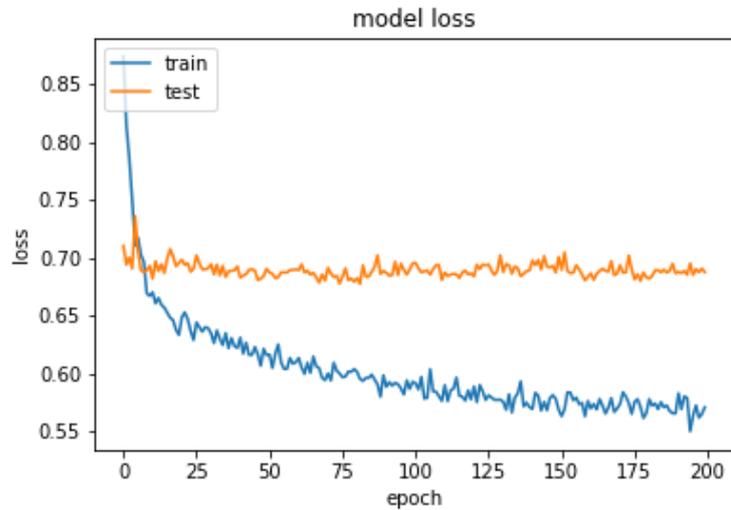
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	86.94%	6.70%
	1 (RM)	13.06%	93.30%

Tabella 2.10: Confusion Matrix relativa al Training Set con 2 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Analizzando i due approcci di normalizzazione, questa volta la rete implementata presentava due layer di convoluzione, rispettivamente di 20 e 30 filtri con dimensione 3x3. Nel Training Set, si osserva subito un drastico aumento delle performance rispetto al caso con un solo layer di convoluzione. Infatti, tutti e quattro i parametri calcolati superano il 90% nel caso di normalizzazione slice per slice (Figura 2.14a), mentre nel caso di normalizzazione paziente per paziente, accuratezza e sensibilità sono appena sotto tale valore (Figura 2.14b). Anche le confusioni matrix nelle Tabelle 2.10 e 2.9 mostrano bassi tassi di FP e FV, mentre TN e TP superano il 90%. Per quanto questi risultati siano promettenti, si nota che in realtà la rete è andata in over-fitting, come dimostrato dai grafici di accuratezza e loss presenti nelle Figure 2.15a, 2.15b, 2.16a e 2.16b e dai bassi valori di performance ottenuti nel Validation Set.

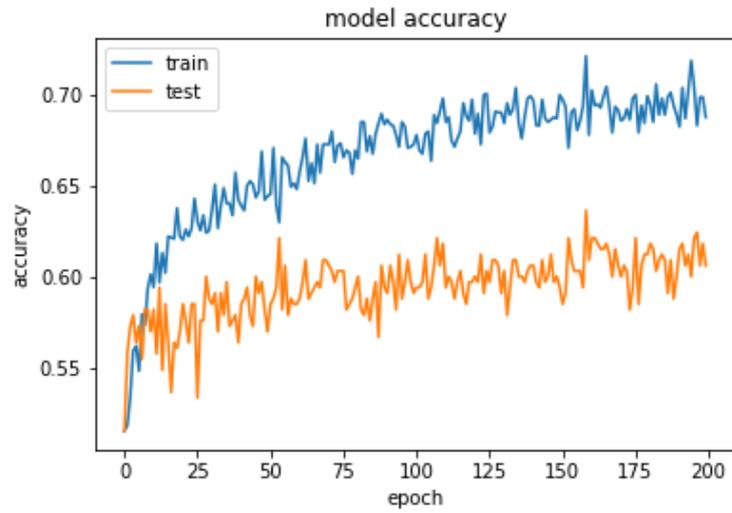


(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 2 layer di convoluzione nel caso di normalizzazione slice per slice.

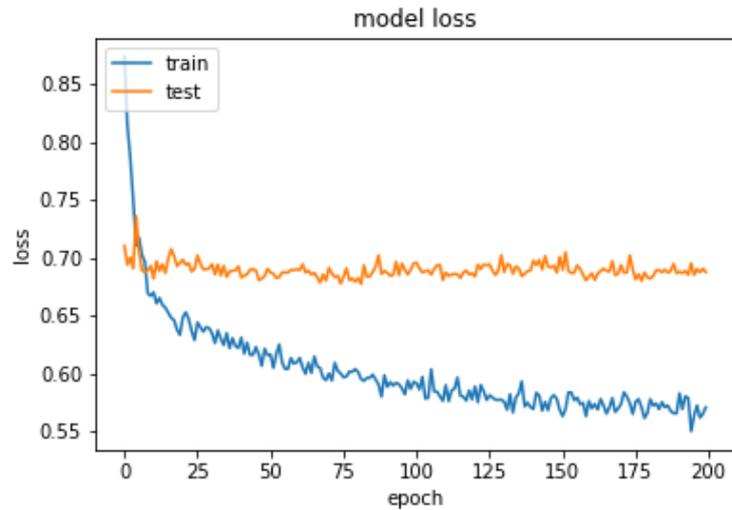


(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 2 layer di convoluzione nel caso di normalizzazione slice per slice.

Figura 2.15

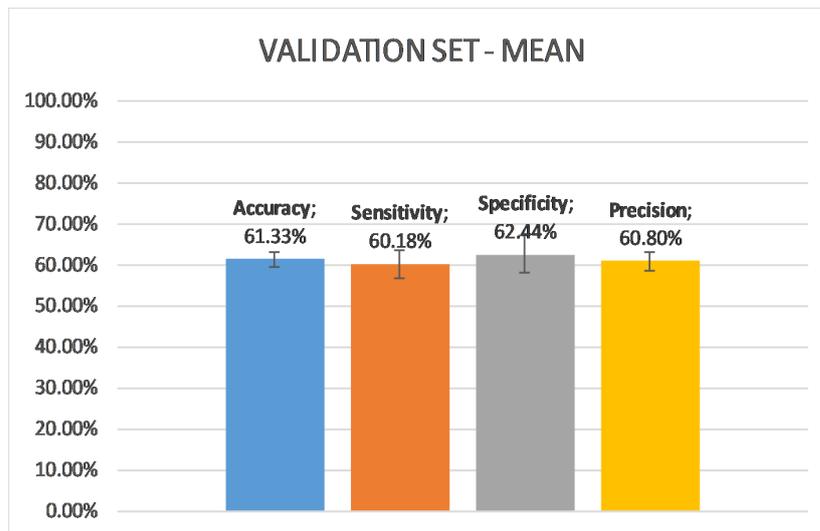


(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 2 layer di convoluzione nel caso di normalizzazione paziente per paziente.

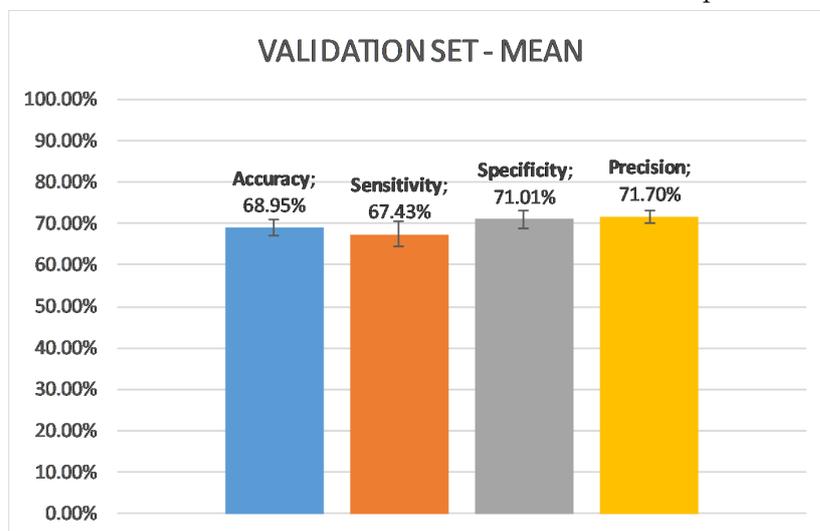


(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 2 layer di convoluzione nel caso di normalizzazione paziente per paziente.

Figura 2.16



(a) Grafico relativo alle performance del Validation Set con 2 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Validation Set con 2 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Figura 2.17

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	61.92%	39.29%
	1 (RM)	38.08%	60.71%

Tabella 2.11: Confusion Matrix relativa al Validation Set con 2 Layer di Convoluzione considerando la normalizzazione slice per slice.

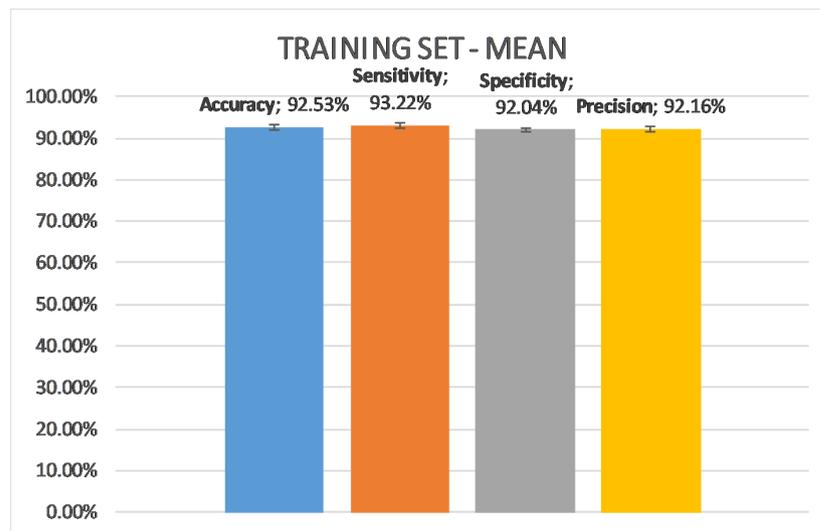
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	66.71%	28.31%
	1 (RM)	33.29%	71.69%

Tabella 2.12: Confusion Matrix relativa al Validation Set con 2 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

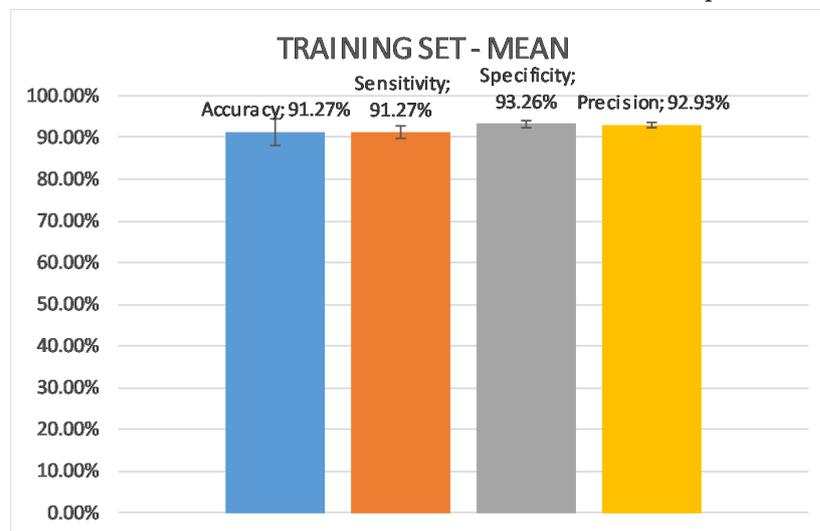
I grafici delle performance nel caso di Validation Set (Figure 2.17a 2.17b), confrontati con i valori del Training Set (Figure 2.14a 2.14b), confermano l'ipotesi di over-fitting della rete. Questo implica che la rete implementata si adatta eccessivamente ai dati del Training Set. Di conseguenza, il modello si adatta alle caratteristiche fornite dai

dati usati per allenarlo e non riesce a generalizzare, con conseguenti performance basse in fase di validazione. Questo fenomeno avviene quando il grado di libertà del modello (il numero di parametri ottimizzabili) è molto più alto comparato alle dimensioni del Training Set. Normalizzando considerando la singola slice, i quattro parametri si assestano sotto il 63% nel Validation Set, mentre con la normalizzazione considerando l'intero paziente, si hanno valori sotto il 72%.

2.4.3 Risultati con 3 layer



(a) Grafico relativo alle performance del Training Set con 3 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Training Set con 3 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Figura 2.18

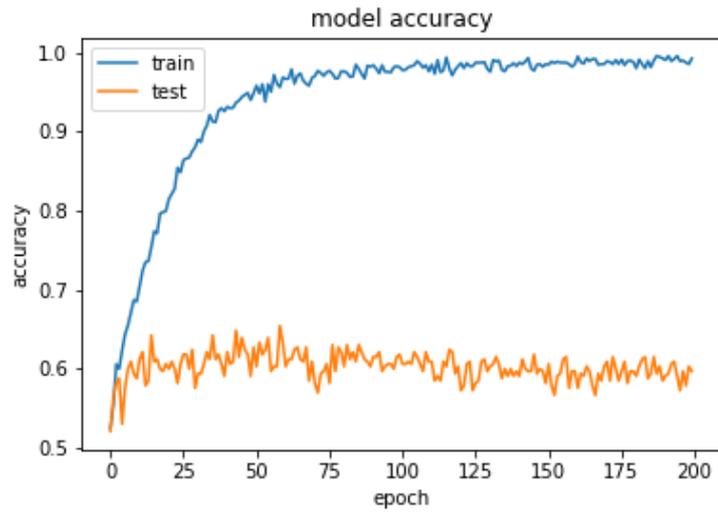
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	93.10%	7.81%
	1 (RM)	6.90%	92.19%

Tabella 2.13: Confusion Matrix relativa al Training Set con 3 Layer di Convoluzione considerando la normalizzazione slice per slice.

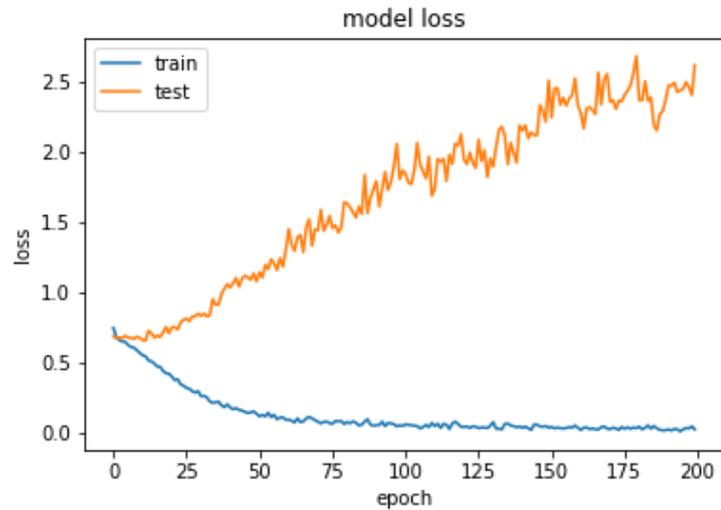
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	91.64%	7.60%
	1 (RM)	8.36%	9.24%

Tabella 2.14: Confusion Matrix relativa al Training Set con 3 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Studiando il caso di un modello di CNN con tre layer di convoluzione con, rispettivamente, 20, 30 e 40 filtri di dimensione 3x3, si osserva lo stesso comportamento del caso con due layer di convoluzione. Infatti, a fronte di promettenti performance che superano il 90% durante la fase di allenamento utilizzando entrambi gli approcci di normalizzazione (Figure 2.14a e 2.14b), si hanno delle basse performance in fase di validazione. Più specificatamente, nella Tabella 2.21a, accuratezza, sensibilità, specificità e precisione sono tutte sotto il 60%. Mentre, nella Tabella 2.21b, le performance sono tutte sotto il 70%. Di conseguenza possiamo nuovamente assumere che la rete sia andata in over-fitting, rendendo così il modello inutilizzabile.

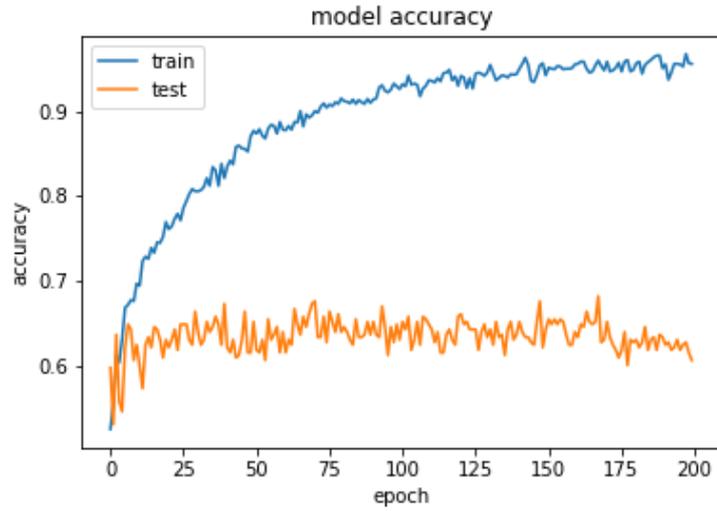


(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 3 layer di convoluzione nel caso di normalizzazione slice per slice.

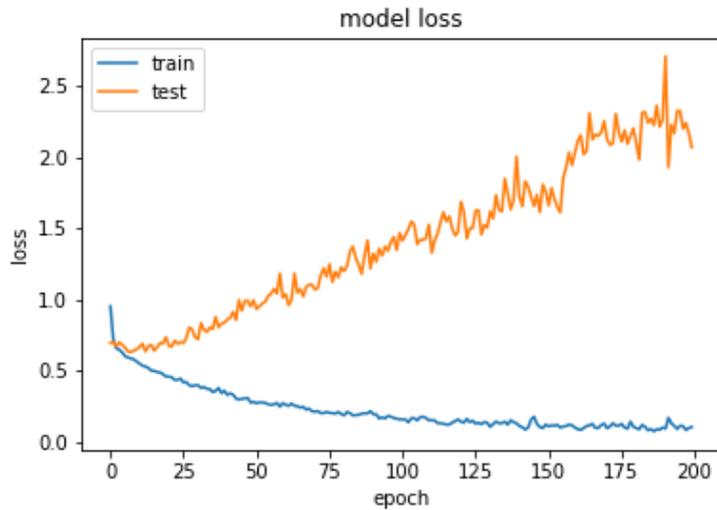


(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 3 layer di convoluzione nel caso di normalizzazione slice per slice.

Figura 2.19

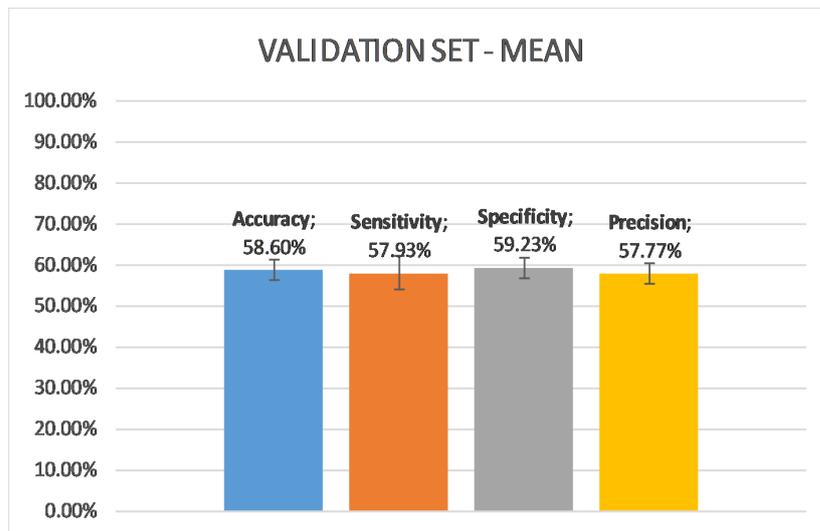


(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 3 layer di convoluzione nel caso di normalizzazione paziente per paziente.

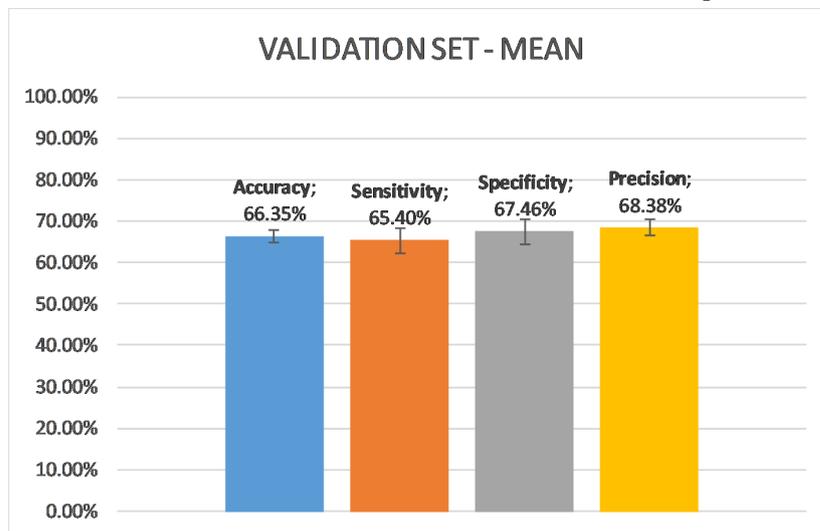


(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 3 layer di convoluzione nel caso di normalizzazione paziente per paziente.

Figura 2.20



(a) Grafico relativo alle performance del Validation Set con 3 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Validation Set con 3 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Figura 2.21

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	59.39%	42.20%
	1 (RM)	40.61%	50.00%

Tabella 2.15: Confusion Matrix relativa al Validation Set con 3 Layer di Convoluzione considerando la normalizzazione slice per slice.

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	64.18%	31.36%
	1 (RM)	35.82%	68.64%

Tabella 2.16: Confusion Matrix relativa al Validation Set con 3 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

2.4.4 Discussioni sui 3 modelli di CNN

Si osserva che l'implementazione della rete neurale convoluzionale con un solo layer di convoluzione non ha portato a performance sufficienti per un'applicazione reale. Questo perché il modello non è in grado di imparare utilizzando i dati forniti: la loss non diminuisce e l'accuratezza si arresta dopo solo 80 epoche su 200 impostate. Questo si riflette nei valori di accuratezza, sensibilità, specificità e precisione che, in ambedue gli approcci di normalizzazione, hanno valori bassi e non raggiungono nemmeno il 75% nel Validation Set. Inoltre, dalle confusion matrix si osservano dei valori molto alti di FP e FN, rispettivamente uguali a 39.09% e 43.55% nel caso di normalizzazione slice per slice (Tabella 2.7) e 36.10% e 29.59% nel caso di normalizzazione paziente per paziente (Tabella 2.8).

Per quanto riguarda i restanti due modelli di CNN con 2 e 3 layer di convoluzione, il comportamento è analogo poiché tutte e due tendono all'over-fitting; questo indipendentemente dalla normalizzazione scelta. Ciò evidenzia come, l'applicazione dell'algoritmo di min-max sulla singola slice o sull'intero paziente non porta ad un evidente beneficio. Anche in questo caso, i due modelli non riescono a fornire un adeguato riconoscimento delle zone di midollo osseo rosso e necessitano di ulteriore studio e conseguente ottimizzazione. Inoltre, l'aggiunta di un terzo layer di convoluzione rispetto a due, comporta un lieve peggioramento delle performance, infatti le performance scendono di circa il 10%. Ciò potrebbe suggerire che utilizzare un numero aggiuntivo di layer non fa che aumentare la complessità della nostra rete, un aspetto dal quale spesso non si riesce a trarre beneficio, soprattutto quando si dispone di pochi elementi con cui allenare la rete, come in questo caso.

2.5 Implementazione CNN con Training Set diviso al 50%

Visti i risultati non soddisfacenti relativi ai precedenti modelli di CNN (Capitolo 2.4), si è pensato di implementare la rete neurale convoluzionale tentando di compensare il problema dell'over-fitting che si è presentato nei modelli che prevedevano l'implementazione di due e tre layer di convoluzione. L'over-fitting è stato causato dal fatto che il modello si adattava troppo agli elementi presenti nel Training Set ed infatti presentava degli ottimi risultati in termini di accuratezza, specificità, sensibilità e precisione in fase di addestramento.

Quando si è provato a testare la rete con il Validation Set, le performance sono scese drasticamente. Una possibile soluzione a questo problema, non avendo a disposizione altri dati, è stata quella di dividere in parti eque Training Set e Test Set. Più nello specifico, il Test Set è il gruppo di ROI che si è venuto a creare nel momento dell'allenamento a seguito della funzione shuffle. In questo caso, la suddivisione è stata a metà, dunque 50% sia Training Set che Test Set. Il Test Set permette di valutare le performance della rete durante l'allenamento, in questo modo si può capire meglio come si comporta in termini di accuratezza e loss.

Per quanto riguarda il modello di CNN implementato, si è scelto di usare la rete con 2 layer di convoluzione. Difatti, seppur il comportamento tra questa rete e quella con 3 layer di convoluzione è simile, la prima aveva performance leggermente migliori nel Validation Set, come visibile nelle Figure 2.17a e 2.17b. Ciò evidenziava come l'aggiunta di un terzo layer di convoluzione, e quindi di un ulteriore livello di complessità, lede

l'apprendimento della rete nel caso di un Training Set di piccole dimensioni. In aggiunta, si è deciso di continuare a procedere con ambedue le normalizzazioni, quindi si sono considerate le ROI normalizzate considerando la singola slice e quelle considerando tutte le slice dell'intero paziente.

La struttura della rete neurale convoluzionale, dunque, è simile a quella studiata nel Capitolo 2.4.2. Sono stati usati due layer di convoluzione con rispettivamente 20 e 30 filtri di dimensione 3x3, layer di Dropout con rate del 50%, layer di Flatten ed infine un layer Dense con 1 come dimensione di output. La rete è stata allenata per un totale di 10 volte, per verificarne la stabilità ed in ogni prova, il modello è stato addestrato per 200 epoche. In aggiunta, in ogni prova è stato mantenuto sempre lo stesso Validation Set e quindi soltanto il Training Set è stato fatto variare durante l'allenamento tramite il comando shuffle.

2.5.1 Risultati della CNN con Training Set e Test Set equi



(a) Grafico relativo alle performance del Training Set al 50% con 2 Layer di Convoluzione considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Training Set al 50% con 2 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Figura 2.22

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	75.46%	24.41%
	1 (RM)	24.54%	75.59%

Tabella 2.17: Confusion Matrix relativa al Training Set al 50% con 2 Layer di Convoluzione considerando la normalizzazione slice per slice.

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	79.61%	24.33%
	1 (RM)	20.39%	75.67%

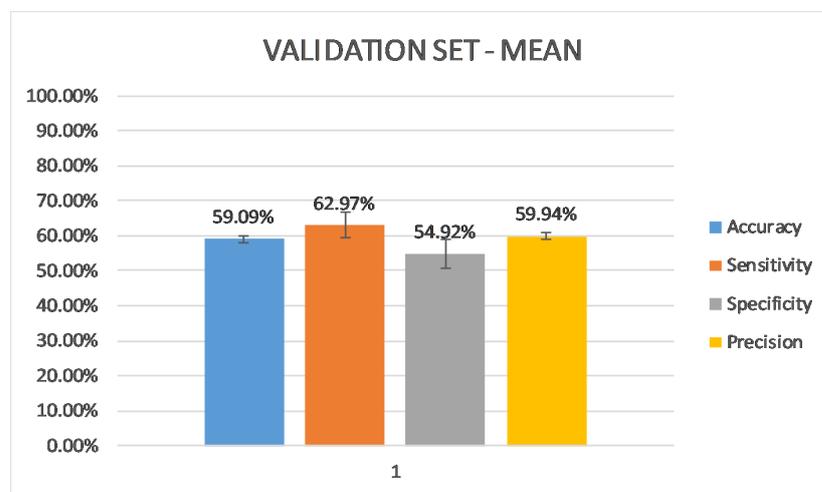
Tabella 2.18: Confusion Matrix relativa al Training Set al 50% con 2 Layer di Convoluzione considerando la normalizzazione paziente per paziente.

Mettendo a confronto i due grafici relativi al Training Set (Figure 2.22a e 2.22b), si osserva come in ambedue i casi, si ha una certa stabilità dei parametri. Nel caso della normalizzazione slice per slice, tutti e quattro i parametri calcolati si aggirano

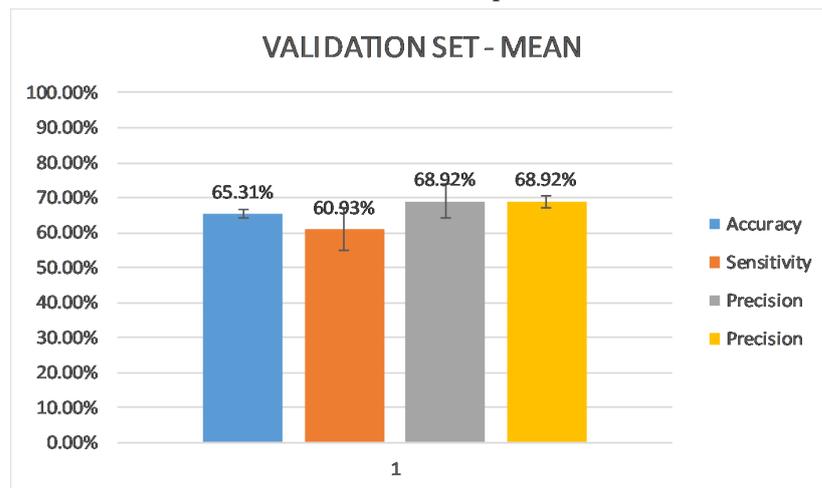
intorno al 75%, mentre nel caso di normalizzazione considerando l'intero paziente, sono lievemente più alti e la specificità arriva a circa 80%.

In generale quindi, la rete si comporta in maniera quasi identica, ciò è suggerito anche dalle confusion matrix che riassumono le 10 prove effettuate, infatti le percentuali di TP, TN, FP e FN sono simili nei due casi di normalizzazione. Un altro aspetto da sottolineare, è che non si hanno più valori nel Training Set che si discostano eccessivamente da quelli del Validation Set (come accadeva nel Capitolo 2.4.2). Ciò è indice che l'effetto di over-fitting è stato decisamente ridimensionato, la rete tende ad adattarsi di meno, poiché si allena sul 50% invece che il 70%.

Osservando invece i grafici di accuratezza e loss durante l'allenamento 2.24a 2.25a, si evince che, pur dividendo equamente Training Set e Test Set, la rete non riesce ad imparare in maniera sufficiente. Infatti, si ha che l'accuratezza nel test non accenna ad aumentare e rimane costante nelle 200 epoche, mentre si nota un aumento della loss.



(a) Grafico relativo alle performance del Validation Set considerando il Training Set al 50%, con 2 Layer di Convoluzione e considerando la normalizzazione slice per slice.



(b) Grafico relativo alle performance del Validation Set considerando il Training Set al 50%, con 2 Layer di Convoluzione e considerando la normalizzazione paziente per paziente.

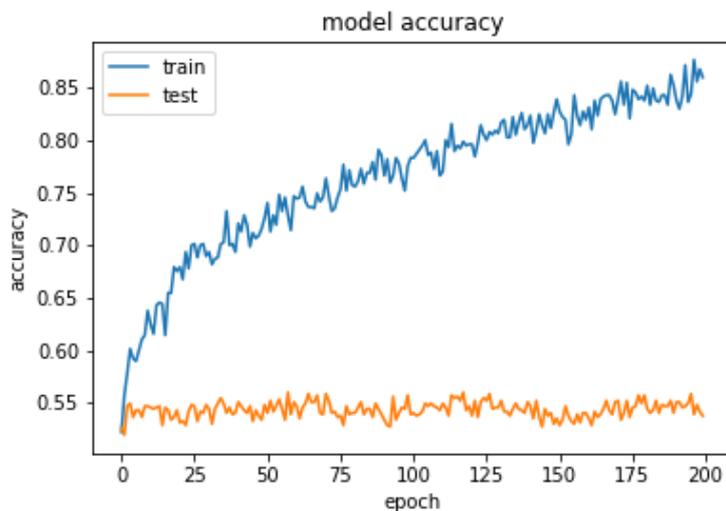
Figura 2.23

		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	54.92%	37.01%
	1 (RM)	45.08%	62.99%

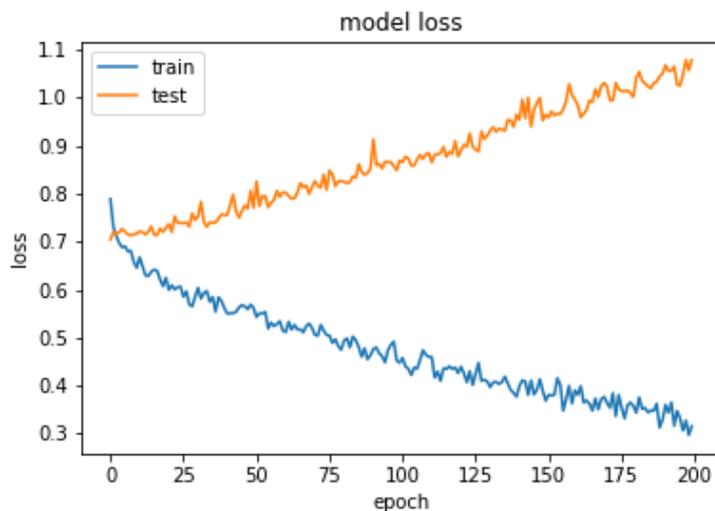
		Classe Attuale	
		0 (YM)	1 (RM)
Classe Predetta	0 (YM)	70.06%	39.01%
	1 (RM)	29.94%	60.93%

Tabella 2.19: Confusion Matrix relativa al Validation Set considerando il Training Set al 50%, con 2 Layer di Convoluzione e considerando la normalizzazione slice per slice.

Tabella 2.20: Confusion Matrix relativa al Validation Set considerando il Training Set al 50%, con 2 Layer di Convoluzione e considerando la normalizzazione paziente per paziente.

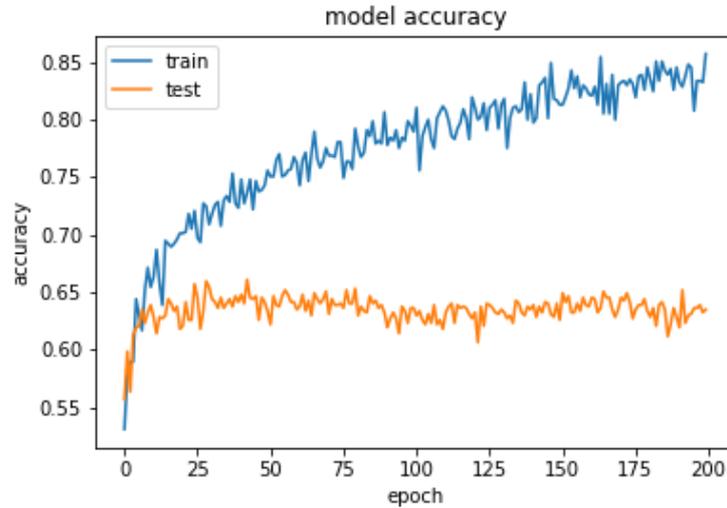


(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 2 layer di convoluzione e Training Set al 50% nel caso di normalizzazione slice per slice.

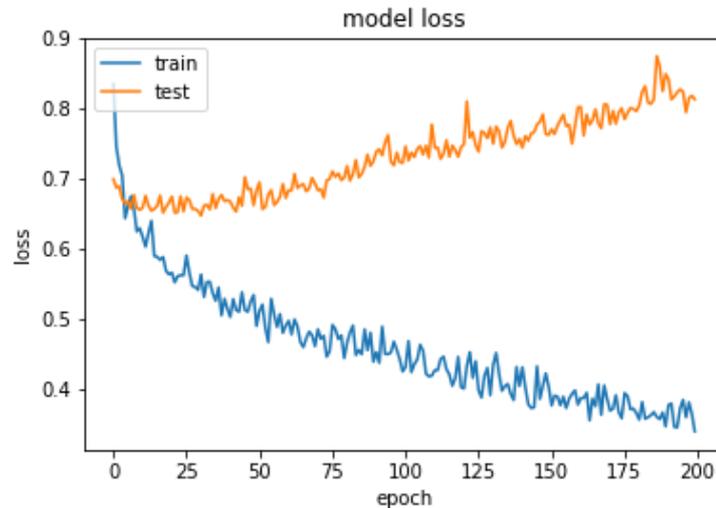


(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 2 layer di convoluzione e Training Set al 50% nel caso di normalizzazione slice per slice.

Figura 2.24



(a) Grafico relativo all'accuratezza della prima prova durante l'allenamento della CNN con 2 layer di convoluzione e Training Set al 50% nel caso di normalizzazione paziente per paziente.



(b) Grafico relativo alla loss della prima prova durante l'allenamento della CNN con 2 layer di convoluzione e Training Set al 50% nel caso di normalizzazione paziente per paziente.

Figura 2.25

Lo studio delle prestazioni delle due reti nel caso di Validation Set, evidenzia come il modello continua a non essere performante in ambedue i casi. Accuratezza, sensibilità, specificità e precisione si aggirano intorno al 60% quando vengono mostrati nuovi elementi alla rete normalizzando slice per slice. Mentre a seguito di una normalizzazione paziente per paziente, le performance sono intorno al 65%

Questi risultati suggeriscono come, probabilmente, la configurazione di rete neurale convoluzionale attuale, non riesce ad estrapolare informazioni abbastanza caratterizzanti da delle immagini così piccole. Oltre alle ridotte dimensioni delle ROI 5x5, bisogna considerare anche le poche immagini disponibili. Infatti, la CNN è una rete che, solitamente, per avere performance valide, deve essere allenata con molti più elementi rispetto a quelli forniti dal Data Set.

Considerando che i due layer di convoluzione sembrano essere il punto di equilibrio tra le tre implementazioni provate, si potrebbe studiare come tale struttura si comporta con più dati, per vedere se è possibile ottenere un miglioramento. Questo significa estrarre ulteriori ROI aggiungendo altri pazienti. Si potrebbe analizzare se, il maggiore popolamento del Data Set va di pari passo con l'aumento di complessità della rete di convoluzione e quindi analizzare l'effettivo comportamento della rete quando si hanno più immagini e quando si implementano tre layer di convoluzione. Questo per poter confermare l'ipotesi fatta nel Capitolo 2.4.3, ovvero che aumentando la complessità della CNN, si rischia effettivamente di peggiorare le performance.

2.6 Analisi dei dati tramite Dendrogrammi

A seguito delle basse performance ottenute nei Capitoli 2.4 2.5, si è reso necessario uno studio più dettagliato delle ROI che costituiscono il Data Set. Più nello specifico, si è cercato di capire se i possibili problemi avuti implementando le reti neurali convoluzionali, siano stati causati da una non omogeneità tra le immagini. Per fare ciò è stata utilizzata la tecnica del dendrogramma.

Il dendrogramma è un diagramma a cascata in grado di evidenziare la correlazione e le relazioni tra gli elementi raggruppandoli in cluster. Difatti, partendo dall'alto verso il basso, si osserva un partizionamento a cascata fino ad arrivare all'elemento specifico; l'asse delle ascisse indica la distanza logica tra i cluster, mentre l'asse delle ordinate definisce il livello di partizionamento. Quindi l'obiettivo è quello di fornire le ROI 5x5 relative alla sotto-regione LSBM del midollo osseo e vedere se quest'ultime possono essere raggruppate in cluster omogenei.

Come indicato nel Capitolo 2.3, i dati con cui si è lavorato nei Capitoli 2.4 e 2.5, sono gli stessi implementati negli studi sulla radiomica [14]. Per ottenere uno studio più completo sul dendrogramma, si sono considerate le 5 soluzioni fornite da GA per il classificatore NN. Quest'ultimo forniva performance abbastanza buone in tutte e 3 le sotto-regioni considerate, inoltre, lo studio delle 5 soluzioni GA permette di osservare l'omogeneità delle ROI con diverse combinazioni di feature estratte negli studi di radiomica [14] e definite nel Capitolo 2.3.

Nella Tabella 2.21, sono indicate le 5 soluzioni dell'algoritmo genetico associate al classificatore NN:

#1 order feats	GLCM	GLDM	GLRLM
Sol 1 1 1 1 1	0 1 0 1 1 0 1 0 1 1 1 1 0 0 0 0 0 1	0 0 0 0 1	0 1 1 0 1
Sol 2 1 1 1 0	1 0 1 0 1 0 0 1 0 0 0 1 0 1 1 0 1 0 0 1 0 1	1 0 1 0 0	0 1 1 0 0
Sol 3 1 0 0 0	1 0 1 0 0 0 0 0 0 1 1 1 0 0 1 0 0 0 1 0 1 1	0 1 0 1 0	0 0 0 1 1
Sol 4 1 1 1 0	0 1 1 1 1 1 0 0 1 0 1 1 1 1 0 1 1 1 1 1 0	0 0 1 1 1	1 1 0 1 1
Sol 5 1 1 0 0	1 0 0 1 1 1 1 1 1 0 1 0 0 0 1 1 0 0 1 0 1 0	0 0 0 0 0	0 0 1 0 1

Tabella 2.21: 5 Soluzioni GA ottenute per l'ottimizzazione del reti neurali.

Le 5 soluzioni GA sono state utilizzate per implementare 5 diversi dendrogrammi, in modo da analizzare, più nel dettaglio, la correlazione tra le 2200 ROI che compongono il Data Set relativo alla sotto-regione LSBM. È importante osservare che, per massimizzare le informazioni ottenute, l'analisi dei dati tramite dendrogrammi è stata condotta studiando i 5 dendrogrammi costruiti considerando tutte e 2200 ROI, indipendentemente dalla classe di appartenenza.

Prima di utilizzare le 36 feature estratte dalle ROI, esse sono state normalizzate sfruttando l'algoritmo di min-max scaling (Equazione 2.7). Questo passo ha permesso di

portare l'intervallo di valori delle feature tra 0 e 1. Inoltre, i valori di massimo e minimo sono stati estratti considerando tutti i valori della feature presa in considerazione nelle ROI appartenenti all'intero paziente.

$$f' = \frac{f - \min(V)}{\max(V) - \min(V)} \quad (2.7)$$

dove f' indica la feature normalizzata ottenuta a seguito dell'applicazione dell'algoritmo di min-max scaling, f definisce una delle 36 feature prese in considerazione, mentre V definisce il vettore-colonna con tutti i valori della feature nelle ROI appartenenti all'intero paziente. Di tale vettore vengono presi il valore massimo e minimo.

2.6.1 Dendrogrammi con ROI RM e YM

In questo caso, visto l'alto numero di elementi utilizzati nel dendrogramma, si è preferito lavorare direttamente con tagli unici per vedere se in questo caso si ottenevano informazioni rilevanti circa l'omogeneità del Data Set. Più specificatamente, si sono considerati due approcci:

- Tagliare i 5 dendrogrammi con un taglio unico formando 2 cluster;
- Tagliare i 5 dendrogrammi con un taglio unico formando 3 cluster, tranne per la soluzione 3 nel quale si implementa un taglio a 4.

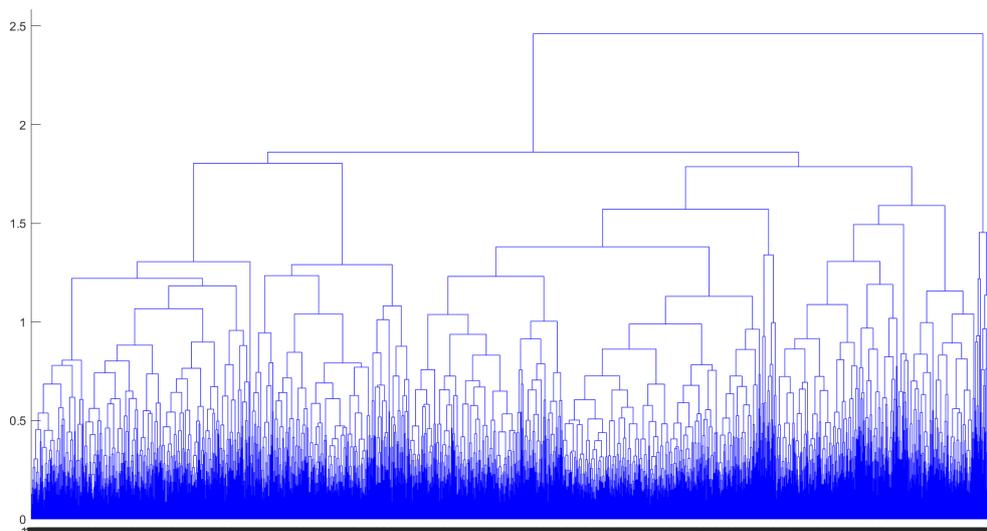


Figura 2.26: Dendrogramma delle 2200 ROI nella prima soluzione GA.

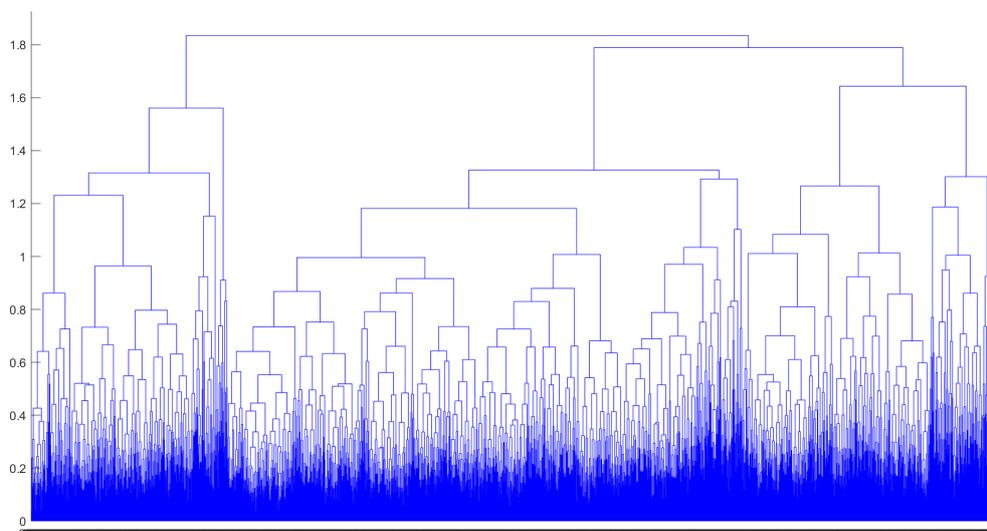


Figura 2.27: Dendrogramma delle 2200 ROI nella seconda soluzione GA.

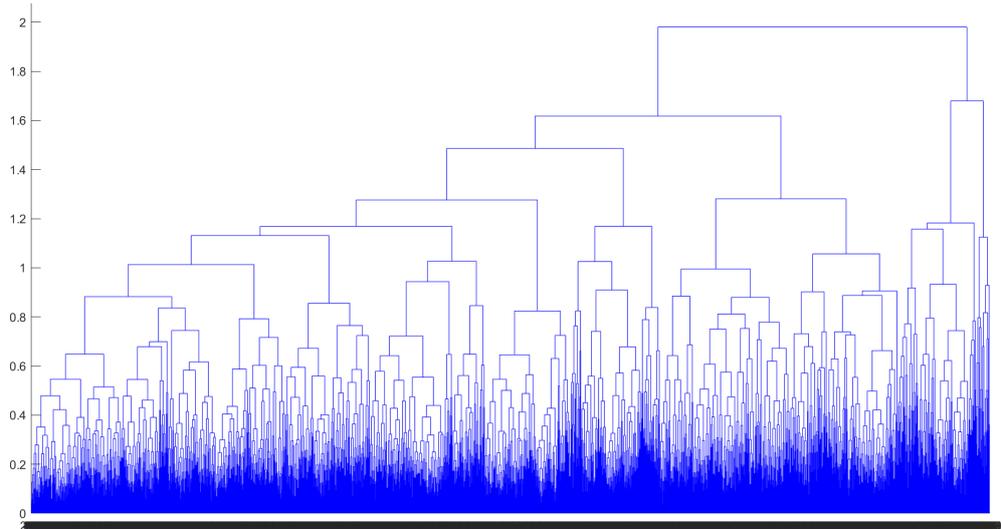


Figura 2.28: Dendrogramma delle 2200 ROI nella terza soluzione GA.

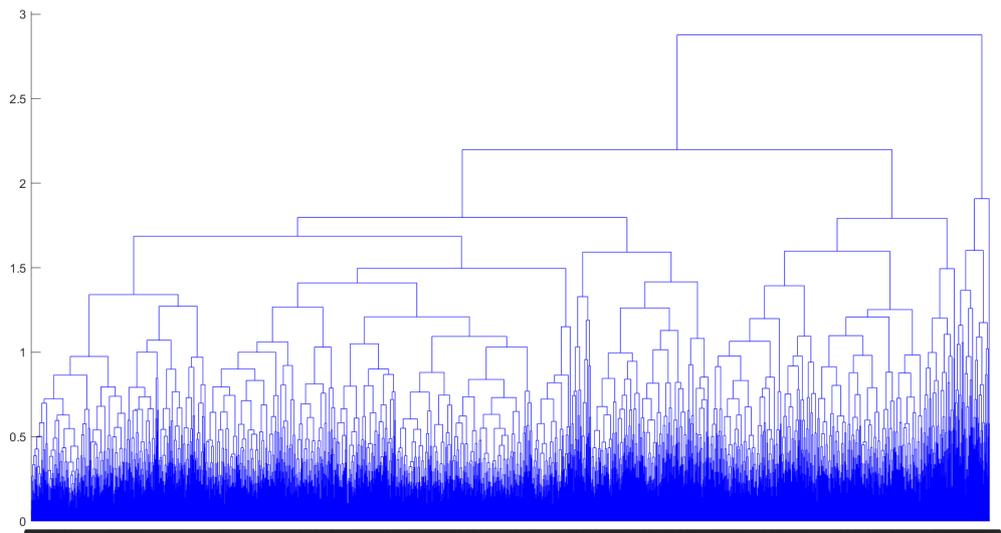


Figura 2.29: Dendrogramma delle 2200 ROI nella quarta soluzione GA.

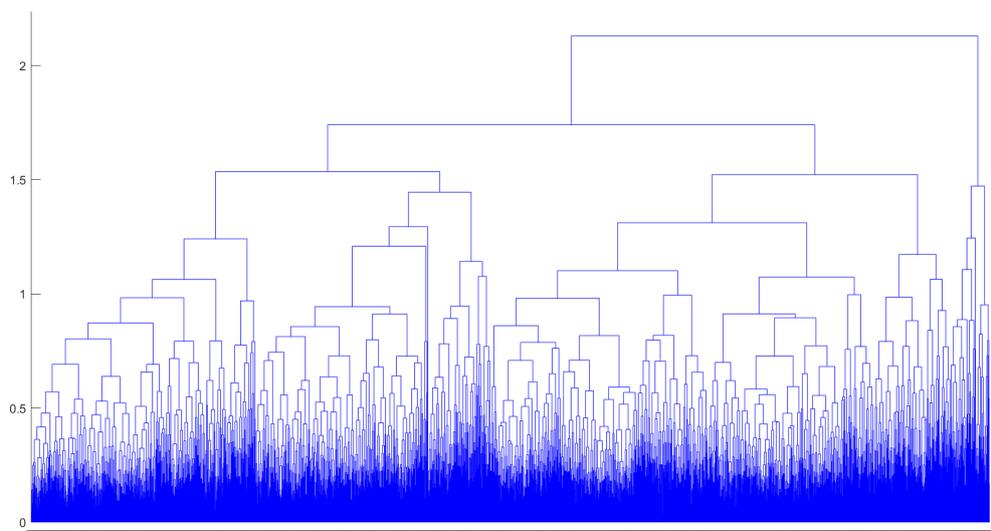


Figura 2.30: Dendrogramma delle 2200 ROI nella quinta soluzione GA.

	Cluster	Classe		ELEMENTI TOTALI	SOMMA
		0	1		
SOLUZIONE 1	Cluster 1	24	10	34	2200
	Cluster 2	1076	1090	2166	
SOLUZIONE 2	Cluster 1	239	212	451	2200
	Cluster 2	861	888	1749	
SOLUZIONE 3	Cluster 1	959	1029	1988	2200
	Cluster 2	141	71	212	
SOLUZIONE 4	Cluster 1	56	24	80	2200
	Cluster 2	1044	1076	2120	
SOLUZIONE 5	Cluster 1	59	30	89	2200
	Cluster 2	1041	1070	2111	

Tabella 2.22: Tabella riassuntiva dei cluster ottenuti applicando un taglio unico pari a 2.

Si osserva già da ora come, da questo primo taglio a due, i cluster delle varie soluzioni siano omogenee poiché presentano quasi lo stesso numero di ROI appartenenti alla classe RM che a quella YM. Si è deciso di fare un ulteriore studio con un taglio a 3, tranne per la soluzione 3 che è stata tagliata a 4 poiché a differenza degli altri presenta un cluster nell'estrema destra più grande.

	Cluster	Classe		ELEMENTI TOTALI	SOMMA
		0	1		
SOLUZIONE 1	Cluster 1	665	633	1298	2200
	Cluster 2	411	457	868	
	Cluster 3	24	10	34	
SOLUZIONE 2	Cluster 1	601	580	1181	2200
	Cluster 2	260	308	568	
	Cluster 3	239	212	451	
SOLUZIONE 3	Cluster 1	226	319	545	2200
	Cluster 2	733	710	1443	
	Cluster 3	28	4	32	
	Cluster 4	113	67	180	
SOLUZIONE 4	Cluster 1	251	313	564	2200
	Cluster 2	793	763	1556	
	Cluster 3	56	24	80	
SOLUZIONE 5	Cluster 1	494	564	1058	2200
	Cluster 2	547	506	1053	
	Cluster 3	59	30	89	

Tabella 2.23: Tabella riassuntiva dei cluster ottenuti applicando i tagli unici a 3, tranne per il dendrogramma con soluzione 3 dove si è effettuato un taglio a 4.

Pure in questo caso, le ROI delle due classi sono distribuite in maniera omogenea tra i cluster. Questa è un'ulteriore conferma che il Data Set è distribuito in maniera omogenea e quindi, le performance scarse ottenute con la CNN non sono dovute al Data Set di partenza. Ne consegue che è necessario cambiare l'architettura di rete neurale utilizzata.

3. U-Net

3.1 Introduzione a U-Net

Come già scritto nel Capitolo 2, le reti neurali convoluzionali, sono state in grado di raggiungere obiettivi che fino ad ora non erano stati raggiunti con altre metodologie. L'architettura U-Net è un modello di rete neurale convoluzionale che ha trovato grande applicazione e sviluppo nel campo biomedico [29] [30] [31], soprattutto nel risolvere problemi di classificazione. È doveroso precisare che questo tipo di rete si basa sulla segmentazione semantica, quindi necessita che ogni pixel dell'immagine che si vuole classificare sia discriminato. La sua struttura base è rappresentata nella Figura 3.1 ed è divisa in due sezioni:

1. La sezione di compressione, ovvero la sezione a sinistra della rete e molto più simile alla struttura base di una generica rete neurale convoluzionale. Questa consiste in una catena di Convolutional Layers ripetuta più volte, seguita da un Activation Layer (solitamente di tipo ReLu) ed infine un Max-Pooling Layer per effettuare il sotto-campionamento (la struttura di questi tre layer è descritta nel Capitolo 2.1). Generalmente, ad ogni ciclo, questa sezione tende ad estrapolare sempre più feature;
2. La sezione di espansione, con cui si intende la parte a destra della struttura. Tale percorso permette la creazione di una segmentazione ad alta risoluzione e consiste in una sequenza ripetuta di Up-Convolution ed una concatenazione con le corrispondenti features ad alta risoluzione del percorso di compressione, osservando che il crop si rende necessario a causa della perdita dei bordi. La "Up-Convolution" usa il kernel per mappare ogni vettore feature in una finestra di data dimensione sempre con una specifica funzione di attivazione (solitamente ReLu). Il layer finale è definito da un Convolutional Layer di dimensione 1x1 e permette di mappare il vettore di feature ottenuto nel layer precedente nella classe desiderata.

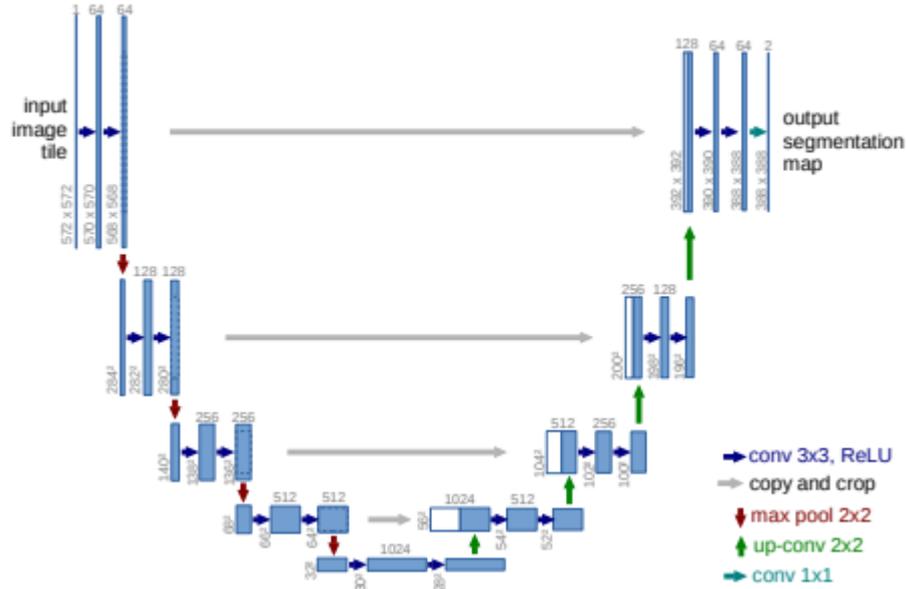


Figura 3.1: Rappresentazione di una rete U-Net base.

3.2 Applicazione U-Net

Per lavorare con la rete neurale convoluzionale U-Net, si sono raccolte tutte le slice, di dimensione 512×512 , ottenute tramite esame di tomografia computerizzata da ogni paziente; il numero di slice totale è di 356 e la suddivisione per ogni paziente è riassunta nella Tabella 3.1.

Paziente	Num. Slice
1	73
2	67
3	73
4	71
5	72

Tabella 3.1: Numero di slice ottenuto per ogni paziente a seguito dell'esame TC.

Nei precedenti studi [14], per ogni slice di tutti i pazienti sono state elaborate:

- La maschera binaria del midollo osseo presente nella regione pelvica (MaskBM) senza considerare l'osso compatto;
- La maschera binaria del midollo osseo rosso (RM) presente nella regione pelvica (MaskRM) senza considerare l'osso compatto;
- La maschera binaria del midollo osseo giallo (YM) presente nella regione pelvica (MaskYM) come sottrazione delle precedenti due.

In questo caso, a differenza del Capitolo 2, la maschera YM è stata utilizzata nell'addestramento perché oltre alla predizione della maschera RM, si è provato a lavorare anche alla predizione della maschera di midollo osseo giallo nel Capitolo 3.3.

Inoltre, si è effettuata una normalizzazione della slice al range $[0,1]$ utilizzando l'algoritmo di min-max scaling (Equazione 2.1), utilizzando il primo approccio descritto,

ovvero di considerare la singola immagine senza far influenzare i valori normalizzati dalle altre slice del paziente. In particolare, ad ogni slice è stata applicata la Mask BM corrispondente, il cui risultato può essere visto nella Figura 3.2; in questo modo, nell'applicare la normalizzazione si è tenuto conto soltanto della parte di midollo osseo, ignorando lo sfondo.



Figura 3.2: Esempio che mostra il risultato dell'applicazione della Mask BM sulla slice

La rete U-Net, come scritto in precedenza, necessita la discriminazione di tutti i pixel. Dunque, in questo caso volendo implementare un allenamento di tipo supervisionato, il Training Set composto da una porzione delle immagini ed ottenuto dal Data Set deve essere affiancato alle maschere RM o YM che identificano le classi di ogni pixel e formando quindi la "matrice" delle etichette.

La costruzione del Training Set è avvenuta anche questa volta in modo casuale, infatti è stato prelevato circa il 70% dal Data Set, quindi 251 immagini, mentre il restante 30%, ovvero 105 immagini, viene utilizzato per il Validation Set.

A questo punto si è provveduto a generare un batch del Training Set. Questo passo può essere effettuato in due maniere:

- Allocando tutto il batch nella memoria RAM;
- Tramite una funzione Generator, che permette di effettuare l'allenamento della rete allocando le immagini in memoria solo quando necessario.

Quest'ultima opzione può rivelarsi molto utile nel caso in cui l'allenamento avvenga su una macchina con una memoria sufficientemente capiente.

La rete implementata consiste in:

- Sezione di compressione, dove si ha un blocco di doppia convoluzione, Max-Pooling Layer (di dimensione 2x2) e un Dropout (con rate pari a 0.1). Più nel dettaglio, il blocco di doppia convoluzione consiste in due sotto-blocchi uguali, processati in sequenza e composti da un Convolution Layer (con kernel 3x3 e numero di filtri iniziale pari a 32), Batch Normalization e Activation Function (con ReLu come funzione di attivazione). Questa catena viene ripetuta ed ogni volta i filtri nel blocco di doppia convoluzione vengono incrementati fino a raggiungere il valore di 32768 nell'ultimo stadio;
- Sezione di espansione, dove si ha in sequenza un Deconvolution Layer (con kernel 3x3), Concatenate Layer e un blocco di doppia convoluzione. Più nel dettaglio, il layer di concatenazione si occupa di unire l'output del Deconvolution Layer

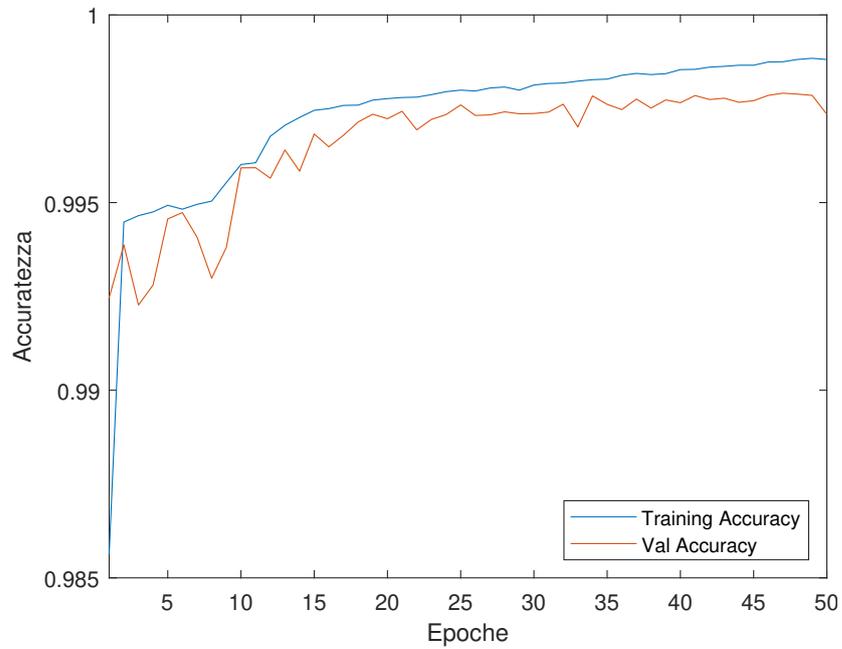
con l'output del blocco di doppia convoluzione nella fase di compressione corrispondente, il suo output viene inviato ad un altro blocco di doppia convoluzione che, nella fase di espansione, presenta un numero di filtri che decrementa fino a tornare a 32;

- Sezione output, che consiste in un singolo layer di convoluzione con un kernel e numero di filtri pari ad 1.

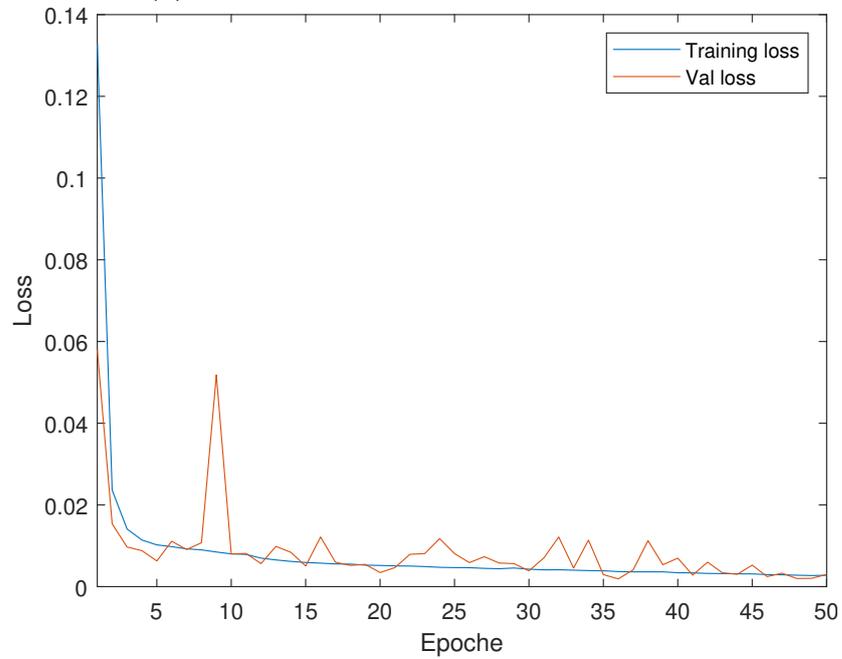
Per poter implementare una rete così complessa, è stato necessario affidarsi ad un servizio esterno di Cloud Computing. Ciò è stato possibile senza modifiche addizionali al codice, poiché la rete è stata sviluppata utilizzando un linguaggio ed una API portabili, come descritto nel Capitolo 2.

La rete è stata allenata per 50 epoche. Tale numero, in relazione all'elevatissimo numero di parametri elaborati dal modello (circa otto milioni), risulta essere un fattore limitante per la rete. Tuttavia, in fase di progettazione, è stato necessario giungere ad un compromesso a causa dell'elevato costo computazionale.

Alla fine dell'addestramento, il modello è stato salvato e utilizzato nella predizione del Validation Set. L'output della predizione è una matrice 512x512 dove in ogni cella è presente la percentuale di appartenenza ad una classe e tramite un threshold (con soglia 0.5) si è ottenuta la conversazione e l'ottenimento della maschera binaria. In Figura 3.3 sono rappresentati i grafici di accuratezza e loss del modello preso in considerazione. Si osserva che l'accuratezza e la loss nel caso della validazione, segue l'andamento dell'apprendimento.



(a) Grafico dell'andamento dell'accuratezza



(b)

Figura 3.3: Grafico dell'andamento della loss

Per la valutazione sono stati calcolati coefficiente di Dice (DSC), Precision (Pr) e Recall (Re). Gli ultimi due parametri sono già stati descritti nel Capitolo 2.4, mentre il primo è stato calcolato come nell'Equazione 3.1 e ci permette di avere un'idea della somiglianza tra la predizione e l'effettiva maschera.

$$DSC = 2 \cdot \frac{Pred \cap Target}{Pred + Target} \quad (3.1)$$

Dove Pred è la maschera binaria del midollo osseo rosso/giallo ottenuta come output dalla predizione della rete, mentre Target indica la maschera binaria del midollo osseo rosso/giallo effettiva. Si osservi che, in questo caso, la classe 1 indica la parte di midollo

osseo interessata (RM) mentre la classe 0 definisce la parte di midollo osseo giallo, ma anche lo sfondo (background, indicato con BG).

Nella Figura 3.4, sono riassunti i valori di DSC, Pr e Re per tutti e 5 i pazienti tramite boxplot. In generale si osservano dei valori molto alti in gran parte dei casi. È doveroso premettere che da tali box plot sono state rimosse quelle slice che si presentavano come outlier perché avevano valori di DSC, Pr e Re pari a 0; queste slice erano tutte caratterizzate da immagini in cui si aveva una netta predominanza di sfondo nero e l'area di RM consisteva in al più 3 pixel.

Specificatamente per il paziente 1, nel Training Set non si ottengono valori al di sotto del 90% di DSC, ad eccezione degli outlier. In aggiunta, la precisione e recall hanno valori molto alti; in particolare, quest'ultimo presenta un primo quartile che si assesta al 95%, indice che la rete riesce a predire bene l'area di RM presente nel primo paziente. Anche nel Validation Set, si hanno risultati molto promettenti, infatti la rete mostra degli ottimi valori di DSC, Pr e Re. Infatti, il 25esimo quartile di questi tre parametri è sempre superiore al 80% e comunque il minimo si assesta al 65%. Per quanto riguarda la recall si assesta al valore di 90% nel primo quartile. Tuttavia, questo primo paziente, sia nel Training Set che nel Validation Set non è esente da outlier che evidenziano alcune slice con criticità.

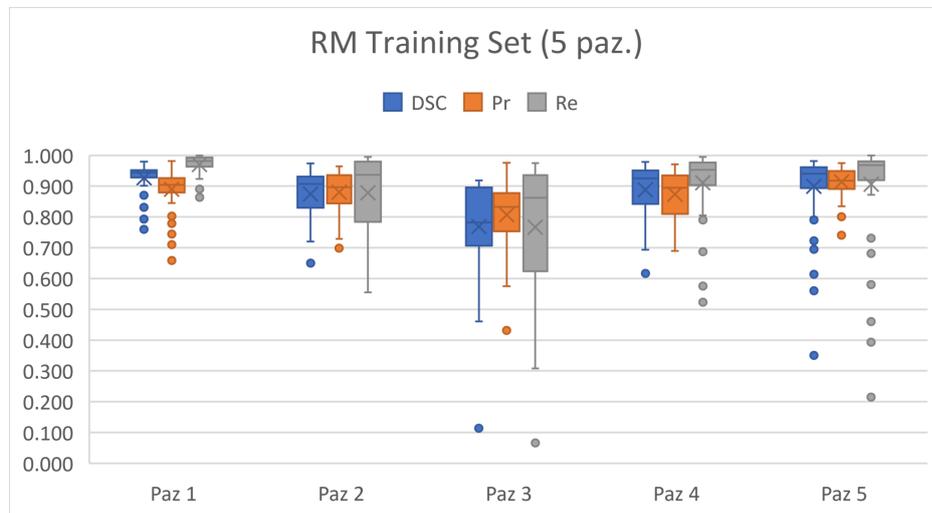
Il paziente 2 sia nel Training Set che nel Validation Set presenta meno outlier rispetto al paziente 1. Nel caso del Training Set, i valori di DSC, Pr e Re si assestano maggiori al 90% nel terzo quartile. Più specificatamente, la DSC ha un valore minimo superiore al 75%, l'unica eccezione è data da un outlier con valore intorno al 65%. La Re, invece, presenta un box plot molto asimmetrico e con una performance inferiore rispetto al paziente 1. Il Validation Set del secondo paziente presenta un DSC che si assesta intorno al 75% per il primo percentile, mentre il terzo percentile arriva a superare il 90%. Inoltre, non si hanno outlier ed il valore minimo è al 55%. In aggiunta, Pr e Re sono comparabili poiché si assestano tutti e due quasi agli stessi valori di terzo percentile e massimo, ovvero il 95%.

Il paziente 3 è quello in cui si evidenziano maggiormente, sia del Training Set che nel Validation Set, l'asimmetria e la presenza di outlier con valori bassi. Nel Training Set, il massimo valore di DSC è intorno al 90% e anche se è il più basso tra tutti i pazienti, risulta comunque molto promettente. In generale, si osserva che questo paziente presenta valore del primo quartile e del terzo quartile, per i tre parametri, basso rispetto a tutti gli altri pazienti.

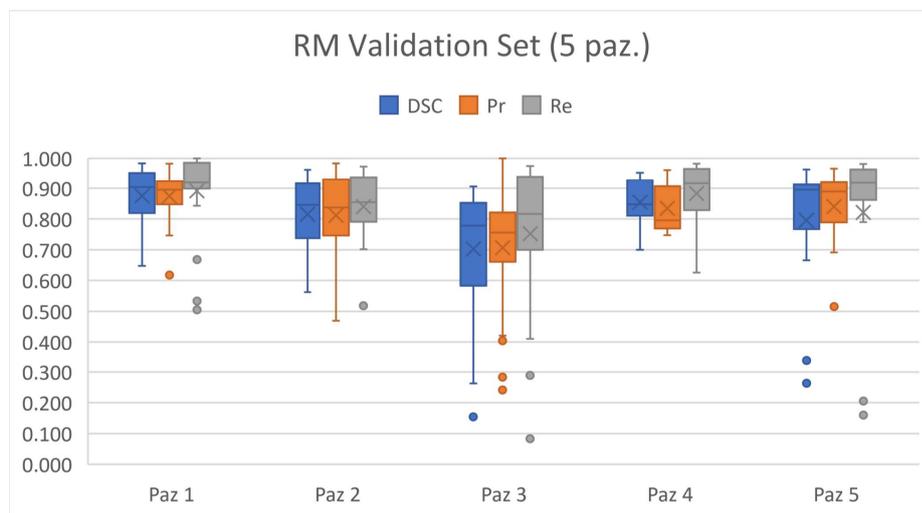
Il paziente 4 nel Training Set, ha degli ottimi risultati. Più nello specifico, il valore minimo di DSC, ad esclusione dell'outlier, pari al 70%, mentre il primo quartile si assesta al 85% ed il terzo quartile è superiore al 90%. Anche Pr e Re sono alte, indicando che la rete riesce bene ad identificare le zone di RM e differenziarle da YM e sfondo nero. Infatti, si ha un valore di recall nel 25esimo percentile pari al 90%. Anche nel Validation Set si hanno risultati promettenti. Infatti, il terzo percentile di tutti e tre i parametri calcolati è maggiore e/o uguale al 90%.

Infine, anche con il paziente 5 si sono ottenuti dei buoni risultati. Più nello specifico, il Training Set evidenzia dei valori di DSC, Pr e Re superiori al 90% a partire già dal primo quartile. Tuttavia, sono presenti molte criticità che evidenziano slice in cui la rete non è riuscita a predire bene la zona di RM e differenziarla dal resto. Anche nel Validation Set si sono ottenuti degli buoni risultati, in linea con i pazienti 1 e 4. Infatti, si ha una buona simmetria dei parametri DSC e Pr ed il primo quartile è superiore

al 75%. Più specificatamente, la Re si assesta al 85%; indicando che la rete ha una promettente capacità nel predire le zone di RM in questo paziente.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.4

3.2.1 Risultati Paziente 1

Il Training Set del paziente 1 presenta, tra i valori di DSC, 5 outlier; Tuttavia, il valore più basso è intorno al 75%. La slice corrispondente è la n147.5, rappresentata in Figura 3.5. Come si osserva, è una slice particolare perché gran parte dell'immagine è sfondo nero e le aree di RM sono molto strette e frastagliate. Tuttavia, la rete riesce a identificare queste zone ad eccezione di pochi pixel che però influenzano in maniera maggiore il risultato quando si ha una predominanza così netta di background. Inoltre, si ha un leggerissimo sovra-campionamento in quanto è presente un pixel di RM dove non è presente midollo osseo ematopoietico.

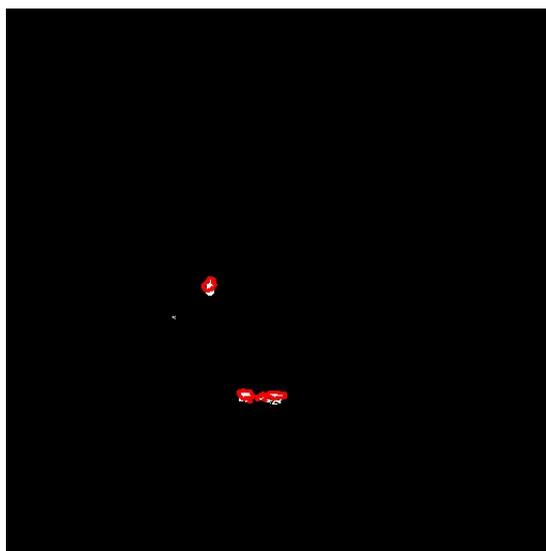


Figura 3.5: Slice n147.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

Andando in ordine decrescente, la seconda slice critica è la n36.5, con un DSC pari a 77.5% e visibile nella Figura 3.6. In questo caso la rete ha lavorato con una slice che presenta più aree ristrette di RM rispetto alla slice n147.5, ma resta ancora netta la predominanza dello sfondo nero.

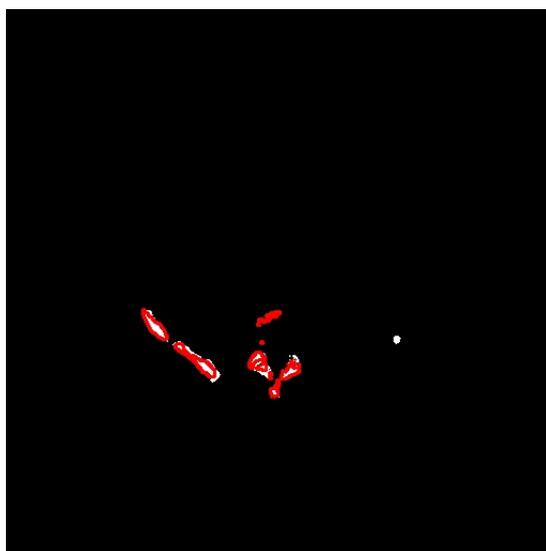


Figura 3.6: Slice n36.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

Invece, il valore di DSC più alto è 97.9% nella slice n222.5, visibile nella Figura 3.7 e rappresentante la sezione del bacino, in questo caso la rete riesce a predire quasi perfettamente la slice sbagliando soltanto una piccola porzione di bordo. In questo caso la zona di RM, seppur ridotta rispetto allo sfondo, è compatta e non frastagliata.

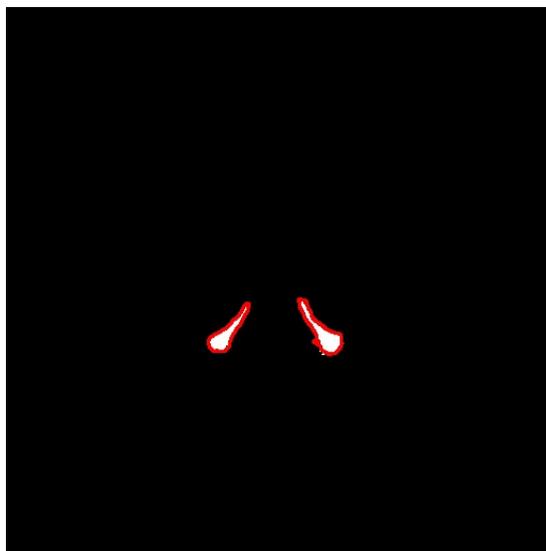


Figura 3.7: Slice n222.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

La seconda migliore predizione è pari a 97.2%, corrisponde alla slice n183.5, visibile in Figura 3.8. Questa sezione anatomica è relativa alla sinfisi pubica ed anche in questo caso si hanno aree ridotte di RM, ma compatte.

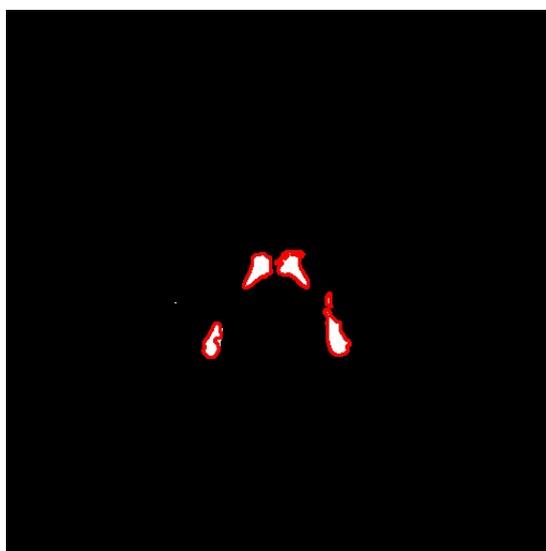


Figura 3.8: Slice n183.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

Nel Validation Set del paziente 1, non sono presenti outlier nel DSC, ma sono presenti in Pr e Re. In quest'ultimo il valore più basso è 50.5% ottenuto dalla slice n144.5, visibile in Figura 3.9. Tale slice è simile alla n147.5 (Figura 3.5) presente nel Training Set. Infatti, è netta la predominanza di sfondo e le aree di RM sono strette e frastagliate.



Figura 3.9: Slice n144.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

Il secondo valore di Re critico in ordine decrescente è 53.5%, ottenuto dalla slice n21.5, rappresentata in Figura 3.10. In questo caso, si osserva che la rete ha difficoltà a predire la parte centrale della colonna vertebrale, mostrando un evidente problema di sotto-campionamento. Tale fenomeno è decisamente più predominante rispetto al sovra-campionamento, poiché in quel caso si trattava di pochi pixel, invece, come si vede in questa slice, la presenza di FN è netta rispetto ai FP visti nelle slice precedenti.

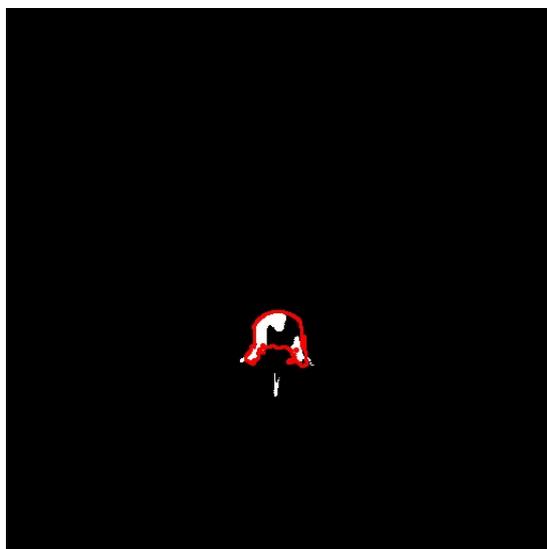


Figura 3.10: Slice n21.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

Nel Validation Set, il valore più alto di DSC è pari a 96.3%, corrispondente alla slice n228.5 (visibile in Figura 3.11). Essa presenta la stessa sezione anatomica della slice 222.5 in Figura 3.7 e ha un valore di Re pari al 99.9%. Ciò è indice che questa particolare sezione anatomica, anche se presenta una vasta predominanza di sfondo nero, riesce ad essere predetta dalla rete in maniera ottimale.

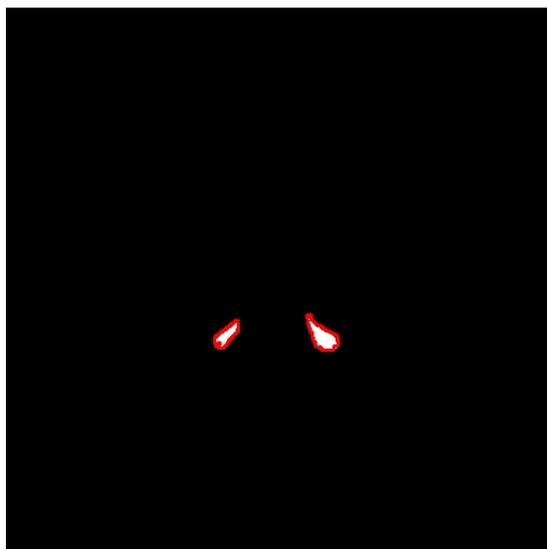


Figura 3.11: Slice n228.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

Anche la slice n195.5 ha un valore alto di DSC pari a 96.8%. Questa slice è visibile in Figura 3.12 e mostra la stessa sezione anatomica della Figura 3.8. Ciò conferma che la rete riesce a predire in maniera ottimale specifiche sezioni anatomiche.

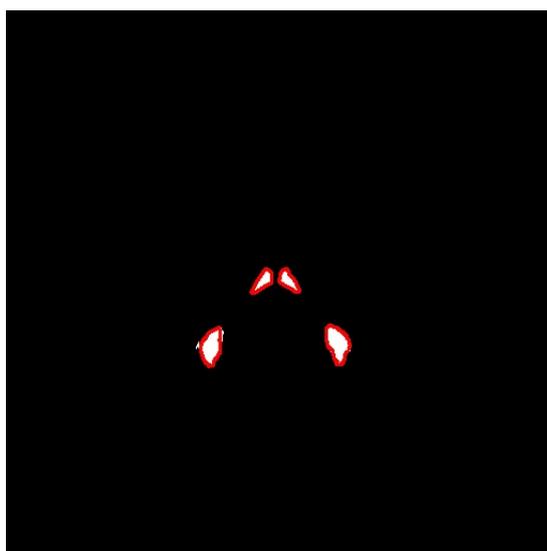


Figura 3.12: Slice n195.5 del paziente 1, il bordo rosso indica i contorni della MaskRM effettiva.

3.2.2 Risultati Paziente 2

Il Training del paziente 2 presenta pochi outlier, l'unico presente nel DSC ha valore pari al 65% ed è corrispondente alla slice n51 (visibile in Figura 3.13). Questa immagine presenta delle aree di RM molto ristrette e sono presenti, anche se in forma lieve, fenomeni di sovra-campionamento che di sotto-campionamento di midollo osseo attivo, quest'ultimo incide molto sul valore di recall che si assesta intorno al 55.5%

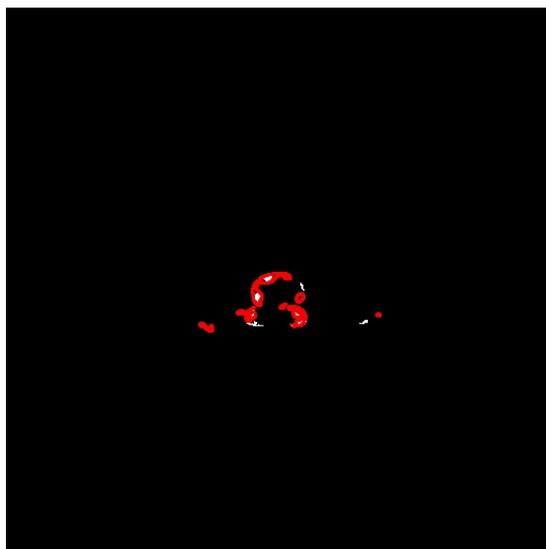


Figura 3.13: Slice n51 del paziente 2, il bordo rosso indica il contorno della MaskRM

Un'altra criticità è stata riscontrata anche nella slice n189 (in Figura 3.14); questa immagine presenta la stessa sezione anatomica delle Figure 3.8 e 3.12 del paziente 1, esse venivano predette molto bene, ma in questo caso, a causa delle aree di RM non compatte e molto piccole, la predizione non riesce ad essere ottimale sotto-campionando e sovra-campionando.

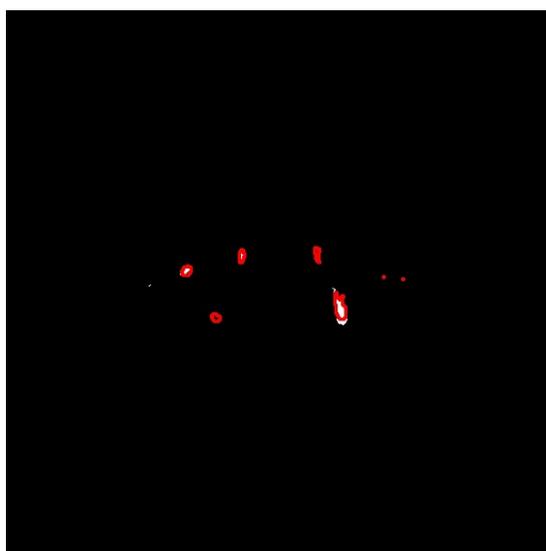


Figura 3.14: Slice n189 del paziente 2, il bordo rosso indica il contorno della MaskRM.

La porzione di Training Set relativa al paziente 2 raggiunge come valore massimo 97.4% nel DSC. Più specificatamente nella slice n99 (Figura 3.15). Si osservano i due forami otturatori presenti nel pube e ai lati le spine iliache. Le aree di RM sono compatte e molto vicine tra loro e tutte le slice vicine alla n99 sono predette molto bene.

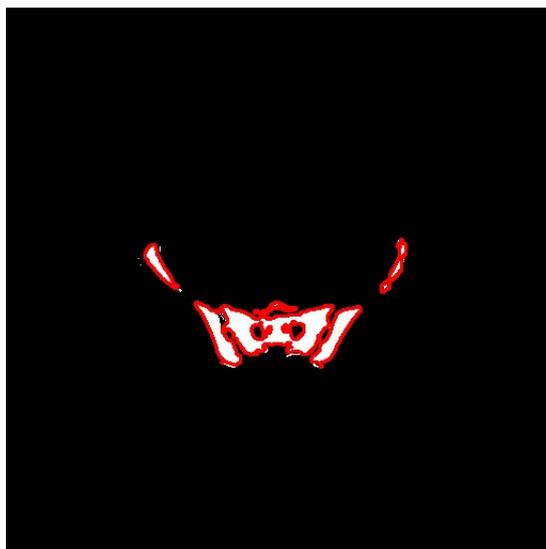


Figura 3.15: Slice n99 del paziente 2, il bordo rosso indica il contorno della MaskRM.

La slice n210 (Figura 3.16) viene predetta molto bene e con un valore di DSC pari al 92.2%. Quest'immagine mostra nuovamente la sezione della sinfisi pubica che già nel paziente 1 aveva mostrato ottime performance in termini di predizione.

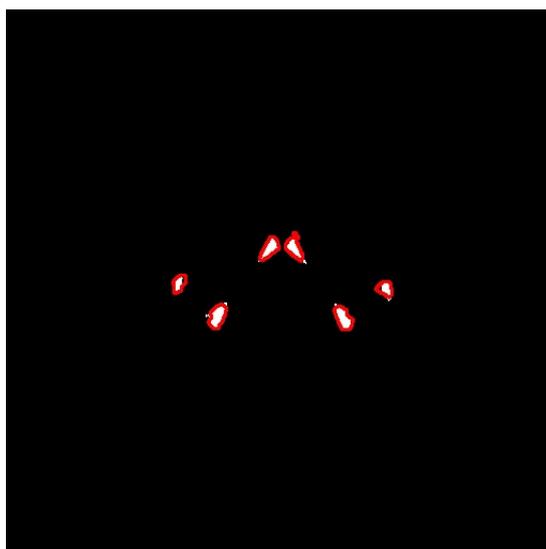


Figura 3.16: Slice n210 del paziente 2, il bordo rosso indica il contorno della MaskRM.

Per quanto riguarda il Validation Set, il valore più basso di DSC si ottiene nella slice n192 (Figura 3.17) pari al 56.3%. In questo caso si verifica lo stesso problema della slice n189 (Figura 3.14) vista in precedenza. Seppur la slice rientra nella sezione anatomica della sinfisi pubica, che è sempre stata predetta bene, la presenza di aree troppo piccole e sparse di RM porta ad una predizione bassa rispetto alla letteratura.

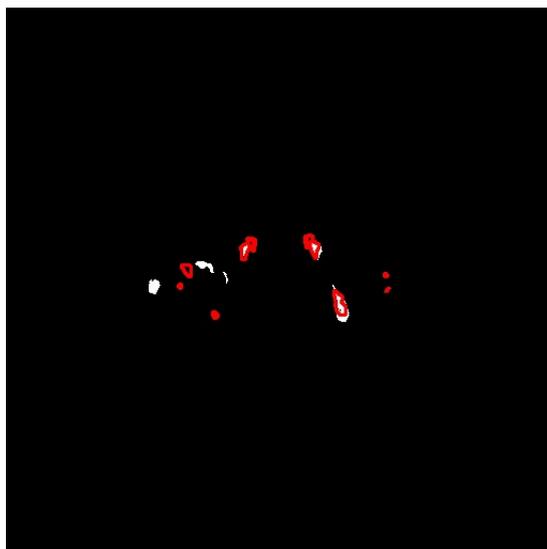


Figura 3.17: Slice n192 del paziente 2, il bordo rosso indica il contorno della MaskRM.

Inoltre, nella Re è presente un outlier di valore 51.8%. Tale percentuale è corrispondente alla slice n159, visibile in Figura 3.18. Si osserva un lieve fenomeno di sotto-campionamento del RM.

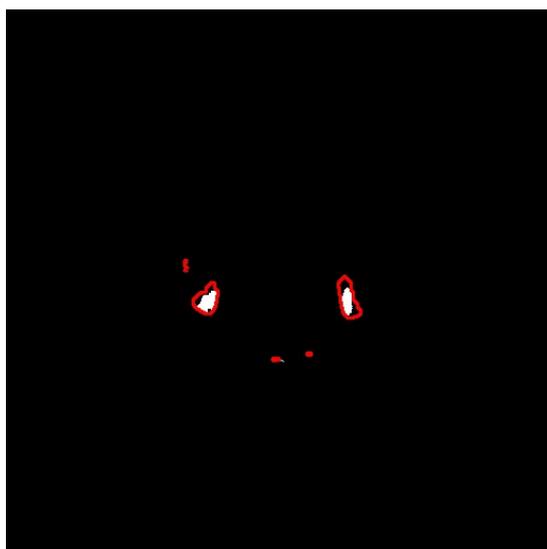


Figura 3.18: Slice n159 del paziente 2, il bordo rosso indica il contorno della MaskRM.

Nel Validation Set, il valore più alto di DSC è 96.3% e si ha nella slice n108 a cui corrisponde un valore di Re pari al 97.1%. La slice (Figura 3.19) presenta la sezione anatomica del bacino in cui si vede tutto l'osso compatto.

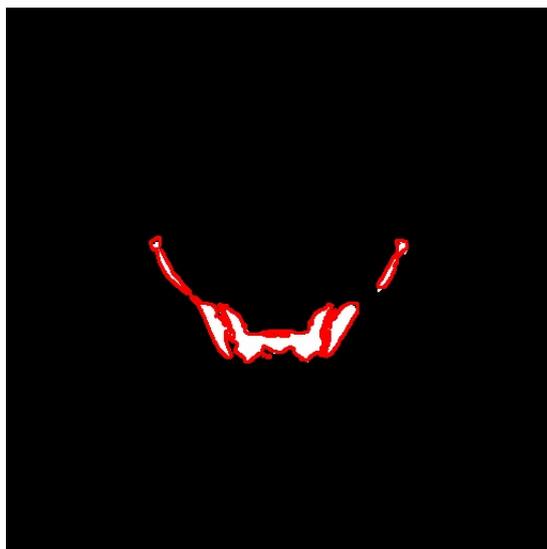


Figura 3.19: Slice n108 del paziente 2, il bordo rosso indica il contorno della MaskRM.

Anche la slice n63 (Figura 3.20) ha un valore di DSC molto alto pari a 93.9%; in questo caso si evidenziano le punte delle creste iliache ai lati, mentre al centro si ha la colonna vertebrale.

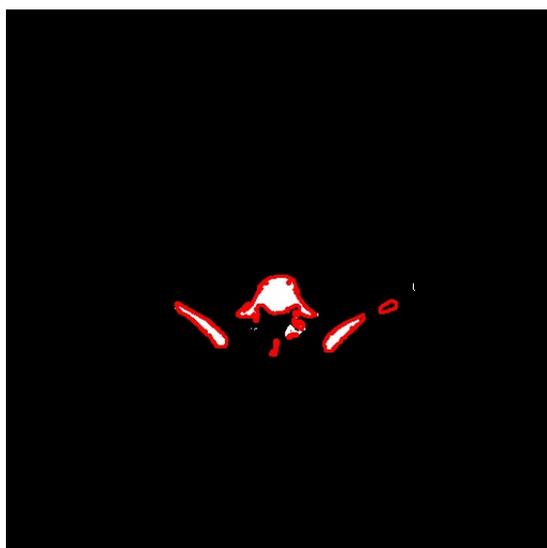


Figura 3.20: Slice n63 del paziente 2, il bordo rosso indica il contorno della MaskRM.

3.2.3 Risultati Paziente 3

La porzione di Training Set del paziente 3 è quella che mostra più asimmetria rispetto a tutti gli altri pazienti. Più nello specifico, i valori di Re sono molto vari, ciò evidenziato da un box plot molto lungo e asimmetrico e con un valore di minimo pari a 30.8%. Questo minimo è raggiunto dalla slice n202 (Figura 3.21) e nel quale si evidenzia un sotto-campionamento di midollo osseo ematopoietico. Ciò è dovuto ai bordi frastagliati delle due aree RM, unito alla forte predominanza di sfondo nero.

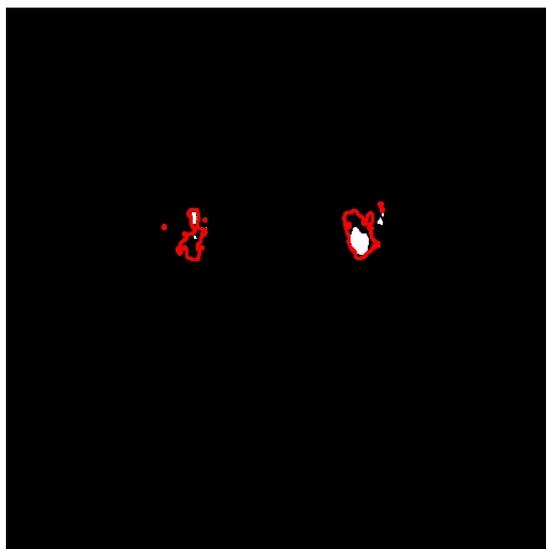


Figura 3.21: Slice n202 del paziente 3, il bordo rosso indica il contorno della MaskRM.

É visibile un outlier tra i DSC che presenta un valore basso pari a 11.4%. Corrisponde alla slice n289 (Figura 3.22) in cui si evidenzia, nuovamente, una difficoltà da parte della rete di riconoscere le zone di RM quando sono troppo piccole, sparse e si ha una forte predominanza di nero.



Figura 3.22: Slice n289 del paziente 3, il bordo rosso indica il contorno della MaskRM.

Il fenomeno appena descritto, si presenta pure nel Validation Set. Infatti, la slice n286 (Figura 3.23) evidenzia delle aree ristrette e molto sparse di RM su uno sfondo nero molto predominante. Si osservano ambedue fenomeni di sovra-campionamento e di sotto-campionamento di midollo osseo rosso.

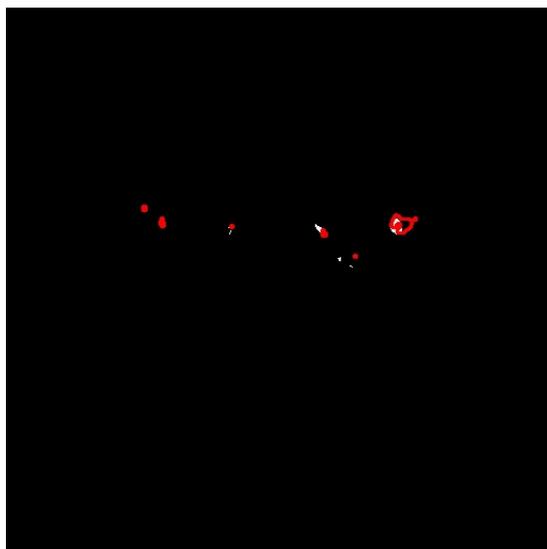


Figura 3.23: Slice n286 del paziente 3, il bordo rosso indica il contorno della MaskRM.

La slice n85 in Figura 3.24 presenta un forte fenomeno di sovra-campionamento di midollo osseo ematopoietico. Infatti, seppur le aree sono piccole e frastagliate, la rete predice come RM tutta la parte centrale.

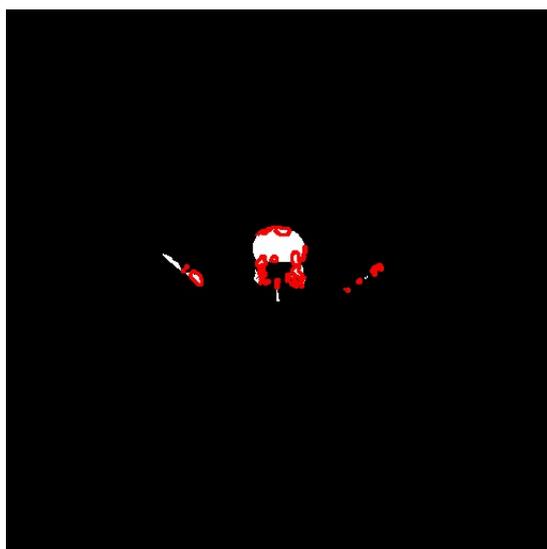


Figura 3.24: Slice n85 del paziente 3, il bordo rosso indica il contorno della MaskRM.

La porzione di Validation Set del terzo paziente mostra valori di DSC inferiori rispetto agli altri pazienti. Infatti, il massimo si ottiene nella slice n127 ed è pari a 90.6% (Figura 3.25). La rete riesce a predire bene la parte centrale del bacino anche se si osservano alcune difficoltà nei bordi.

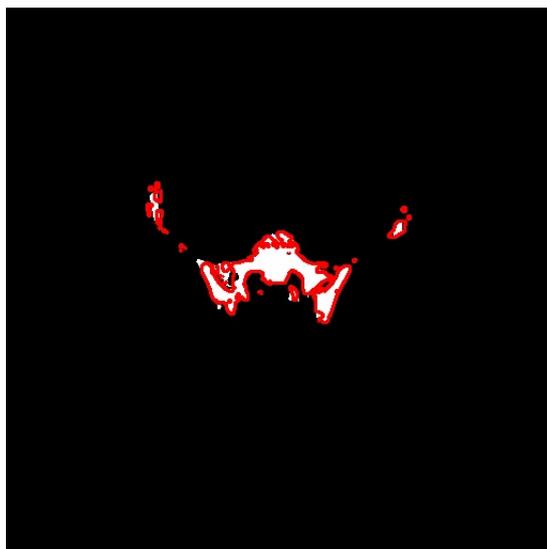


Figura 3.25: Slice n127 del paziente 3, il bordo rosso indica il contorno della MaskRM.

Nel paziente 3, solo questa sezione anatomica del LSBM viene predetta con buoni valori di DSC; come viene confermato anche dalla slice n136 in Figura 3.26 in cui si hanno comunque problemi di predizione ai bordi.



Figura 3.26: Slice n136 del paziente 3, il bordo rosso indica il contorno della MaskRM.

3.2.4 Risultati Paziente 4

La porzione di Training Set del quarto paziente presenta un outlier tra i DSC di valore pari al 61.6% associato al più basso valore di Re. Ciò corrisponde alla slice n128.5 (Figura 3.27). Questa immagine evidenzia nuovamente la criticità relativa alla difficoltà da parte della rete nel riconoscimento di midollo osseo rosso quando le aree sono ridotte e frastagliate.



Figura 3.27: Slice n128.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

Anche la slice n119.5 del quarto paziente mostra questa criticità (Figura 3.28). In generale, la rete riesce a predire bene le slice di questa porzione di Training Set, come visibile nel box plot. Le uniche criticità riscontrate sono dovute ad aree molto strette e non compatte di RM.

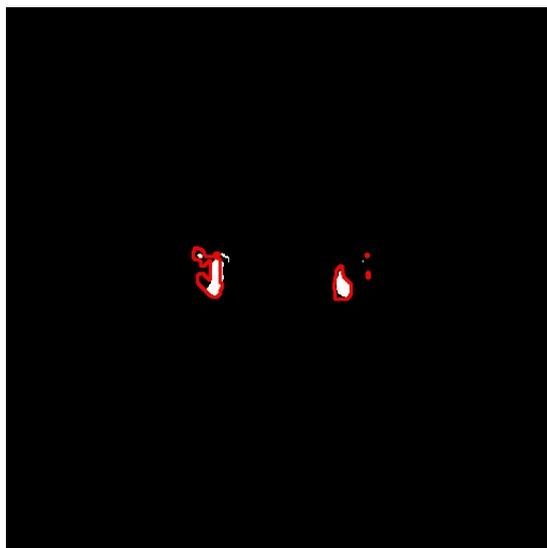


Figura 3.28: Slice n119.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

Il valore massimo di DSC è 97.9% ottenuto nella slice n8.5 (Figura 3.29). In questo caso, si è riuscito a predire molto bene la parte centrale poiché compatta anche se nella slice 21.5 (Figura 3.10) del paziente 1 aveva dato alcuni problemi di sotto-campionamento delle aree RM.

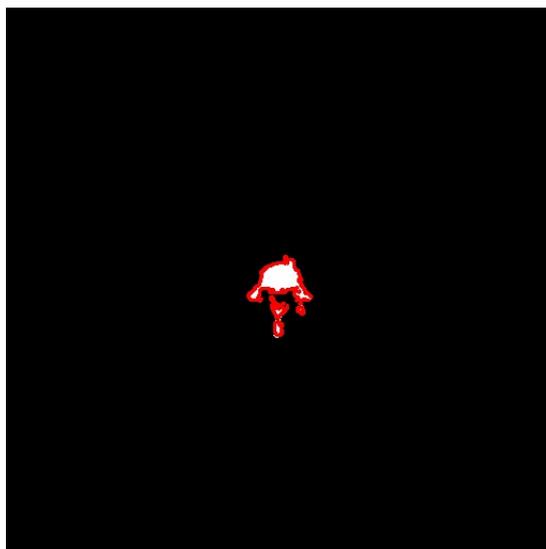


Figura 3.29: Slice n8.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

Anche la slice n83.5 del paziente 4 (Figura 3.30) ha un buon valore di DSC pari a 95.7%. Tale slice è relativa alla parte di spina che viene spinta indietro dalla presenza dell'osso sacro.

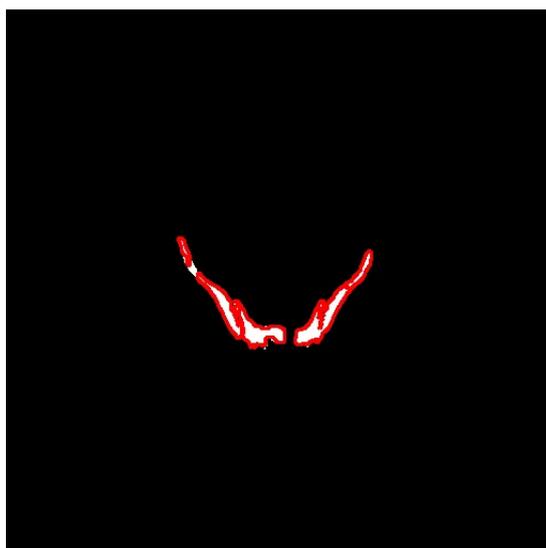


Figura 3.30: Slice n83.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

La porzione del Validation Set corrispondente al quarto paziente non presenta outliers. L'unica evidente criticità era relativa alla slice n125.5 che è stata rimossa dal box plot perché era un'immagine con al più 2-3 pixel di RM e poi tutto sfondo nero e, conseguentemente, aveva valore di DSC nullo. Per questo motivo la slice ad avere valore minimo di dice è la n116.5 (Figura 3.31) che mostra la stessa sezione anatomica della slice n119.5 (Figura 3.28) e che aveva anch'essa una predizione non performante.

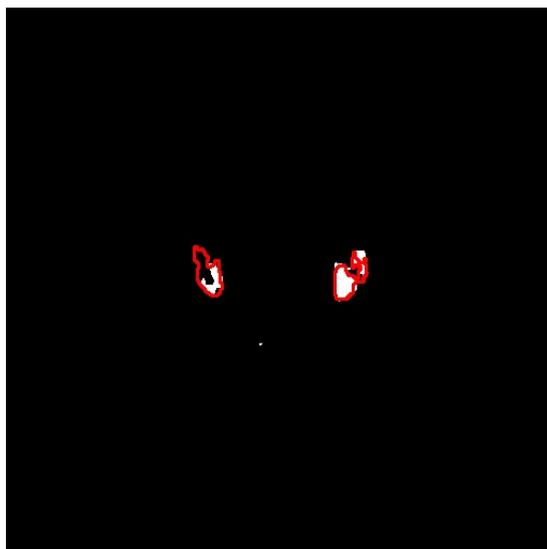


Figura 3.31: Slice n116.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

In media, il Validation Set presenta un valore di DSC di circa 85% nel quarto paziente. Tuttavia, riesce a raggiungere un valore di DSC massimo pari a 95.4% nella slice n74.5 (Figura 3.32); si osservano i due forami otturatori presenti nel pube e ai lati le spine iliache. Le aree di RM sono compatte e molto vicine tra loro e tutte le slice vicine alla n74.5 sono predette molto bene.

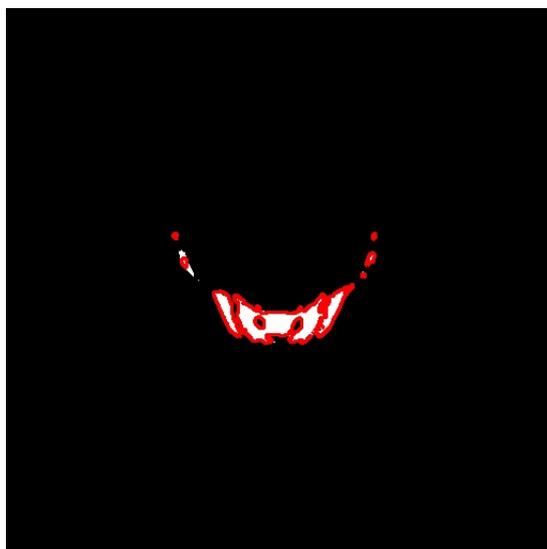


Figura 3.32: Slice n74.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

Anche la slice n101.5 riesce a raggiungere un valore di DSC pari al 93.5% (Figura 3.33). Più specificatamente, si osserva che la rete riesce a predire molto bene le punte delle creste iliache presenti ai lati.



Figura 3.33: Slice n101.5 del paziente 4, il bordo rosso indica il contorno della MaskRM

3.2.5 Risultati Paziente 5

Il paziente 5 presenta delle performance molto simili al paziente 1. Infatti, in generale la rete riesce a predire molto bene le performance di questa porzione di Training Set e Validation Set, ma sono presenti alcuni outlier che evidenziano delle criticità. Più specificatamente, nel Training Set i valori di DSC e Re più bassi (rispettivamente 35.1% e 25.1%) sono ottenuti dalla slice n6 (Figura 3.34). In questo caso, le zone di RM sono molto strette e frastagliate.



Figura 3.34: Slice 6 del paziente 5, il bordo rosso indica il contorno della MaskRM

La slice n132 con un DSC pari al 56% è la seconda criticità. Come visibile in Figura 3.35 è relativa ad una sezione anatomica in cui il RM è presente in aree non compatte, molto strette ed irregolari.

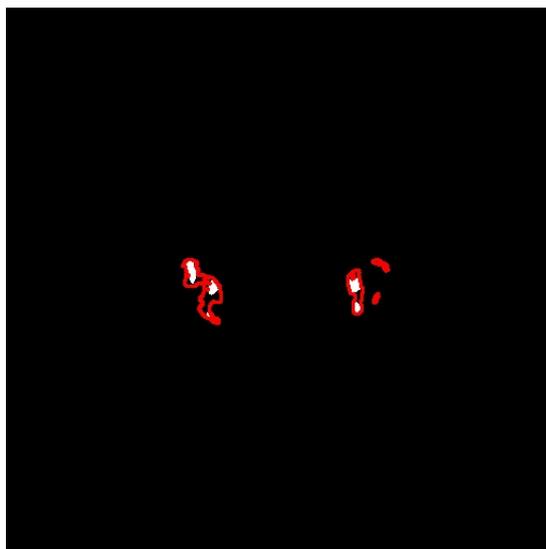


Figura 3.35: Slice n132 del paziente 5, il bordo rosso indica il contorno della MaskRM

A parte queste due criticità, il Training Set nella porzione del quinto paziente raggiunge performance ottimali. Più nello specifico, il valore massimo di DSC viene raggiunto dalla slice n15 (Figura 3.36) ed è pari a 98.1%. In questa slice vengono predette molto bene le creste iliache e la parte centrale spina dorsale.

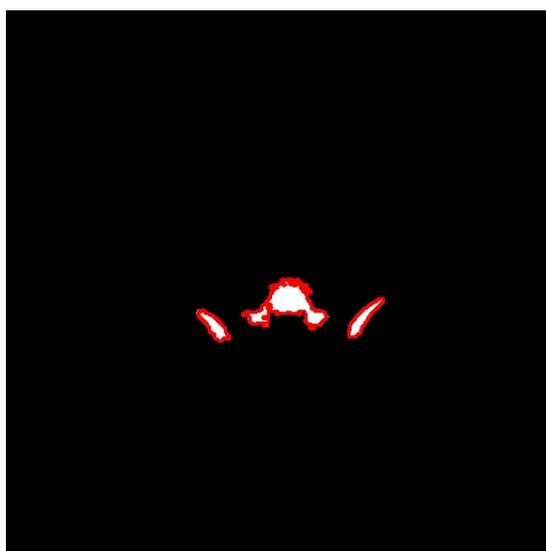


Figura 3.36: Slice n15 del paziente 5, il bordo rosso indica il contorno della MaskRM

Inoltre, dal box plot si osserva che una slice riesce a raggiungere una recall pari al 100%, questo è il caso della slice n114 (Figura 3.37) in cui viene predetta in maniera perfetta la parte delle creste iliache.



Figura 3.37: Slice n114 del paziente 5, il bordo rosso indica il contorno della MaskRM

Dalla porzione di Validation Set relativa al quinto paziente sono state rimosse le slice n210 e n213 poiché erano immagini con al più 2-3 pixel di RM ed il resto tutto sfondo nero. Le criticità quindi sono presenti nella slice n9 (e di conseguenza nella n12 poiché rappresentano la stessa sezione anatomica) e nella slice n168.

La slice n9 è visibile in Figura 3.38 e mostra la parte centrale della colonna vertebrale come nel caso nella slice n6. Anche in questo caso, le aree di RM sono molto irregolari e strette.

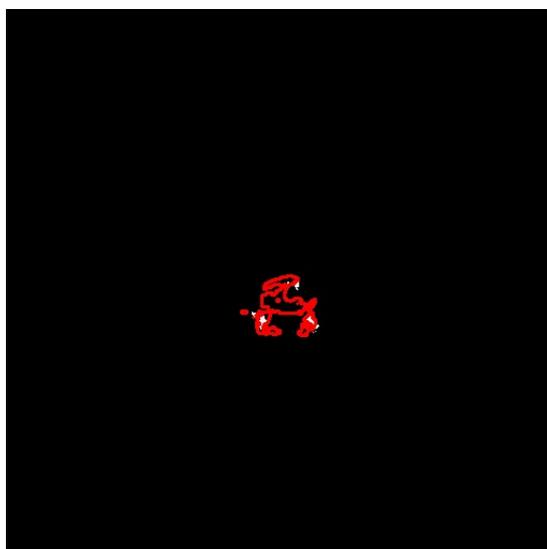


Figura 3.38: Slice n9 del paziente 5, il bordo rosso indica il contorno della MaskRM

Per quanto invece riguarda la slice n168 (Figura 3.39). In questa immagine si osserva una tendenza a sovra-campionare il midollo osseo rosso.

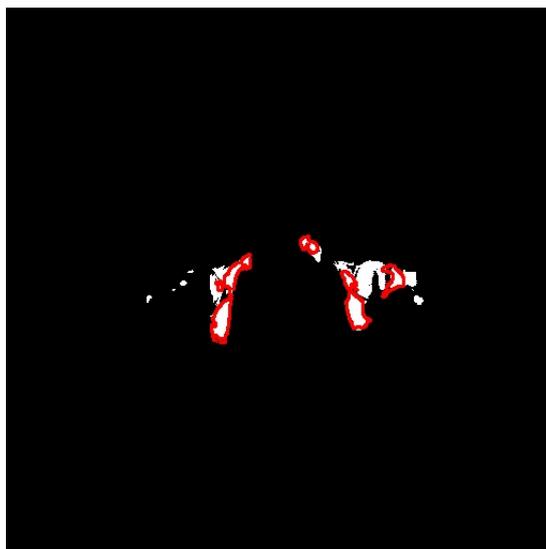


Figura 3.39: Slice n168 del paziente 5, il bordo rosso indica il contorno della MaskRM

U-Net riesce a raggiungere un valore massimo di DSC pari a 96.4% nel quinto paziente. Ciò avviene in corrispondenza della slice n45 (Figura 3.40); la rete riesce a predire bene la parte centrale del bacino.

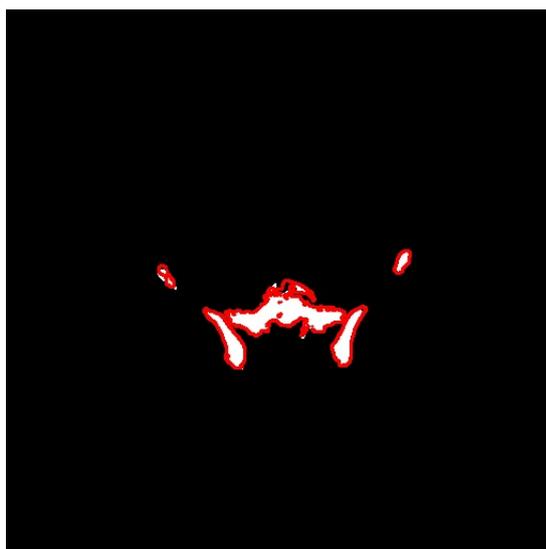


Figura 3.40: Slice n45 del paziente 5, il bordo rosso indica il contorno della MaskRM

La slice n120 (Figura 3.41), invece, conferma nuovamente l'ottima capacità della rete di predire la zona di RM nelle punte delle creste iliache ai lati.

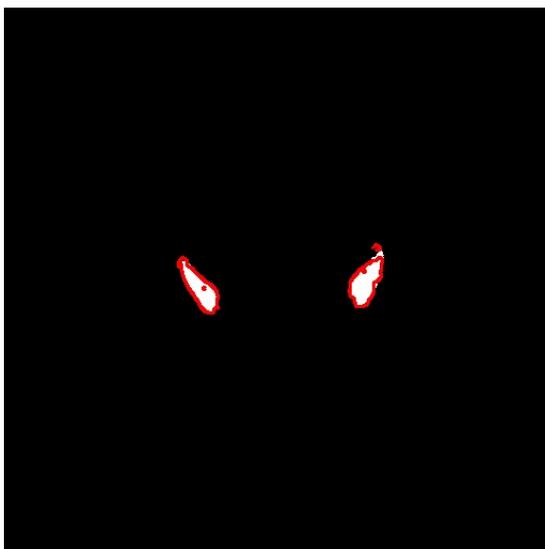


Figura 3.41: Slice n120 del paziente 5, il bordo rosso indica il contorno della MaskRM

3.3 Applicazione U-Net con Data Set suddiviso

Nel Capitolo 3.2, si è osservato il comportamento variabile che la rete U-Net presenta a seconda della sezione anatomica considerata nella slice. Dunque, si è reso necessario uno studio più approfondito di questo aspetto.

In particolare, partendo dal Data Set composto da 356 elementi, si è deciso di suddividerlo in 3 Data Set più piccoli, nei quali è presente 1/3 delle slice per ogni paziente. Ovviamente, non essendo il numero di slice per ogni paziente perfettamente divisibile in tre parti, il terzo Data Set presenterà slice in più rispetto alle altre 2. Nella Tabella 3.2 si ha una panoramica di tale suddivisione. Ogni sotto-Data Set è stato suddiviso 70% Training Set e 30% Validation Set.

Anche in questo caso, le slice di dimensione 512x512 sono state normalizzate come già spiegato nel Capitolo 3.2. Inoltre, la rete U-net è stata mantenuta uguale ed allenata, visto l'eccessivo peso computazionale, in cloud computing.

#Sezione	#Paziente	Slice Inizio	Slice Fine	Totale	Totale Sez	% Slice	% Sez
1	1	21.5	90.5	24	117	20.51%	32.87%
	2	48	111	22		18.80%	
	3	79	148	24		20.51%	
	4	5.5	71.5	23		19.66%	
	5	6	75	24		20.51%	
2	1	93.5	162.5	24	117	20.51%	32.87%
	2	114	177	22		18.80%	
	3	151	220	24		20.51%	
	4	74.5	140.5	23		19.66%	
	5	78	147	24		20.51%	
3	1	165.5	237.5	25	122	20.49%	34.27%
	2	180	246	23		18.85%	
	3	223	295	25		20.49%	
	4	143.5	215.5	25		20.49%	
	5	150	219	24		19.67%	
TOT Slice:					356		

Tabella 3.2: Schema riassuntivo della suddivisione fatta per la creazione dei tre Data Set.

É doveroso far presente che, in questo caso, si sono utilizzati due approcci di apprendimento. Il primo caso, come nel Capitolo 3.2, la rete è stata allenata nel predire la maschera di midollo osseo rosso (RM); nel secondo caso, la rete invece è stata allenata nel predire la maschera di midollo osseo giallo (YM). Tale maschera è stata ottenuta dalla differenza di della maschera binaria dell'intero midollo osseo (MaskBM) e la MaskRM. Questo approccio è stato pensato per provare a compensare le criticità anatomiche che si sono presentate in alcune sezioni anatomiche dei pazienti, infatti si è pensato che potendo predire in maniera più accurata la MaskYM, da essa sarebbe possibile arrivare, sottraendola alla MaskBM, la MaskRM.

3.3.1 Primo Data Set

Il primo Data Set è composto dalle prime 24 slice dei pazienti 1, 3 e 5; le slice del paziente 2 sono 22 (poiché di numero inferiore rispetto alle altre, come visibile nella Tabella 3.2), mentre il paziente 4 è composto da 23 slice in questo primo Data Set.

In generale, in questo Data Set sono presenti (a parte eccezioni dovuti ad arrotondamenti per eccesso e/o difetto delle slice) le zone relative alla colonna vertebrale, le creste iliache fino alla comparsa di una zona compatta centrale che è il bacino vero e proprio e la fossa iliaca.

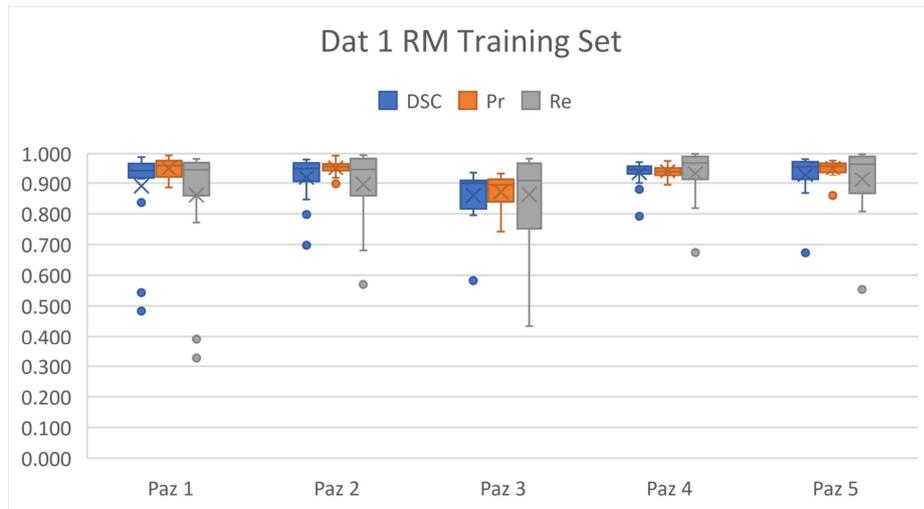
Come visibile dai box plot nelle Figure 3.42a e 3.42b relative alla prima suddivisione, si osserva che nel caso del Training Set, i pazienti 1, 4 e 5 hanno una buona simmetria. Più nello specifico, il paziente 1 mostra risultati di DSC molto alti, infatti il primo quartile si assesta sopra il 90% e mostra solo 2 outlier dove quello con performance minori ha un DSC pari al 50%. Anche la Re è molto alta in quanto il 25esimo percentile si assesta ad un valore maggiore del 85% e raggiungendo valore maggiore di poco sotto il 98%. Nel Validation Set, invece, il primo paziente non mostra più la stessa simmetria e di valori di DSC, Pr e Re sono più distribuiti ed in aggiunta, non sono presenti outlier. Più nello specifico, il primo quartile del DSC è sceso sotto il 70% mentre il terzo quartile si assesta intorno al 90%. Tuttavia, la recall è il parametro che viene meno variato rispetto al Training Set e presenta un 25esimo percentile che si assesta al valore di 80%.

Il paziente 2 evidenzia delle buone performance nel box plot del Training Set; presenta pochi outlier ed i valori di Pr sono molto alti rispetto agli altri pazienti, infatti si aggirano tutti superiori al 90%, indice che i FP sono in minoranza rispetto ai TP. Il DSC, invece, presenta come valore minimo 70%, ma il primo quartile si assesta intorno al 90%. Il Validation Set della porzione del paziente 2, resta coerente con i risultati del Training Set. Infatti, i tre parametri presentano le stesse simmetrie soltanto con valori lievemente inferiori. Il primo quartile del DSC si assesta intorno al 87%, invece la precisione resta comunque molto alta e sopra il 90%. In questo caso, la recall mostra un outlier con valore di poco sopra il 20%.

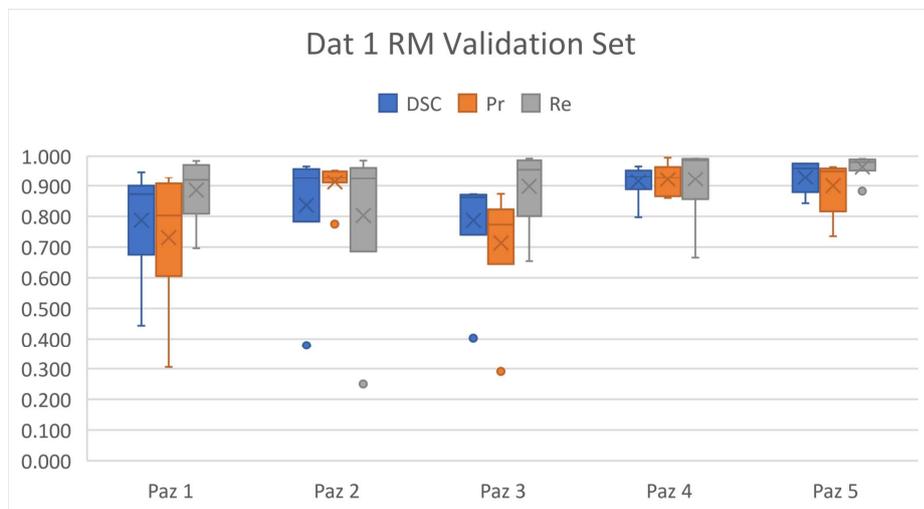
Il paziente 3 è quello che evidenzia più criticità. Come succedeva nel primo approccio descritto nel Capitolo 3.2, questo paziente mostra performance inferiori rispetto agli altri pazienti. Infatti nel Training Set, il primo quartile si assesta intorno al 80% per il DSC e la recall mostra una forma molto asimmetrica con un minimo di poco superiore al 40%. Nel Validation Set, il DSC e Re sono coerenti con i risultati ottenuti dalla rete nel predire RM nel Training Set. Tuttavia, la Pr scende ulteriormente con un primo quartile che si assesta intorno al 65%.

Come già scritto precedentemente, il paziente 4 è quello che mostra le migliori performance in termini di valori e simmetria. Infatti, DSC e Pr sono valore minimo intorno al 90% nel Training Set, mentre la Re intorno al 80%. Anche il Validation Set ha risultati promettenti e con l'assenza di outlier. Infatti il valore minimo del DSC si assesta al 80% ed il terzo quartile è intorno al 95%.

Il paziente 5 nel Training Set ha delle ottime performance; infatti, il valore di DSC già a partire dal primo quartile, ha valori al 90%. L'unico outlier presente ha un valore di DSC intorno al 68% e la sua recall è intorno al 55%. Nel Validation Set, la performance relativa al DSC si mantiene costante e sono assenti outlier. Il boxplot della Re mostra valori superiori al 90% e il DSC continua a mantenere risultati promettenti e superiori al 85% a partire dal 25esimo percentile.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.42

Lo studio del riconoscimento di midollo osseo giallo da parte di U-Net presenta delle performance (mostrate nei box plot nelle Figure 3.43a e 3.43b) inferiori rispetto al caso con RM. Infatti, solo il DSC è sempre inferiore al 90% in tutti e cinque i pazienti.

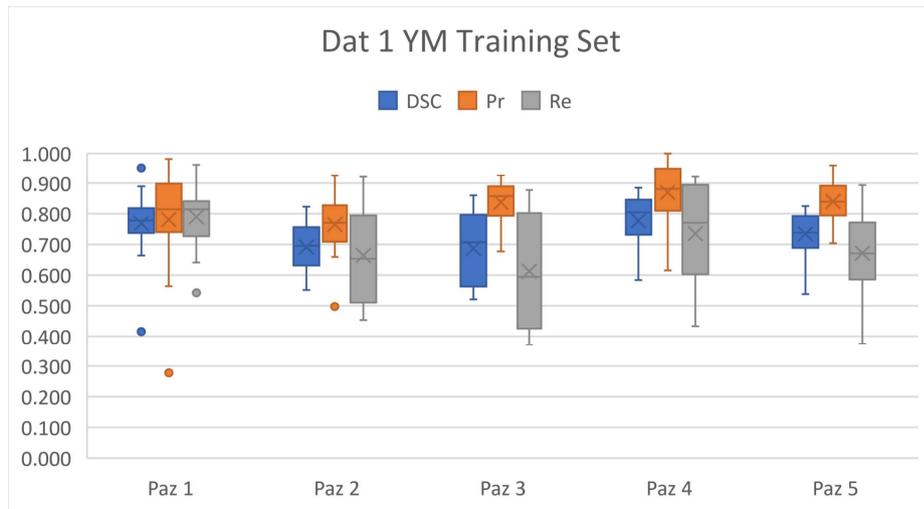
Andato nel dettaglio del paziente 1, il Training Set ha performance nella media; difatti, il DSC ha una distribuzione di valori simmetrica intorno al 75% e 80%, con massimo di poco inferiore al 90% e minimo intorno al 65%. In realtà è presente un outlier con valore di DSC intorno al 40%, ma anche un outlier con valore intorno al 95%. Pr e Re sono meno simmetrici ed il primo ha un valore di minimo appena sotto il 60%. Invece, ambedue hanno valore massimo superiore al 90%. Nel Validation Set, si notano meglio le difficoltà da parte della rete U-Net di individuare la parte di midollo osseo giallo. Più nello specifico le performance del paziente 1 non sono comparabili con il caso di riconoscimento di zone RM. Infatti, il valore massimo di DSC è sotto il 75%, mentre per la Re si ha valore massimo appena sotto il 70%.

La porzione di Training Set YM del paziente 2 presenta performance comparabili con la porzione del paziente 1; anche in questo caso i valori di DSC, Pr e Re sono inferiori rispetto alla controparte nel riconoscimento di RM ed in particolare, il valore massimo di DSC è di poco superiore al 80%. Nella porzione di Validation Set tutti e tre i parametri hanno valori inferiori al 80%, l'unica eccezione è la slice n48 in cui il valore di Pr è pari al 100%, questo perché si ha un drastico fenomeno di sotto-campionamento delle zone di YM che porta ad un numero molto basso di FP.

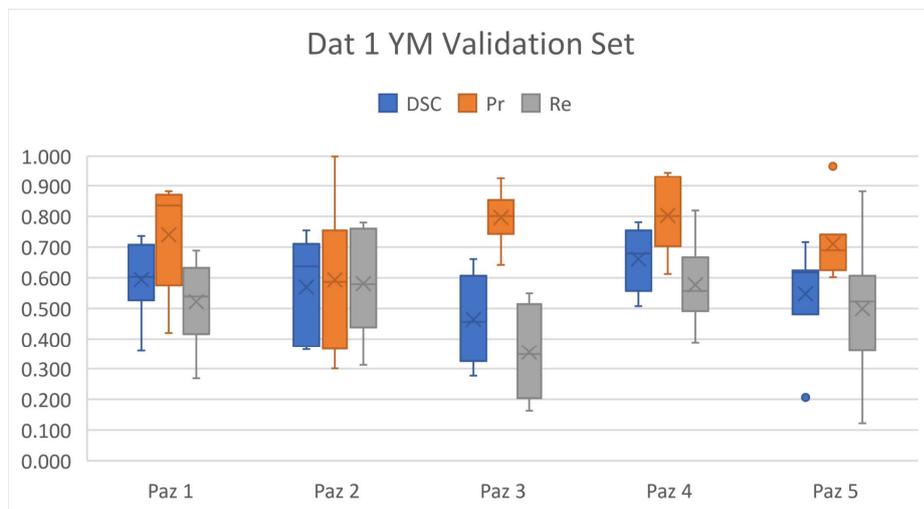
Il terzo paziente è nuovamente quello più critico. La porzione di Training Set ha valori comunque comparabili con tutti gli altri pazienti, infatti il DSC è inferiore al 90% e valore minimo poco superiore al 50%; la distribuzione dei valori di recall è molto varia e va da un minimo pari al 35% fino ad un massimo di poco inferiore al 90%. Nel Validation Set, invece, si osserva che DSC e recall si abbassano e inferiori al 60%, mentre la precisione resta molto alta, ciò è indice che i TP (che in questo caso sono i pixel di YM) sono maggiori rispetto ai FP.

Il quarto paziente è l'unico ad avere performance lievemente superiori rispetto agli altri nel Training Set. Infatti il DSC nel terzo quartile si assesta intorno agli 85%, mentre il valore minimo è quasi al 60%; in aggiunta, non sono presenti outlier. Il Validation Set ha comportamento uguale ai paziente 1 e 5, ma con performance lievemente superiori. Anche in questo caso si ha una slice in cui si raggiunge il valore di 100% nella precisione, mentre il DSC ha valore massimo inferiore al 80%.

Il quinto paziente ha lo stesso andamento del paziente 1, con la differenza che la recall ha un comportamento simmetrico, ma una larga distribuzione di valori. Infatti, il valore massimo è poco inferiore al 90%.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.43

Risultati Paziente 1 - RM

Nel Training Set relativo alla prima suddivisione del paziente 1 usata per allenare la rete all'individuazione del RM, sono presenti 5 outlier. Le due slice che corrispondono ai valori più bassi di DSC sono, rispettivamente, la n45.5 (Figura 3.44) e la n27.5 (Figura 3.45). Nella prima slice presa in considerazione, si ha un DSC pari al 48.3%, infatti si evidenzia un problema di sotto-campionamento di tutte le aree in cui è presente il RM, in particolare la parte centrale.

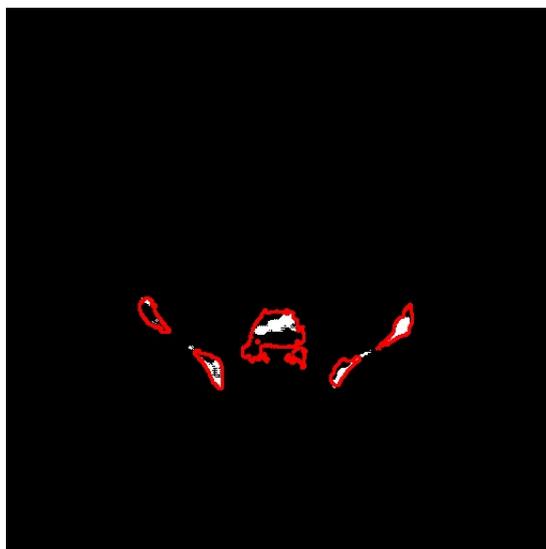


Figura 3.44: Slice n45.5 del paziente 1 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Anche in questa slice si ha un grave problema di sotto-campionamento della stessa di zona RM centrale vista nella slice precedente, e conseguentemente, un alto tasso di FN che abbassa radicalmente il valore di recall fino a 39.1%. Inoltre, il valore di DSC è pari al 54.3%. È doveroso osservare che questa slice era ben predetta quando si aveva l'intero Training Set e raggiungeva un DSC 92.8%.

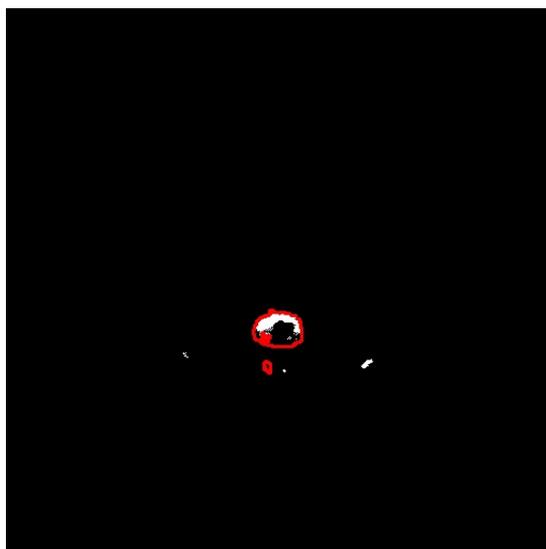


Figura 3.45: Slice n27.5 del paziente 1 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Tuttavia, la rete è riuscita a predire in maniera quasi eccellente la slice n21.5 (Figura 3.46) mostrando un DSC molto alto e pari al 98.8%. La zona rappresentata è la colonna vertebrale e quando era stata predetta in un intero Training Set aveva un valore di DSC pari al 66.8%.

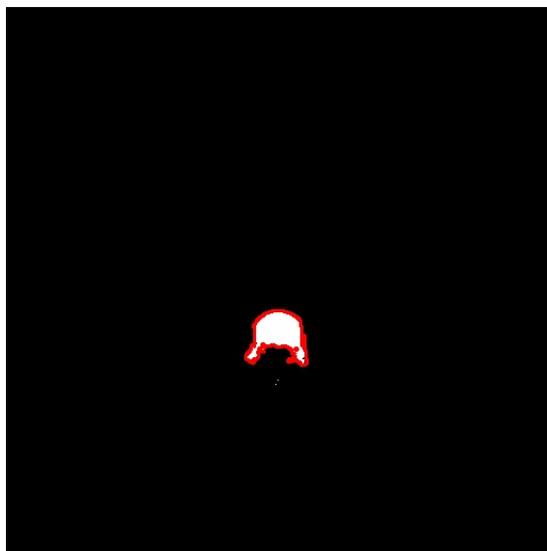


Figura 3.46: Slice n21.5 del paziente 1 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

A parte queste criticità iniziali, la slice n90 è una delle tante slice che invece ha mantenuto un DSC molto alto, in questo caso pari al 97.5%. Ciò fa dedurre che la parte iniziale delle slice, quindi quella in cui si ha prevalentemente la colonna vertebrale, sono fortemente influenzate dal Training Set che si usa.

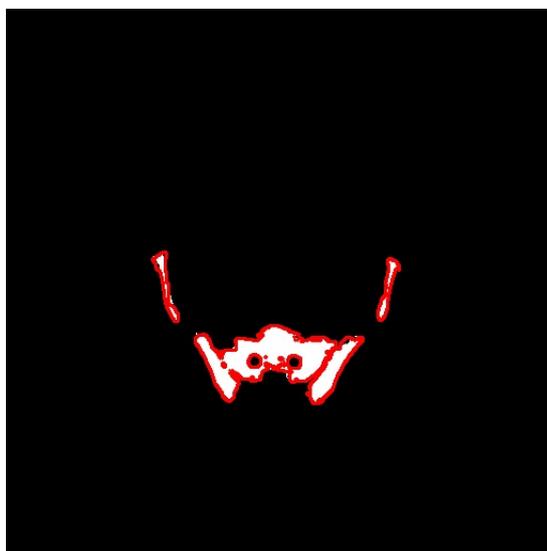


Figura 3.47: Slice n90.5 del paziente 1 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Nel Validation Set la slice con valore minimo di DSC e Pr è la slice n36.5 (Figura 3.48). Questa slice era già stata trattata nel caso dell'intero Training Set, però presentava un DSC pari a 77.5%. In questo caso si evidenzia un sovra-campionamento non visibile nel primo approccio.

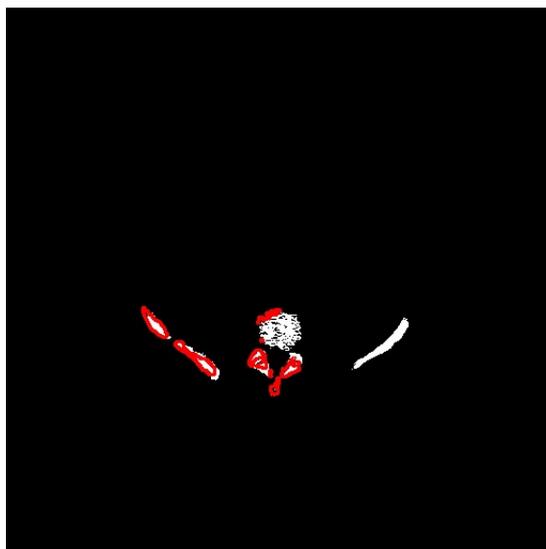


Figura 3.48: Slice n36.5 del paziente 1 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Anche la slice n39.5 (Figura 3.49) che è subito successiva alla precedente presenta gravi difficoltà, indicando che questa prima porzione di slice ha difficoltà rispetto alle slice della porzione finale.



Figura 3.49: Slice n39.5 del paziente 1 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

In conferma di quanto detto prima, le restanti slice, che identificano la parte di bacino compatta hanno valori di DSC promettenti, ne sono un esempio le slice n66.5 (Figura 3.50) e n75.5 (Figura 3.51).



Figura 3.50: Slice n66.5 del paziente 1 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

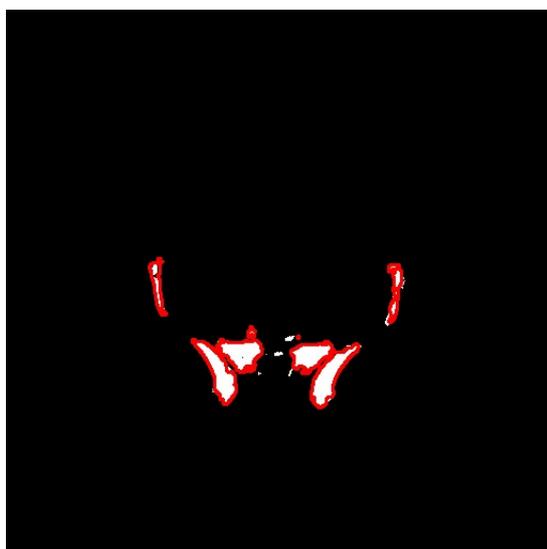


Figura 3.51: Slice n75.5 del paziente 1 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 1 - YM

Come si vede già dal Training Set, allenare la rete al riconoscimento di YM, nella maggior parte dei casi non ha portato a performance comparabili con l'individuazione di RM. Il caso peggiore è la slice n45.5 (Figura 3.52) con DSC pari a 41.5% e Pr pari al 27.8%. Essa era già stata esaminata nel caso RM e non aveva avuto performance ottimali. Si osserva un alto tasso di sovra-campionamento e quindi di FP, inoltre le zone di YM sono troppo strette e frastagliate.

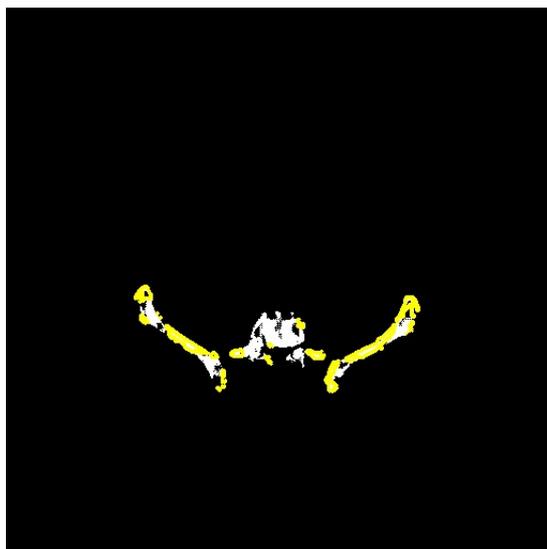


Figura 3.52: Slice n45.5 del paziente 1 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Anche la slice n90.5 precedentemente vista non ottiene performance ottimali, con valore di DSC pari al 66.4%, mentre prima otteneva circa 97.5%.

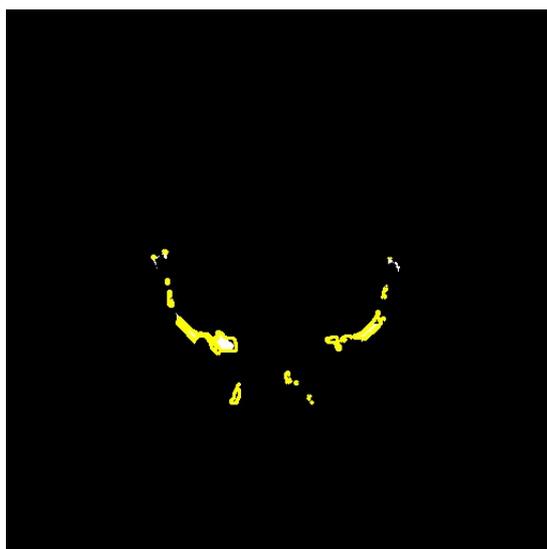


Figura 3.53: Slice n90.5 del paziente 1 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Sono pochi i casi in cui la predizione di YM va a buon fine e sono quei casi in cui la rete non riesce a predire efficacemente le zone di midollo osseo rosso. Due esempi sono le slice n21.5 (Figura 3.54) vista precedentemente e la slice n72.5 (Figura 3.55). La prima slice riesce ad essere predetta bene seppur formata da aree molto piccole; lo stesso avviene nella seconda con la differenza che si ha un certo sovra-campionamento di YM.



Figura 3.54: Slice n21.5 del paziente 1 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

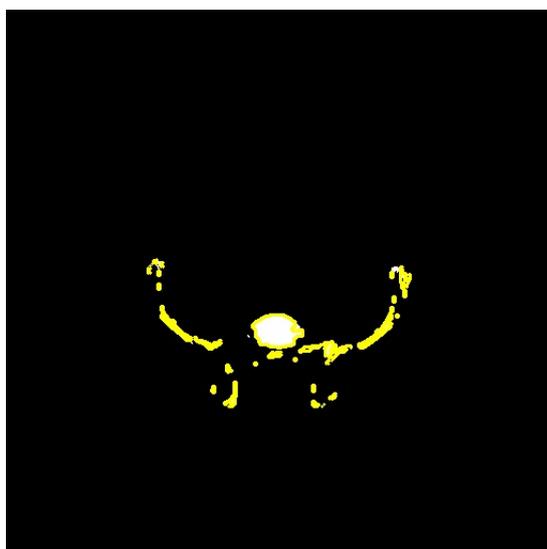


Figura 3.55: Slice n72.5 del paziente 1 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 2 - RM

Nella porzione del Training Set relativa al secondo paziente, soltanto una slice scende sotto il valore di DSC di 80% ed è la slice n54 (Figura 3.56). In essa si evidenzia l'assenza di individuazione di midollo osseo rosso nella parte centrale.

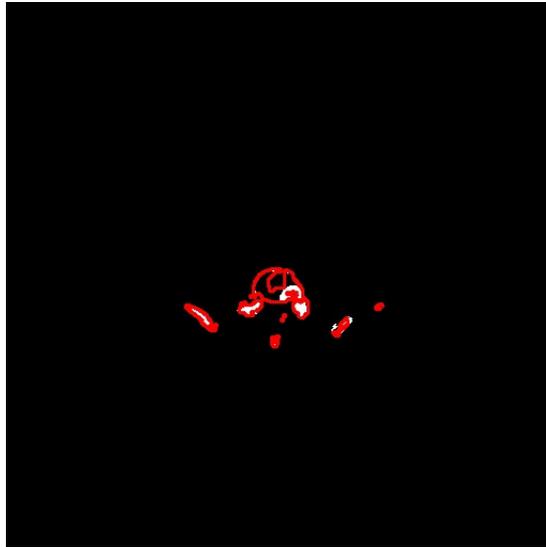


Figura 3.56: Slice n54 del paziente 2 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Nel secondo paziente, la slice con le migliori performance nel Training Set per quanto riguarda il DSC è la n99 (Figura 3.57); questa slice era già stata vista nel caso con l'intero Training Set e aveva sempre ottenuto le performance migliori.

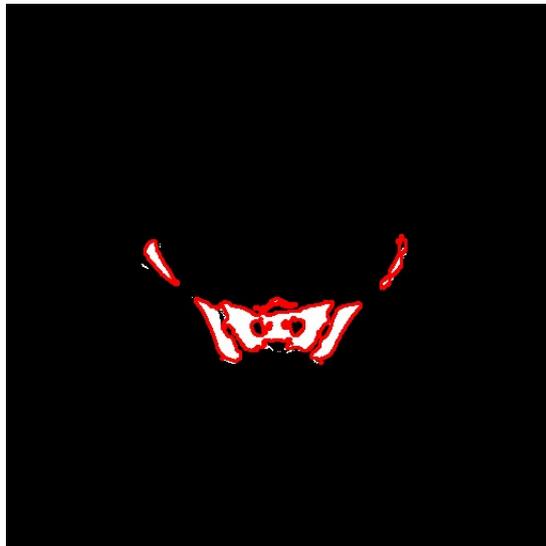


Figura 3.57: Slice n99 del paziente 2 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Un'altra slice che ha ottenuto un ottimo risultato di DSC è la n63 (Figura 3.58); questa slice era stata analizzata nel Validation Set completo con ottimi risultati.

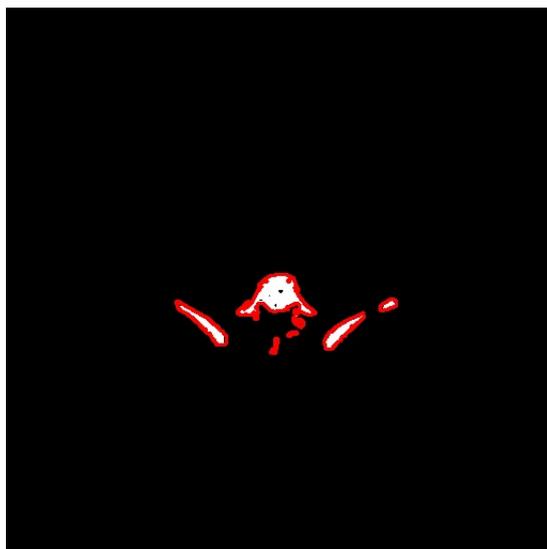


Figura 3.58: Slice n63 del paziente 2 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Nella prima suddivisione del Validation Set il paziente 2 mostra due criticità. Una è stata rimossa dal box plot poiché consistente nella slice n48 quasi totalmente nera e conseguentemente con un DSC pari a 0. La seconda criticità è data dalla slice n51 (Figura 3.59) precedentemente studiata nel Training Set completo con DSC pari a 66%. In questo caso ha un DSC pari a 37.8%.

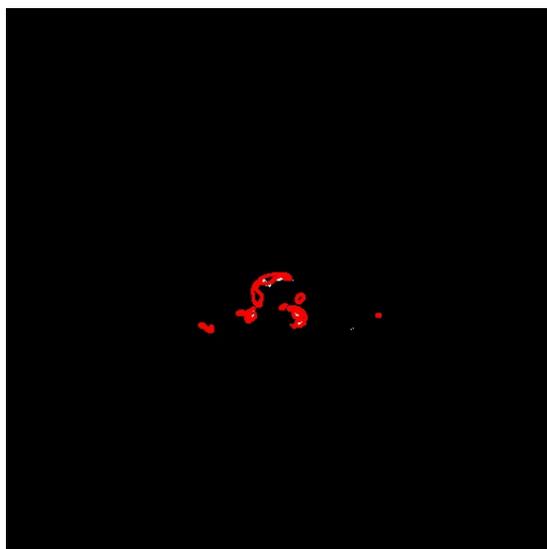


Figura 3.59: Slice n51 del paziente 2 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

La slice in cui si ottiene il valore massimo di DSC pari al 96.6% è la n90 (Figura 3.60). Questa slice e la n105 (Figura 3.61) confermano l'ottima capacità da parte della rete di riconoscere le zone di RM nel parte centrale del bacino e quella del coccige.

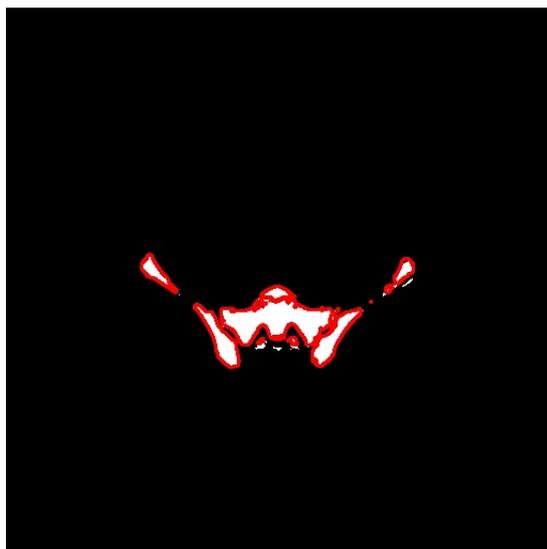


Figura 3.60: Slice n90 del paziente 2 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

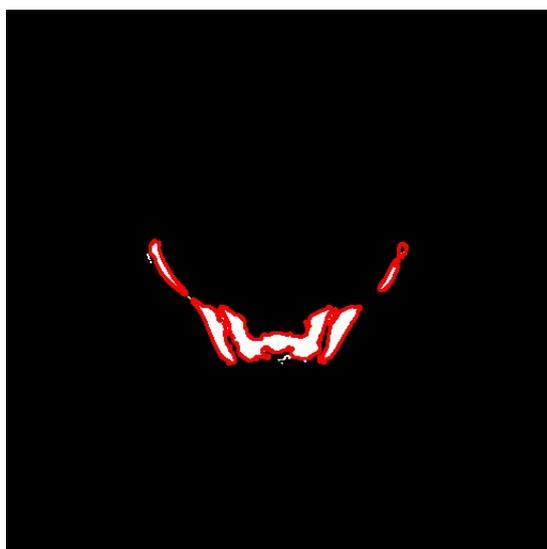


Figura 3.61: Slice n105 del paziente 2 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 2 - YM

In generale, le performance del paziente 2 nel Training Set 2 non sono ottimali, un outlier è dato dalla slice n69 (Figura 3.62) che ha una Re pari a 49.8%.

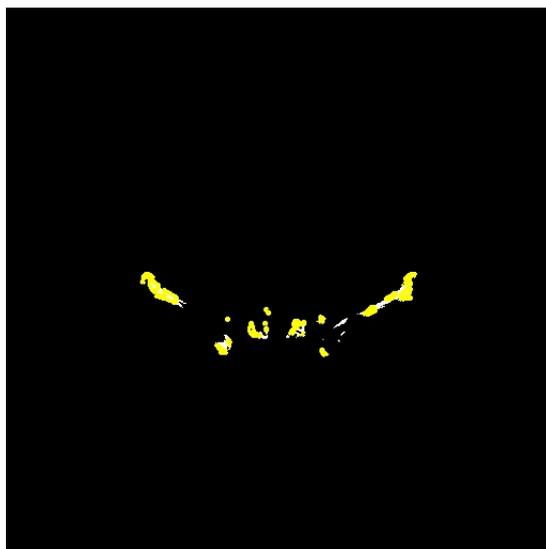


Figura 3.62: Slice n69 del paziente 2 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Un'altra slice con un valore basso di DSC è la n99 (Figura 3.63, vista in precedenza nel caso RM e aveva ottenuto ottime performance. In questo caso, la zona di YM è molto stretta e frastagliata.

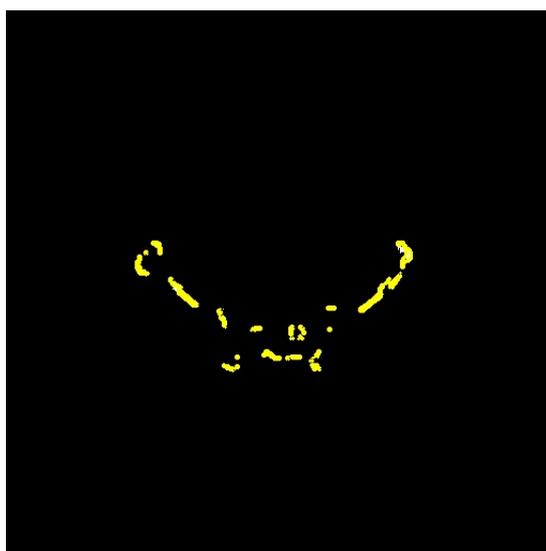


Figura 3.63: Slice n99 del paziente 2 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Pur non raggiungendo mai il valore di 90% del DSC, la slice n63 (Figura 3.64 è da osservare poiché trattata in precedenza e ha ottenuto delle ottime performance. Questa slice ottiene valore pari a 81.3% anche se è particolarmente frastagliata.



Figura 3.64: Slice n63 del paziente 2 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Anche nel Validation Set non si hanno performance ottimali, infatti la DSC massima è pari al 75.5%. Le criticità più evidenti sono date dalla n105 (Figura 3.65) che in precedenza aveva dato ottime performance e n90 (Figura 3.66), anch'essa analizzata in precedenza con risultati promettenti.

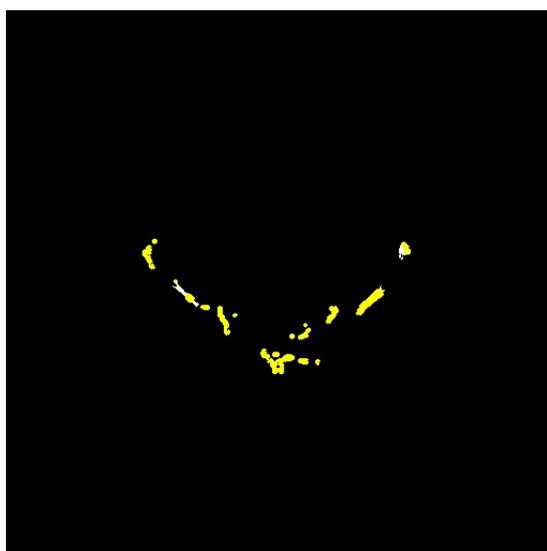


Figura 3.65: Slice n105 del paziente 2 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

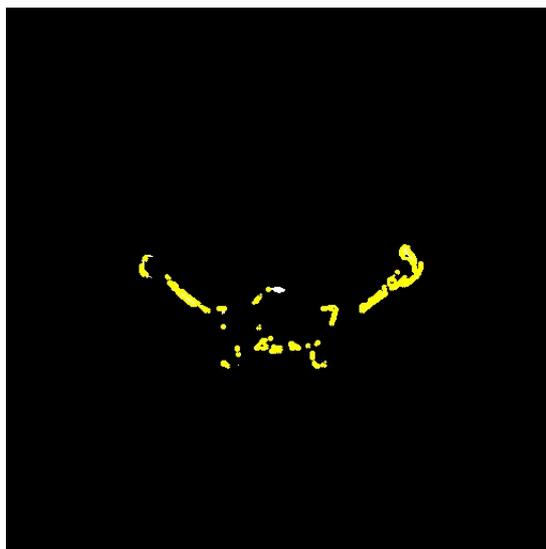


Figura 3.66: Slice n90 del paziente 2 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

L'unica slice che raggiunge un valore di DSC accettabile pari a 75.5% è la n51 (Figura 3.67) che in precedenza aveva dato performance molto basse pari a 37.8%.



Figura 3.67: Slice n51 del paziente 2 relativa alla prima suddivisione del Validation SetYM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 3 RM

Nella porzione di Training Set appartenente al paziente 3, soltanto una slice presenta una criticità, questa è la slice n85 (Figura 3.68) con DSC pari a 58.3%. Come si vede, la zona di RM è molto frastagliata e stretta.

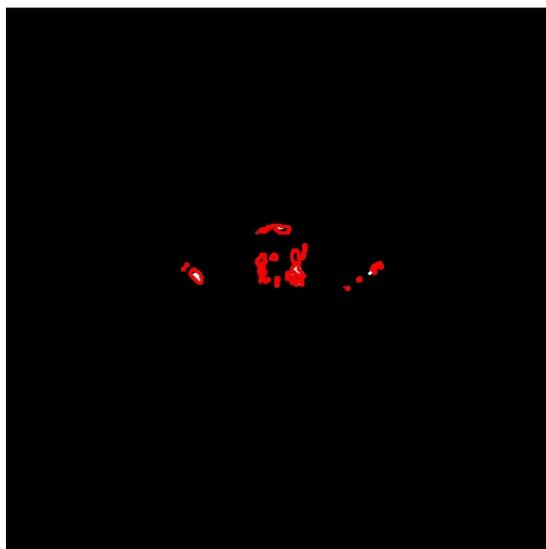


Figura 3.68: Slice n85 del paziente 3 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Le slice n118 (Figura 3.69) e n133 (Figura 3.70) evidenziano, con DSC superiori al 90%, come la zona RM del bacino centrale e le creste iliache siano state predette in maniera performante.

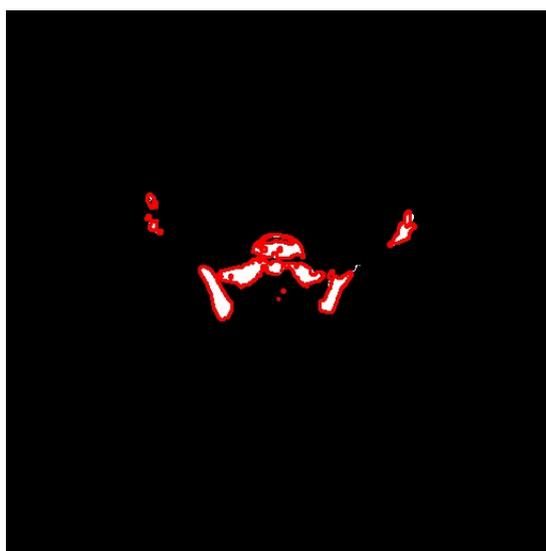


Figura 3.69: Slice n118 del paziente 3 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

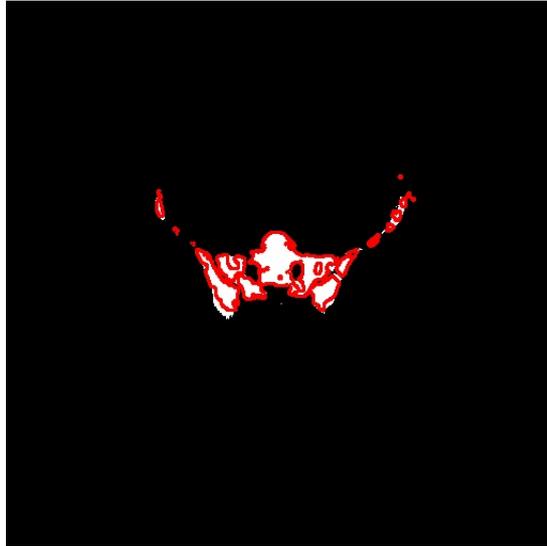


Figura 3.70: Slice n133 del paziente 3 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Anche nel Validation Set, solo una slice porta a criticità, ovvero la n82 (Figura 3.71). Essa è una slice vicina alla n85 vista nel Training Set, confermando che quella particolare zona anatomica è difficile da predire per la rete.

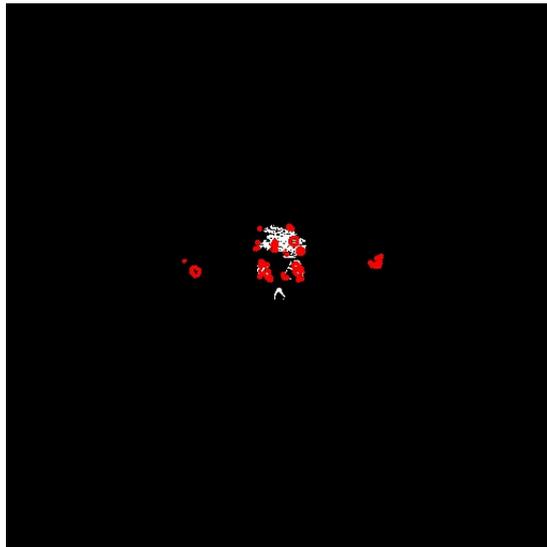


Figura 3.71: Slice n82 del paziente 3 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Le slice n115 (Figura 3.72) e n127 (Figura 3.73), sono due esempi di immagini presenti nel Validation Set che vengono predette molto bene. Tuttavia il DSC non supera il 90%.

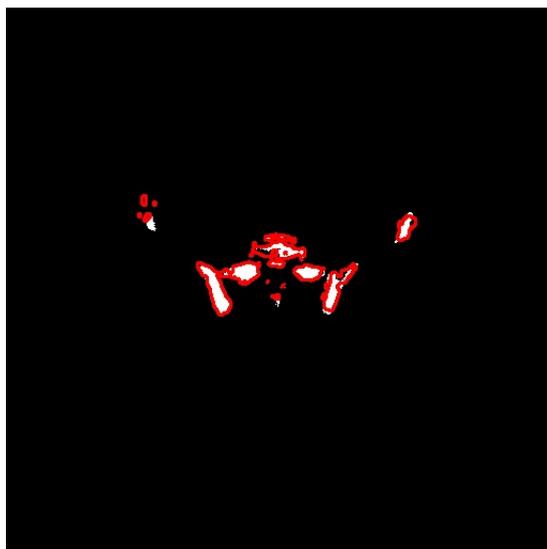


Figura 3.72: Slice n115 del paziente 3 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

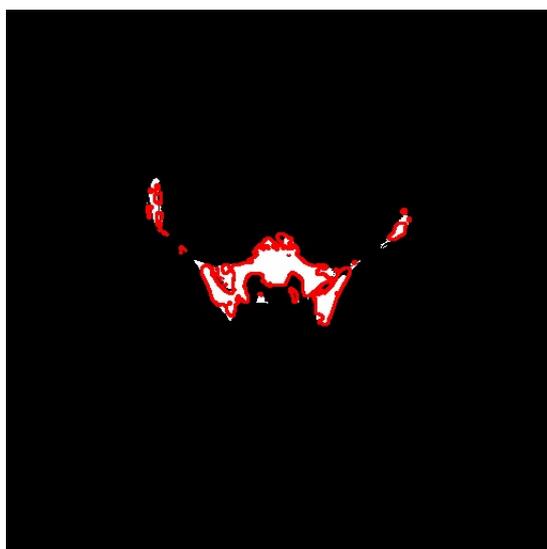


Figura 3.73: Slice n127 del paziente 3 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 3 YM

Il Training Set YM per il paziente 3 non ha evidenziato risultati performanti; infatti, sono più i casi in cui il DSC è inferiore al 70%. La slice n133 (Figura 3.74) era già stata studiata nel caso RM con risultati promettenti, mentre in questo caso presenta un DSC pari a 52.1%.

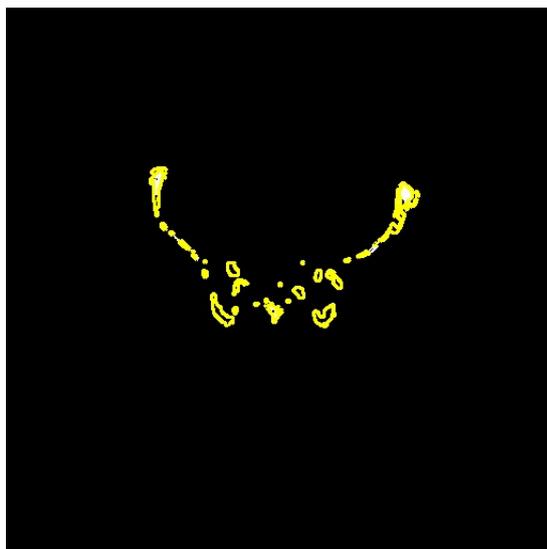


Figura 3.74: Slice n133 del paziente 3 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

La slice n121 (Figura 3.75) ha un DSC pari al 52.9% e quindi molto basso. Ciò è dovuta alle zone troppo strette ed irregolari del YM.

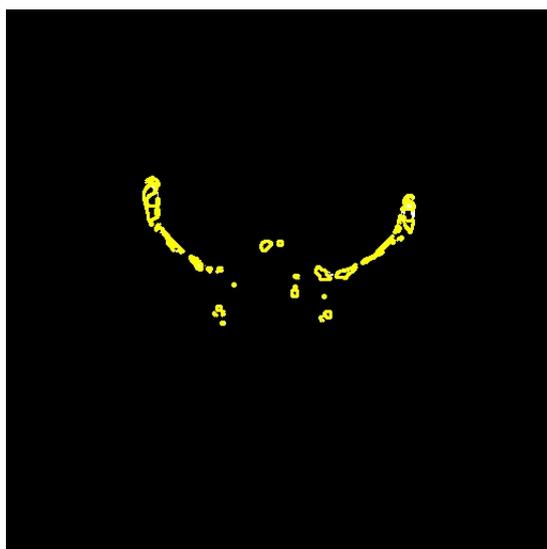


Figura 3.75: Slice n121 del paziente 3 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

La slice n97 (Figura 3.76) è l'unica slice ad ottenere un valore di DSC alto e pari a 86.1%. Infatti in questo caso la rete è riuscita a predire correttamente le zone di YM nelle creste iliache che in questa particolare sezione anatomica sono più visibili.

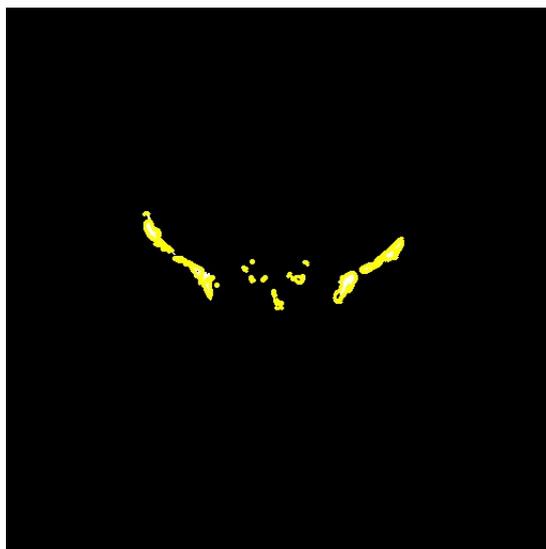


Figura 3.76: Slice n97 del paziente 3 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Il Validation Set per il paziente 3 non mostra risultati performanti e sono tutti con un DSC inferiore al 70%. Le slice n139 (Figura 3.77) e n136 (Figura 3.78) sono un esempio di sotto-campionamento delle zone YM.

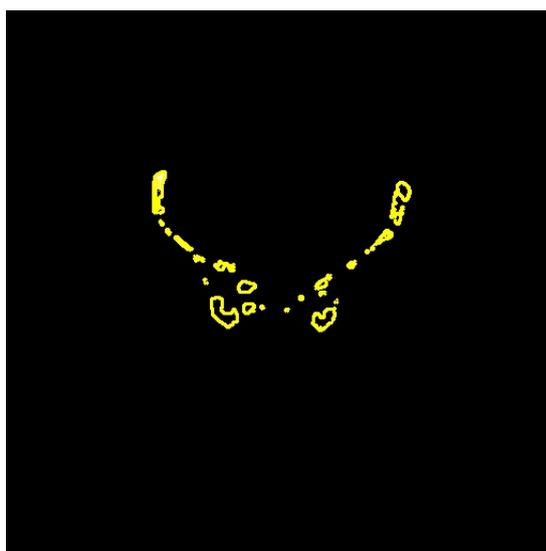


Figura 3.77: Slice n139 del paziente 3 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

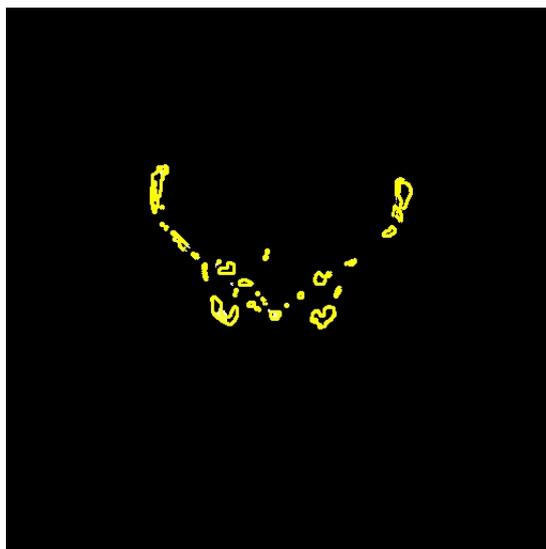


Figura 3.78: Slice n136 del paziente 3 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 4 - RM

Il Training Set RM relativo al quarto paziente presenta tutti valori molto alti di DSC ed in generale superiori al 90%. L'unica criticità è la slice n8.5, anche se in quest caso il DSC è a 79.8%. Si osserva una difficoltà, già riscontrata nelle slice dei precedenti pazienti, nel predire le zone di RM della colonna vertebrale.

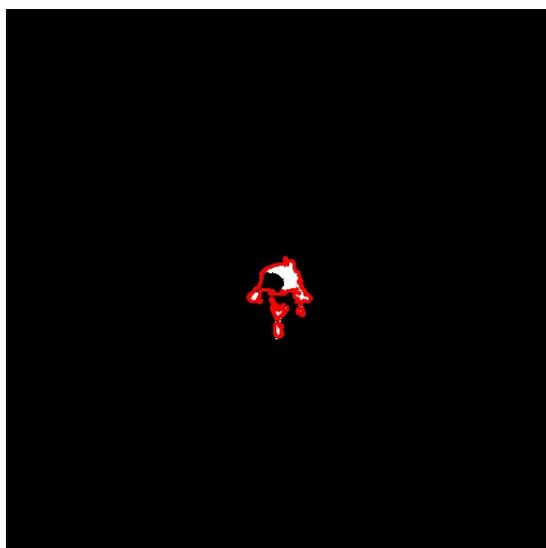


Figura 3.79: Slice n8.5 del paziente 4 relativa alla prima suddivisione del Training Set.

Tutte le altre slice, riesco a predire bene la parte centrale, ma le performance migliori vengono ottenute nella predizione delle aree RM che si hanno nella sezione centrale del bacino, come visibile nelle slice n68.5 e n53.5 (Figure 3.80 e 3.81).

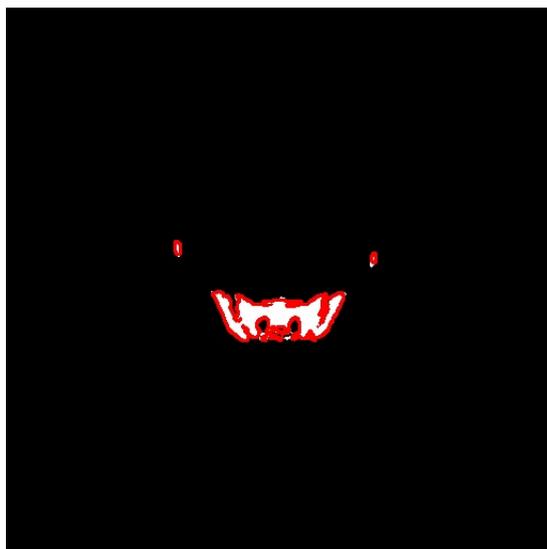


Figura 3.80: Slice n68.5 del paziente 4 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

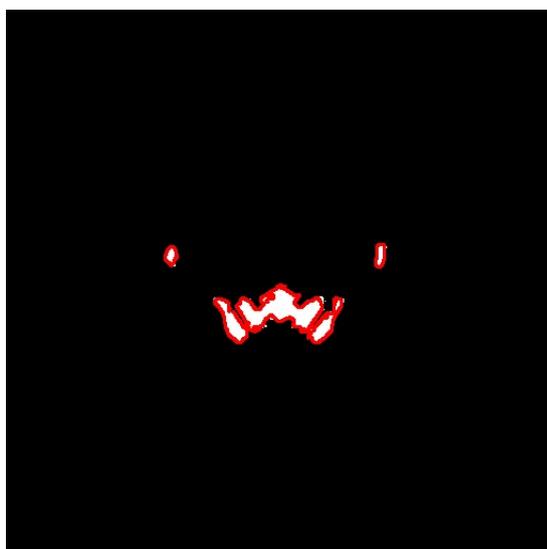


Figura 3.81: Slice n53.5 del paziente 4 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Anche il Validation Set dimostra come l'iniziale sezione della colonna vertebrale, quando ancora non si vedono le creste iliache, crea difficoltà di predizione alla rete; un esempio è la slice n5.5 (Figura 3.82) che ha, tuttavia, un DSC pari al 79.8%.

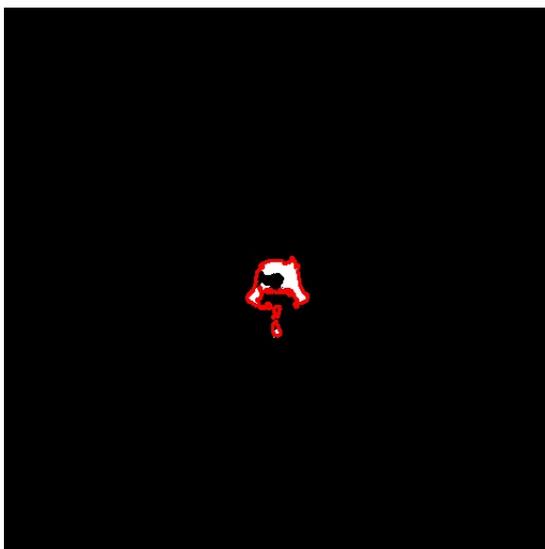


Figura 3.82: Slice n5.5 del paziente 4 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

A parte l'iniziale criticità, il resto delle slice del paziente 4 vengono predette in maniera performante. Più nello specifico, la parte di bacino compatto, come dimostrato ulteriormente dalle slice n59.5 e n65.6 (Figure 3.83 e 3.84).



Figura 3.83: Slice n59.5 del paziente 4 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

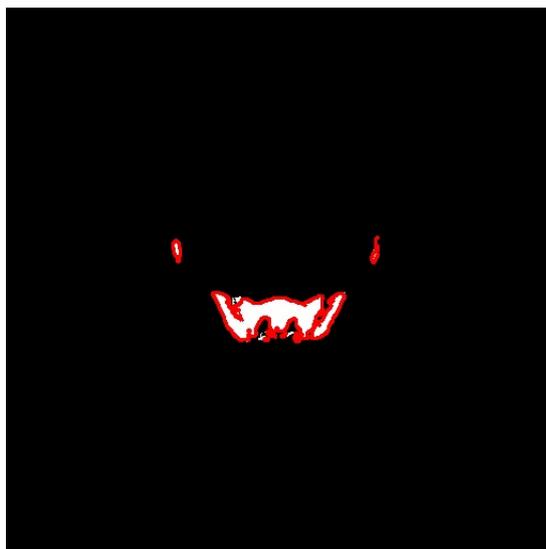


Figura 3.84: Slice n65.5 del paziente 4 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 4 - YM

Le difficoltà nella predizione di zone YM si riflettono anche nel paziente 4, il quale nel caso RM ha sempre ottenuto risultati performanti. Più in dettaglio, nel Training Set YM le slice n47.5 e n56.5 (Figure 3.85 e 3.86) ottengono un DSC rispettivamente di 60.5% e 67.1% e mostrano come, in questa sezione anatomica, la zona di YM è ridotta e frastagliata.

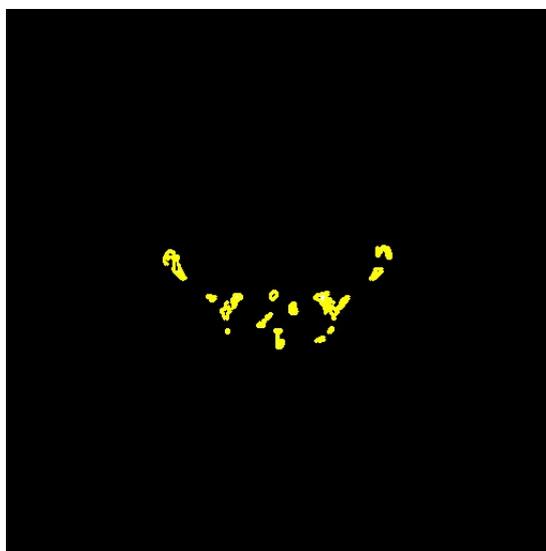


Figura 3.85: Slice n47.5 del paziente 4 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

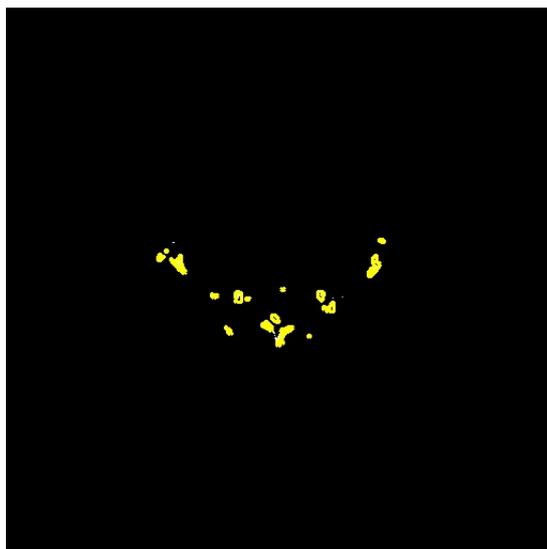


Figura 3.86: Slice n56.5 del paziente 4 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

La rete riesce a predire abbastanza bene la zona di YM presente nelle creste iliache, anche se essa è molto ristretta. Un esempio sono la slice n23.5 (Figura 3.87) e la slice n29.5 (Figura 3.88) con ambedue con DSC pari al 86%.



Figura 3.87: Slice n23.5 del paziente 4 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

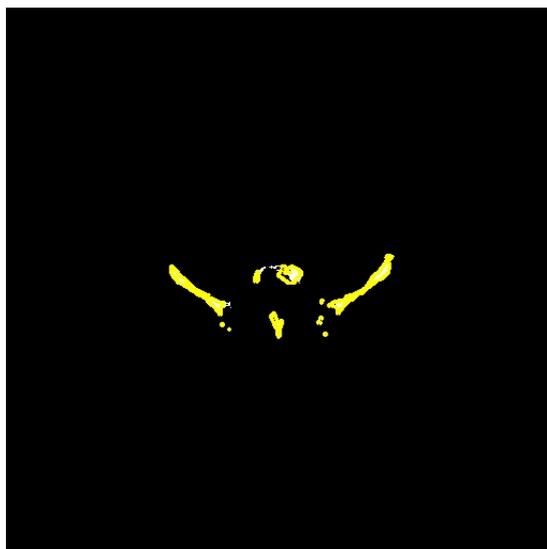


Figura 3.88: Slice n29.5 del paziente 4 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Il Validation Set riflette le difficoltà della rete incontrate nel Training Set, infatti il paziente 4 non riesce a raggiungere un DSC pari ad 80%. Le slice n59.5 e n71.5 sono un esempio delle difficoltà incontrate dalla rete che non riesce a predire queste aree irregolari. Si osservi che nel caso della slice n71.5 si hanno avuto problemi con la predizione di zone YM presenti nelle creste iliache.

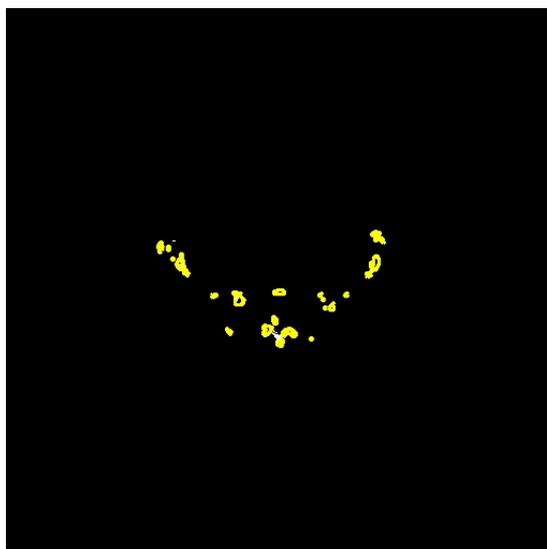


Figura 3.89: Slice n59.5 del paziente 4 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

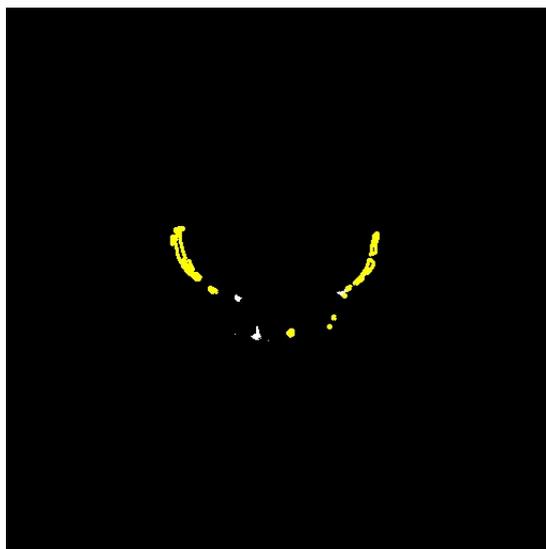


Figura 3.90: Slice n71.5 del paziente 4 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 5 - RM

La prima suddivisione del Training Set RM del paziente 5 ha nuovamente evidenziato una difficoltà nella predizione di zone di midollo osseo attivo nella parte centrale; nella slice n6 (Figura 3.91) ciò è causato da aree che sono molto irregolari e sparse.



Figura 3.91: Slice n6 del paziente 5 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Come già descritto nei precedenti pazienti, la sezione centrale del bacino viene predetta in maniera performante. Questo è il caso delle slice n54 e n60 (Figure 3.92 e 3.93) dove si ha un DSC superiore al 95%.

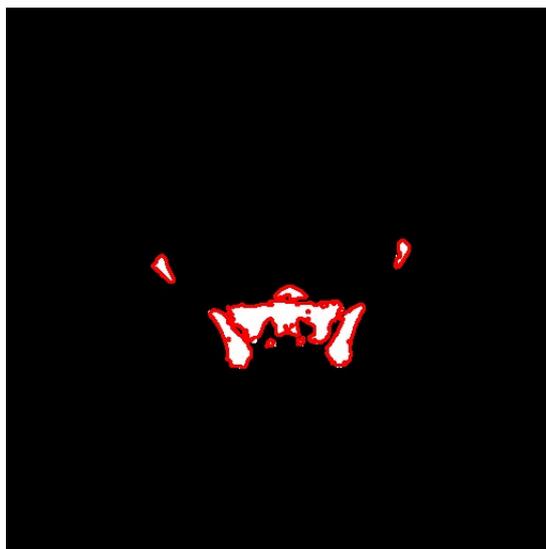


Figura 3.92: Slice n54 del paziente 5 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.



Figura 3.93: Slice n60 del paziente 5 relativa alla prima suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Il Validation Set RM è coerente con i risultati ottenuti nel corrispettivo Training Set. Infatti si hanno problemi di individuazione delle zone RM nella zona centrale della colonna vertebrale, come visibile nella Figura 3.94. Tuttavia, l'eccellente predizione delle zone RM nelle slice n48 e n57 (Figure 3.95 e 3.96) conferma come la rete riesca a lavorare in maniera promettente con queste sezioni anatomiche.

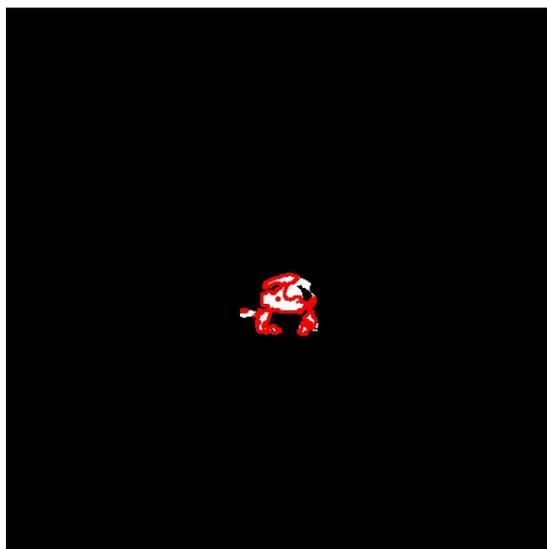


Figura 3.94: Slice n9 del paziente 5 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

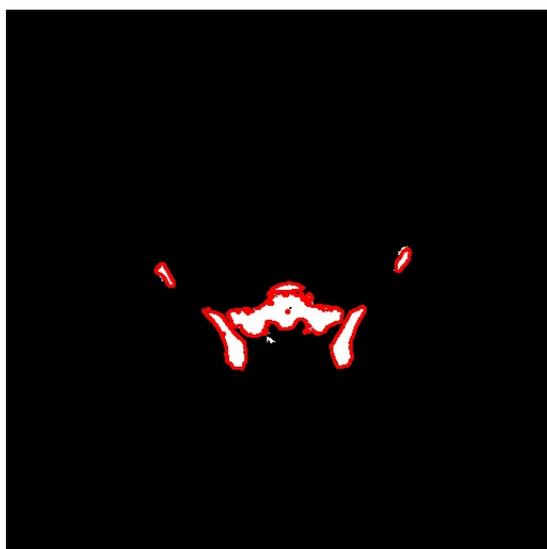


Figura 3.95: Slice n48 del paziente 5 relativa alla prima suddivisione del Validation SetRM, i bordi rossi indicano la MaskRM effettiva.

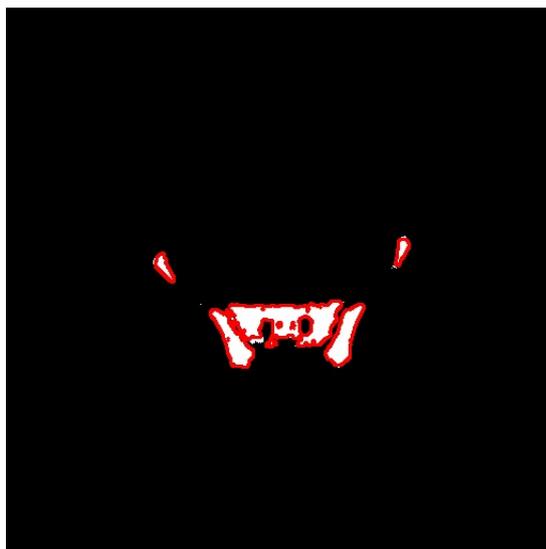


Figura 3.96: Slice n57 del paziente 5 relativa alla prima suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 5 - YM

La porzione di Training Set YM del paziente 5 conferma quanto descritto nei precedenti pazienti. Più nel dettaglio si hanno valori di DSC molto bassi e mediamente inferiori al 80%. Le slice n63 e n54 (Figure 3.97 e 3.98) ne sono la prova con DSC pari, rispettivamente, a 62.6% e 53.8%. La slice n54 era già stata trattata nel caso RM con performance ottimali poiché si trattava della parte del bacino.

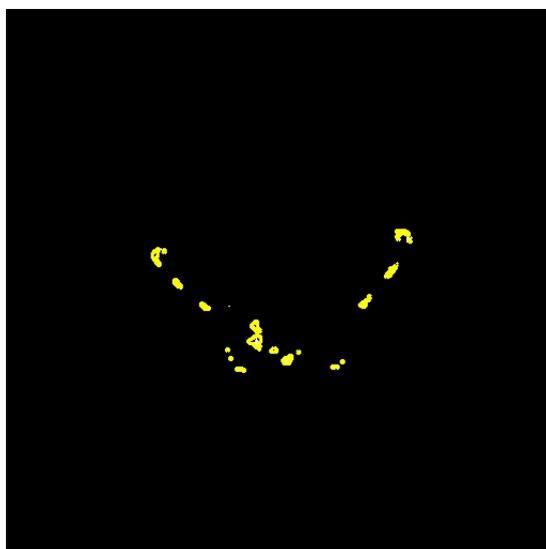


Figura 3.97: Slice n63 del paziente 5 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

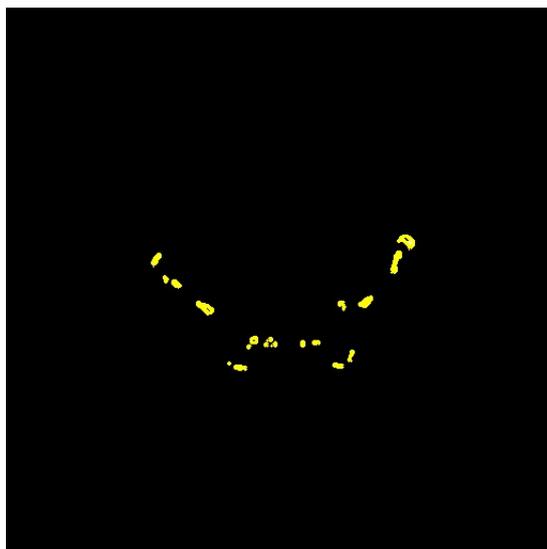


Figura 3.98: Slice n54 del paziente 5 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

La slice n36 è una delle poche con un DSC superiore al 80%, infatti raggiunge valore 82.5%. È importante notare che questa slice rappresenta le zone di YM nelle creste iliache che, seppur strette, vengono predette senza problemi.

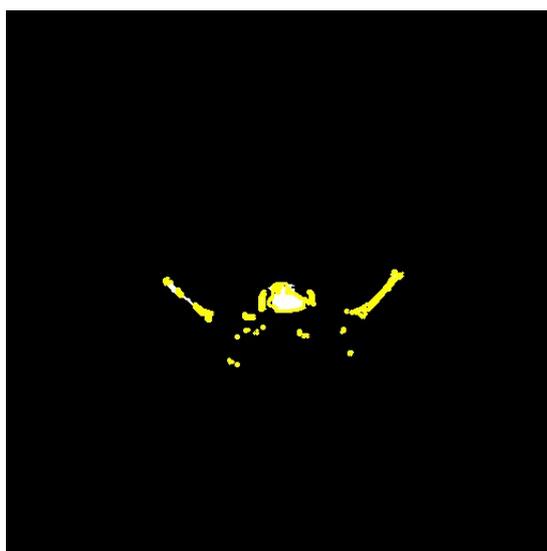


Figura 3.99: Slice n36 del paziente 5 relativa alla prima suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Il Validation Set RM del paziente 5 non ha valori performanti, infatti tutte le slice hanno DSC inferiore al 72%. La slice n9 (Figura 3.100) ha ottenuto un DSC pari a 20.7% poiché la rete non è riuscita a predire tutta la parte centrale.



Figura 3.100: Slice n9 del paziente 5 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.101: Slice n42 del paziente 5 relativa alla prima suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Riassumendo, questa prima suddivisione ha evidenziato che la rete riesce a lavorare molto meglio con l'individuazione di zone RM rispetto alle zone YM. Questo perché spesso le slice relative alle zone di midollo osseo giallo sono strette, irregolari e sparse e quindi di più difficile predizione. Tuttavia, il riconoscimento di zone RM non è esente da criticità. La rete ha alcune difficoltà con le zone anatomiche relative alla parte centrale della colonna vertebrale presente nelle primissime slice. In tutti i pazienti, la predizione di zone RM nelle sezioni centrali del bacino e nelle creste iliache ha mostrato performance promettenti nel quale si è riuscito a raggiungere anche il 98% nel DSC e Re.

3.3.2 Secondo Data Set

Questa seconda suddivisione consiste in 24 slice per i pazienti 1, 3 e 5; il paziente 4 presenta 23 slice, mentre il paziente 2 presenta solo 22 slice e quindi è minoranza (poiché si hanno meno slice di questo paziente). La seconda suddivisione comprende slice relative al pube, la parte bassa del bacino e l'iniziale comparsa dei femori.

La seconda suddivisione di Training Set con cui si è allenata la rete per individuare zone di RM evidenzia una coerenza in tutti e 5 i gruppi di box plot (Figura 3.102). Quindi, è possibile fare un discorso generale; sia DSC che Pr hanno una larga variabilità di valori, mentre la Re si colloca in alto.

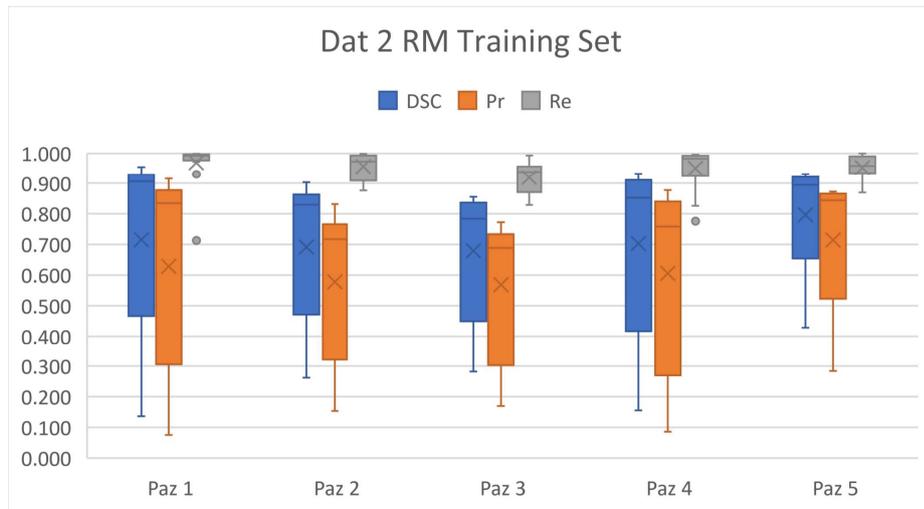
Più nello specifico, il DSC ha una distribuzione di valori che dal primo quartile fino al terzo quartile va da circa il 50% fino al 90%. Ci sono solo due eccezioni: il paziente 3 ha come valore massimo di DSC 85% e il paziente 5 ha un 25esimo percentile intorno al 65%. Anche la Pr segue lo stesso andamento ed infatti si ha che il paziente 3 ha valore massimo intorno al 75%, mentre gli altri superano il 80%; inoltre, il paziente 5 presenta primo quartile più alto rispetto a tutti e sopra il 50%, mentre in media si assesta intorno al 30%. La recall è molto alta e superiore al 90%; tuttavia, questo non si verifica per il terzo paziente poiché, seppur alta, ha una distribuzione di valori che scendono fino al 85%.

La seconda suddivisione di Validation Set mostra, in maniera netta, come il paziente 3 si comporti diversamente rispetto agli altri. Infatti, prendendo in considerazione i box plot dei pazienti 1, 2, 4 e 5 si hanno degli andamenti abbastanza simili anche se non quanto il Training Set visto precedentemente. Il paziente 1 ha un valore massimo di DSC intorno al 95% e si discosta lievemente dal terzo quartile. Come per i paziente 2 e 4, il valore del 25esimo percentile del DSC è intorno al 50%. Anche la Pr ha una distribuzione di valori molto ampia ed il terzo quartile si discosta di poco dal valore massimo che è 90%; il primo quartile è intorno al 45% ed il valore minimo appena sopra il 10%. La Re, invece è molto alta e superiore già dal suo primo quartile al 90%, anche se si nota un outlier con valore al 70%

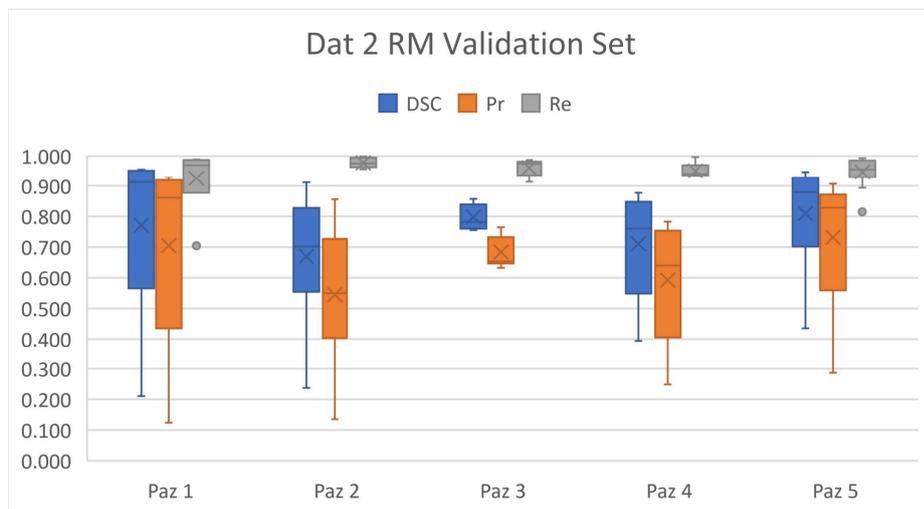
I paziente 2 e 4 sono quelli più simili tra loro. Infatti, ambedue hanno ottenuto per il DSC un 75esimo percentile che si assesta a circa 85% e 25esimo percentile intorno al 55%. Le uniche differenze si notano nei valori di massimo e minimo; il paziente 2 ha valore massimo di DSC al 90%, mentre il paziente 4 arriva a circa 87%. Il valore minimo di DSC nel paziente 2 scende fino al 15%, mentre per il paziente 4 si ferma al 25%. La Re, invece, supera il 95% in ambedue i pazienti e già a partire dal 25esimo quartile.

Il quinto paziente ha un andamento simile ai precedenti, anche se presenta valori lievemente superiori. Infatti, il primo quartile del DSC è al 70% e ha valore massimo intorno al 95%. Questo è indice che la distribuzione di valori è più spostata verso i valori alti rispetto agli altri pazienti. Ciò avviene anche per Pr, in cui si ha primo quartile intorno al 55%, quando tutti gli altri pazienti scendevano sotto il 50%. La Re, invece, è alta come in tutti i pazienti e superiore al 90%. Tuttavia, è presente un outlier con valore 80%.

Il terzo paziente, rispetto alle performance viste in precedenza nel Capitolo 3.2, non ha un'estesa variabilità di valori, non ha outlier e sono tutti abbastanza alti. Infatti, i tre parametri hanno valori superiori al 60%. Dettagliatamente, il DSC ha valore massimo intorno al 85% e primo quartile intorno al valore di 75%; la precisione ha una distribuzione di valori che vanno da un primo quartile al 65% ed un terzo quartile al 75%. Infine, la recall ha valori alti superiori al 90% come il resto dei pazienti.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.102

La seconda suddivisione di Training Set usata per allenare la rete al riconoscimento di zone YM evidenzia delle performance con variabilità nei valori, indice che non si ha una buona coerenza nelle predizioni ed in generale si hanno molte più criticità rispetto al caso con RM. Il primo paziente è quello che presenta maggiore variabilità e ha una distribuzione che va dal 25esimo percentile pari al 45% fino al terzo percentile con valore al 95%; si osserva che il valore massimo si discosta di poco, mentre il valore minimo è molto diverso rispetto al primo quartile poiché è pari al 15% circa. La Re ha una distribuzione di valori molto più ampia; il suo terzo quartile supera il 90% fino a raggiungere un primo quartile poco sotto il 20%. Infine, la Pr è quella che ha minore distribuzione poiché il 25esimo percentile e 75esimo percentile valgono, rispettivamente, 85% e circa 98%; indice che la rete riesce a predire bene la classe YM poiché i FP sono bassi.

Il secondo paziente ha una minore variabilità rispetto al primo paziente, ma in generale ha performance che per quanto riguarda il DSC si assestano ad un valore medio di circa 80%. Soltanto la Pr non ha una grande variabilità e ha valori tutti superiori al 90%.

Il terzo paziente ha valore di DSC che nel 75esimo percentile raggiunge il 90%. Tuttavia, il 25esimo percentile si assesta intorno al 65%. La Re segue l'andamento degli altri pazienti (escludendo il paziente 1 che ha una variabilità maggiore) e ha un valore di 25esimo percentile intorno al 55%.

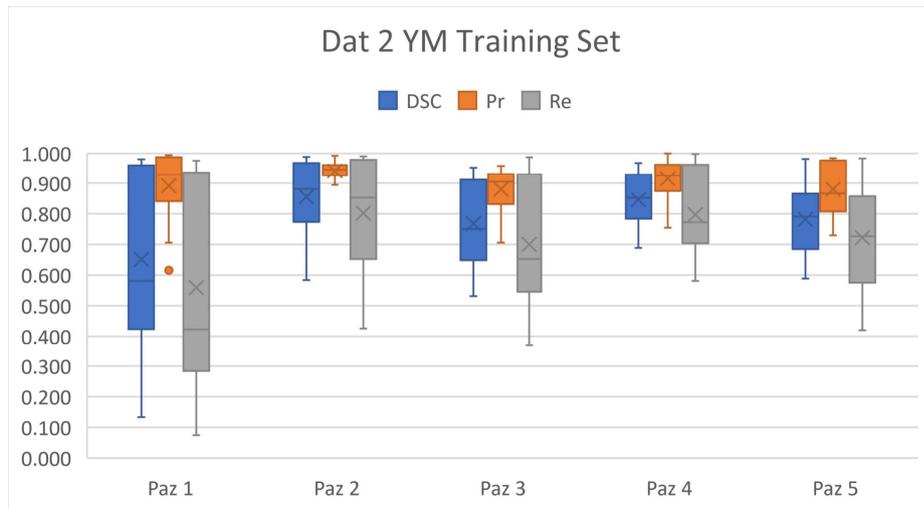
Il quarto paziente presenta delle buone performance, infatti il valore di DSC nel 75esimo percentile è superiore al 90%, il valore minimo è pari al 70% e non sono presenti outlier. La Re mostra una minore variabilità rispetto ai pazienti descritti precedentemente e ha come valore minimo 60% e ha come valore massimo 100%, questa particolare slice verrà in seguito discussa. Il quinto paziente non mostra alcuna evidente variazione rispetto ai pazienti 2, 3 e 4; infatti, la variabilità è nettamente inferiore rispetto al paziente 1, la DSC si assesta con un 75esimo percentile a 85%, viene evidenziato un massimo al 98% ed un minimo pari a 60%. La recall è poco inferiore rispetto ai precedenti pazienti e ha come 75esimo percentile 85% anche se il massimo è circa 98%. La precisione, come nei casi precedenti è sempre molto alta e con valore minimo di 75%.

Nella seconda suddivisione del Validation Set si evidenziano maggiormente le non soddisfacenti performance da parte della rete nel riconoscimento di zone YM. Infatti, seppur il valore massimo riesce anche a superare il 90%; nei pazienti 1, 2,4 e 5 il terzo quartile non riesce a superare il 80%. Più nel dettaglio, il primo paziente presenta una grande variabilità, soprattutto nella recall. Infatti il 25esimo percentile si assesta sotto il 20%, mentre il 75esimo percentile è circa 85%. Il DSC come già detto è sotto il 80% nella maggioranza dei casi e arriva appena al 25% nel primo quartile. Il quinto paziente ha performance simili con la differenza che la variabilità è minore poiché il 25esimo percentile per i tre parametri si assesta intorno al 60%.

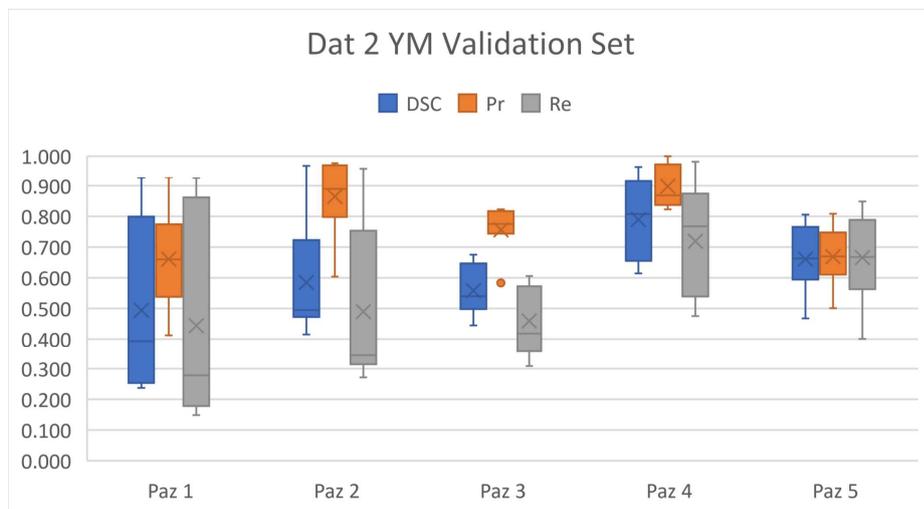
Il secondo paziente presenta meno variabilità rispetto al primo paziente, però le performance restano comunque inferiori. Da notare che la precisione si alza, infatti il valore minimo è 60%; tuttavia, il terzo quartile è intorno al 95%, indice che la rete ha un numero di FP inferiore ai TP.

Il terzo paziente presenta un comportamento non coerente tra DSC, Pr e Re. Più nel dettaglio, quest'ultima ha valore massimo 60%, mentre il DSC non supera il 65%. La precisione presenta una variabilità bassa rispetto agli altri pazienti.

Il quarto paziente evidenzia delle buone performance rispetto agli altri, infatti la DSC ha un terzo quartile più in alto e pari al 90%, mentre per la recall si ha il 70%.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.103

Risultati Paziente 1 - RM

La porzione del paziente 1 relativa al Training Set RM della seconda suddivisione mostra delle criticità soltanto in un gruppo di slice specifico. Come mostrato nelle slice n150.5 e n153.5 (Figure 3.104 e 3.105) ciò è dovuto al sovra-campionamento di FP dove sono presenti le teste femorali.

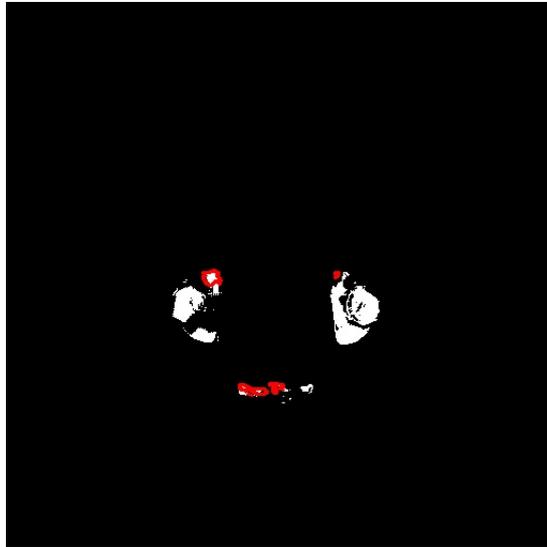


Figura 3.104: Slice n150.5 del paziente 1 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

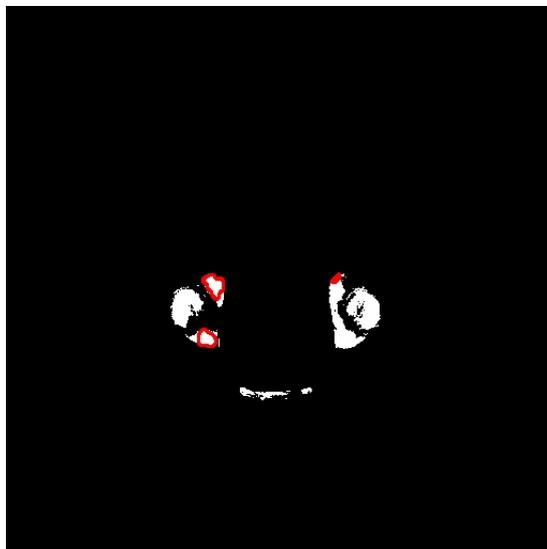


Figura 3.105: Slice n153.5 del paziente 1 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Nel momento in cui i femori non ci sono e si è nella parte terminale della spina dorsale, dove si ha il coccige, la predizione diventa quasi eccellente con valori di DSC superiori al 90%. Ciò accade nelle slice n105.5 e n99.5 (Figure 3.106 e 3.107) che hanno ambedue DSC pari al 95.6%.

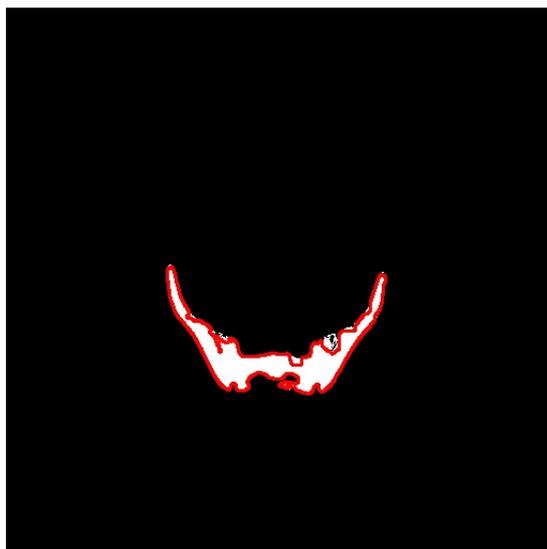


Figura 3.106: Slice n105.5 del paziente 1 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

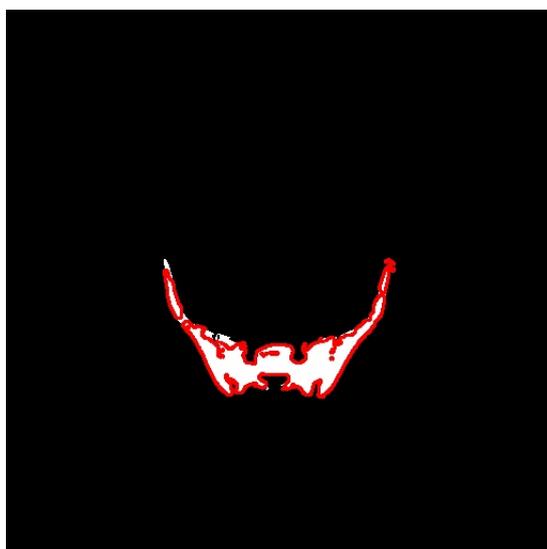


Figura 3.107: Slice n99.5 del paziente 1 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Tale comportamento della rete si riflette sul Validation Set RM che è coerente con le performance ottenute nel Training Set. Infatti la rete riesce a predire le zone di RM, ma si hanno problemi di sovra-campionamento quando le teste femorali diventano predominanti nelle slice. Quando, invece, le slice riguardano la parte finale della colonna vertebrale, la predizione è ottimale.

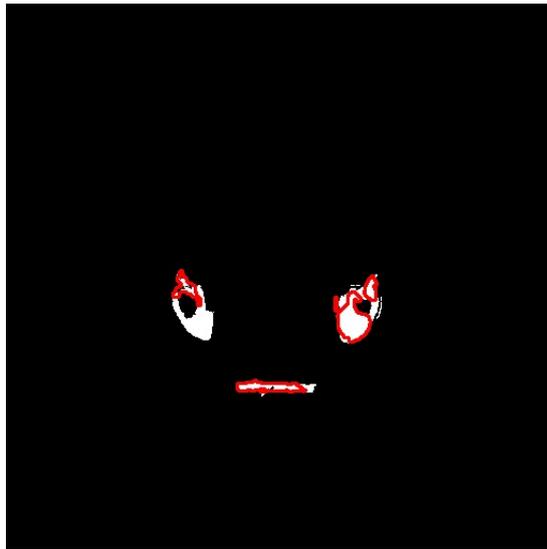


Figura 3.108: Slice n138.5 del paziente 1 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

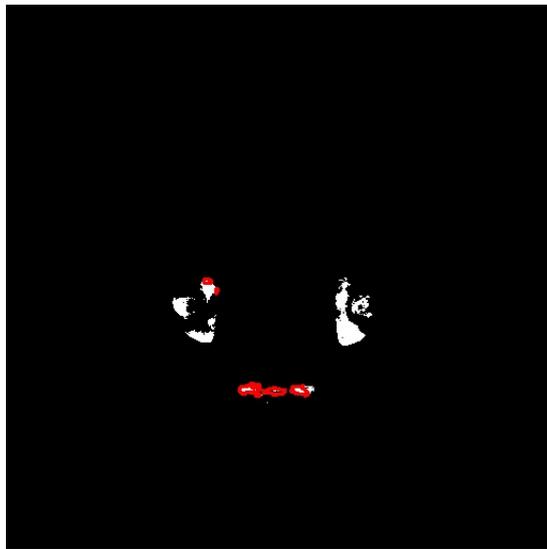


Figura 3.109: Slice n144.5 del paziente 1 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

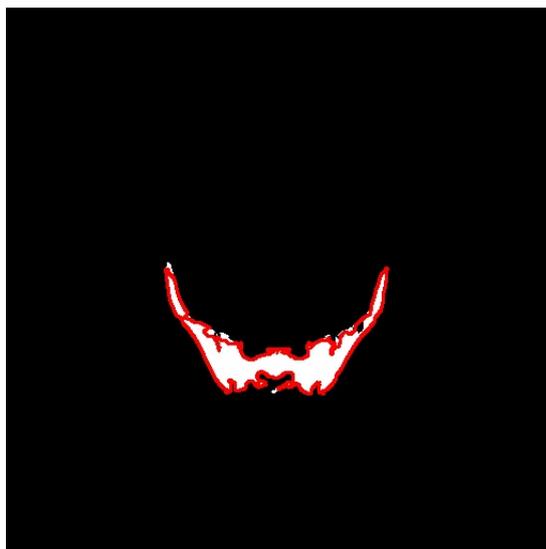


Figura 3.110: Slice n102.5 del paziente 1 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

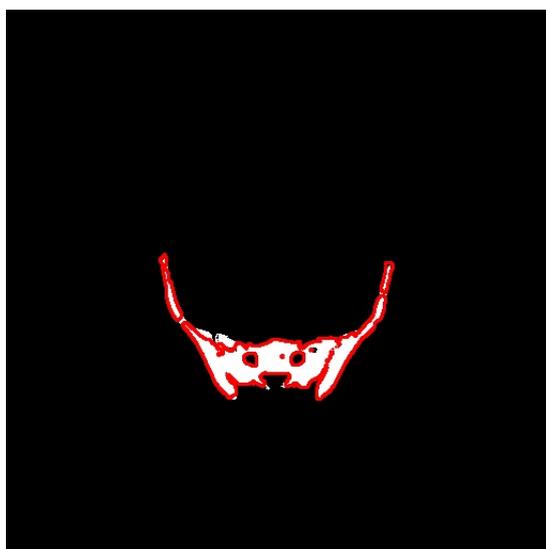


Figura 3.111: Slice n96.5 del paziente 1 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 1 - YM

Le porzioni del primo paziente relative al Training Set YM e Validation YM hanno comportamento completamente opposto ai casi visti in precedenza. Infatti, nel momento in cui si deve predire la parte di midollo osseo giallo nel coccige, si hanno risultati non performanti, invece, quando si devono predire le zone di YM presente nelle teste femorali si hanno risultati eccellenti, come nel caso delle slice nelle Figure 3.114 e 3.115 per il Training Set YM e le slice n102.5 e n126.5 nelle Figure 3.116 e 3.117 per il Validation Set YM, in queste slice si raggiunge un DSC mediamente pari al 98%.



Figura 3.112: Slice n129.5 del paziente 1 relativa alla seconda suddivisione del Training Set, i bordi gialli indicano la MaskYM effettiva.

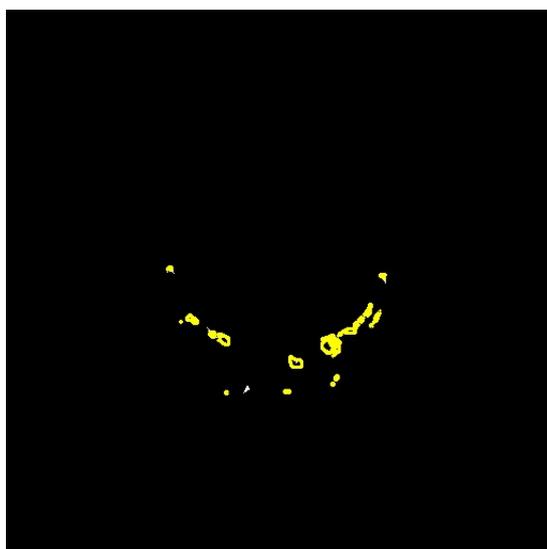


Figura 3.113: Slice n105.5 del paziente 1 relativa alla seconda suddivisione del Training Set, i bordi gialli indicano la MaskYM effettiva.

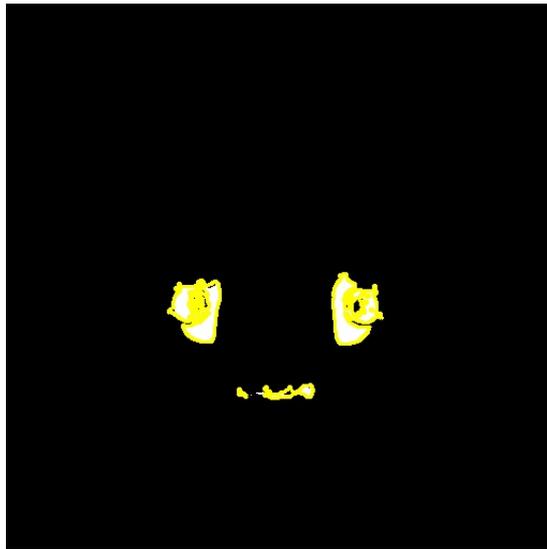


Figura 3.114: Slice n147.5 del paziente 1 relativa alla seconda suddivisione del Training Set, i bordi gialli indicano la MaskYM effettiva.

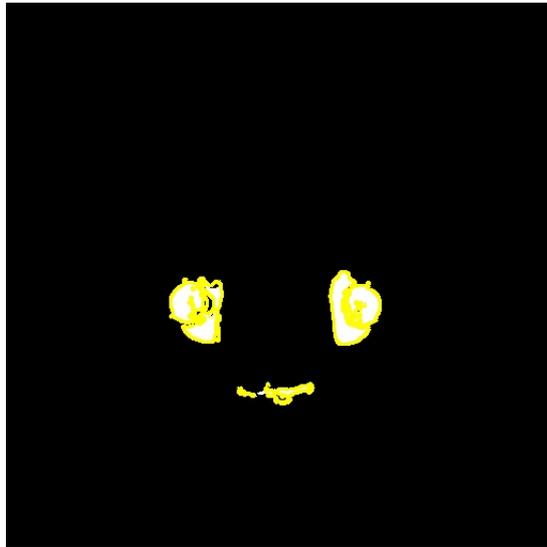


Figura 3.115: Slice n150.5 del paziente 1 relativa alla seconda suddivisione del Training Set, i bordi gialli indicano la MaskYM effettiva.



Figura 3.116: Slice n102.5 del paziente 1 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.117: Slice n126.5 del paziente 1 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskRM effettiva.



Figura 3.118: Slice n144.5 del paziente 1 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskRM effettiva.

Risultati Paziente 2 - RM

Anche il paziente 2 ha comportamento uguale al paziente 1 sia nel Training Set che Validation Set. Anche se in questo caso le performance sono lievemente inferiori, poiché mediamente si ha un DSC intorno al 80%. Infatti, come si vede nelle Figure 3.121 e 3.122, in cui si ha il coccige, si ha un fenomeno di sovra-campionamento che non era presente col paziente 1, ciò comporta un alto tasso di FP e conseguentemente diminuiscono Pr e DSC. Analogamente, nelle slice n171.5 e n177.5 (Figure 3.129 e 3.120) le zone delle teste femorali sono state predette come zone RM quando in realtà non lo sono.

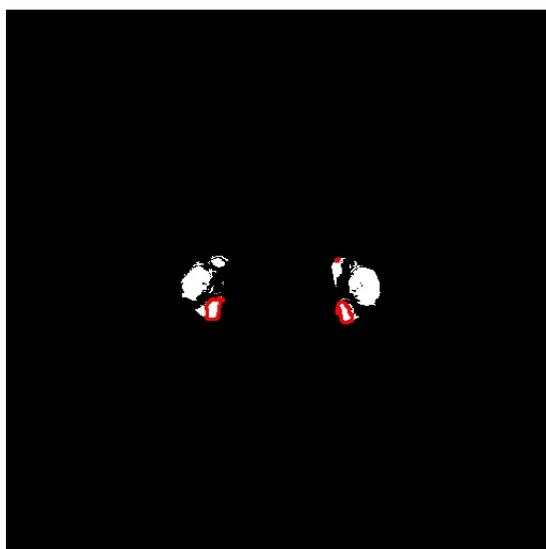


Figura 3.119: Slice n171 del paziente 2 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

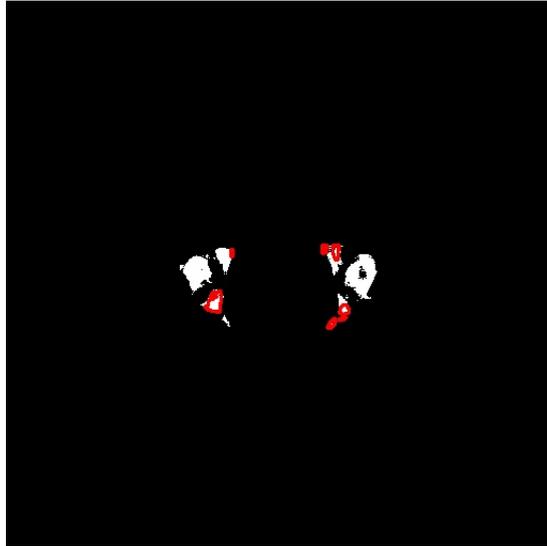


Figura 3.120: Slice n177 del paziente 2 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

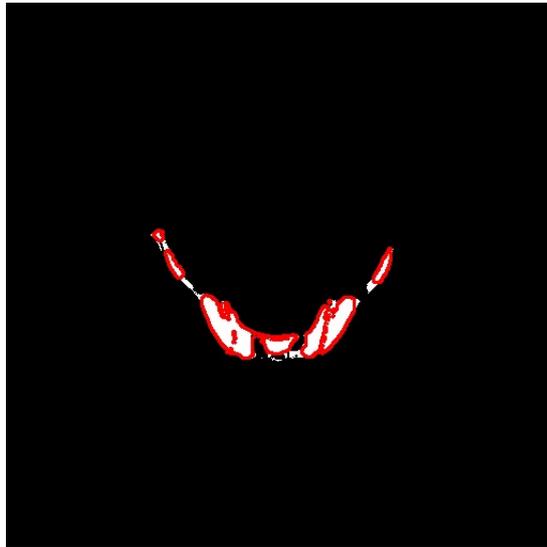


Figura 3.121: Slice n117 del paziente 2 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

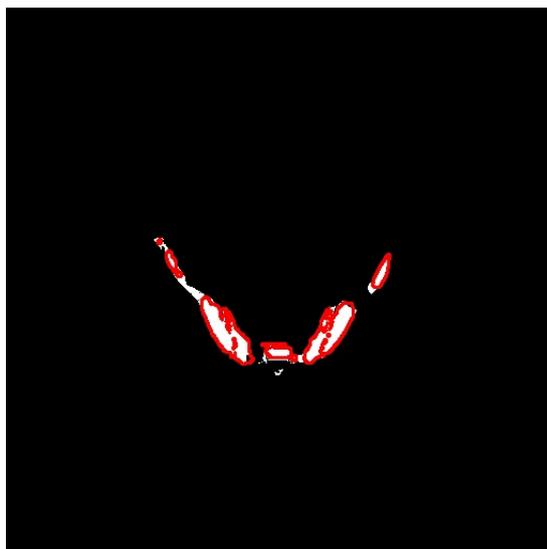


Figura 3.122: Slice n120 del paziente 2 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Nel Validation Set si osserva meglio questo problema nel paziente 2, infatti come si vede nella slice n138 3.123, si ha sovra-campionamento di midollo osseo rosso pure nella sezione anatomica finale della spina dorsale che nel paziente 1 non dava alcun problema.

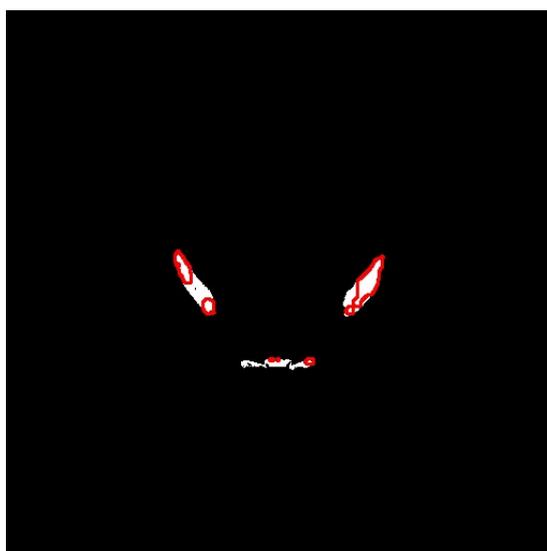


Figura 3.123: Slice n138 del paziente 2 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

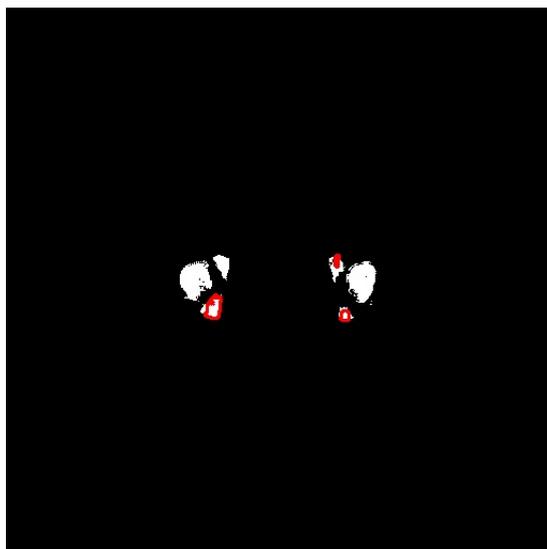


Figura 3.124: Slice n174 del paziente 2 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

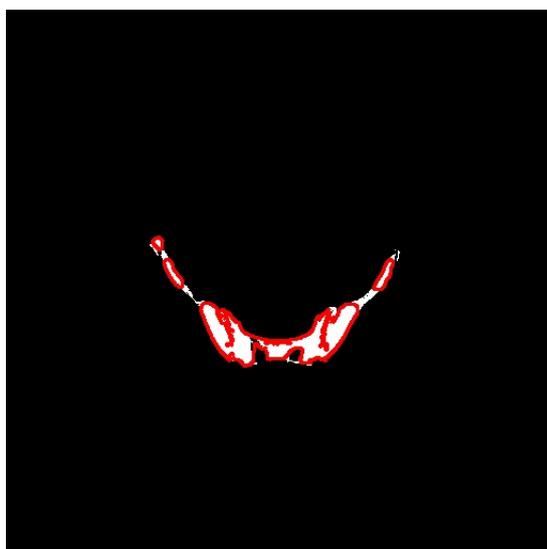


Figura 3.125: Slice n114 del paziente 2 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 2 - YM

Le porzioni del secondo paziente relative al Training Set YM e Validation YM hanno comportamento completamente opposto ai casi visti in precedenza. Infatti, nel momento in cui si deve predire la parte di midollo osseo giallo nel coccige, si hanno risultati non ottimali, invece, quando si devono predire le zone di YM presente nelle teste femorali si hanno risultati eccellenti, come nel caso delle slice nelle Figure 3.128 e 3.129 per il Training Set YM e la slice n174 nella Figura 3.132 per il Validation Set YM, in queste slice si raggiunge un DSC mediamente pari al 97%.



Figura 3.126: Slice n117 del paziente 2 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.127: Slice n147 del paziente 2 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

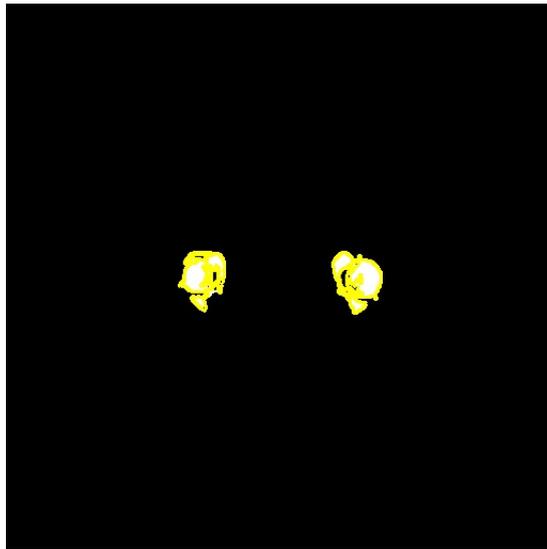


Figura 3.128: Slice n168 del paziente 2 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

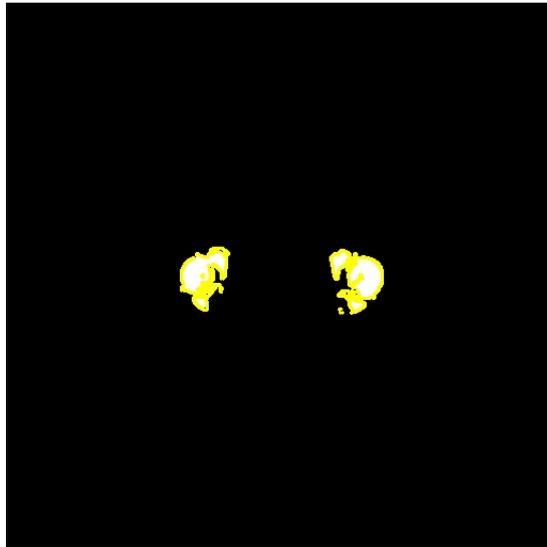


Figura 3.129: Slice n171 del paziente 2 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

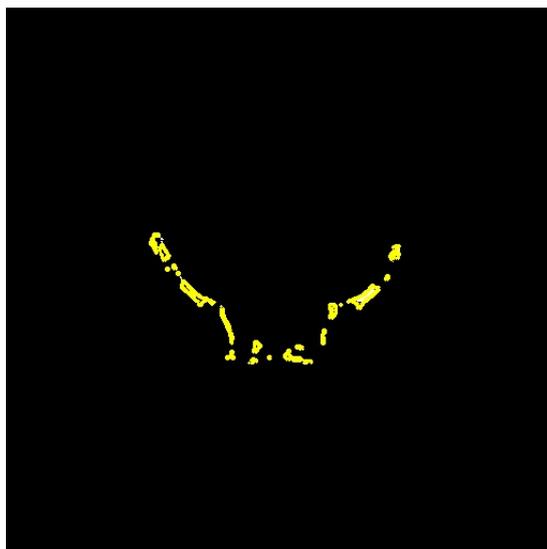


Figura 3.130: Slice n114 del paziente 2 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

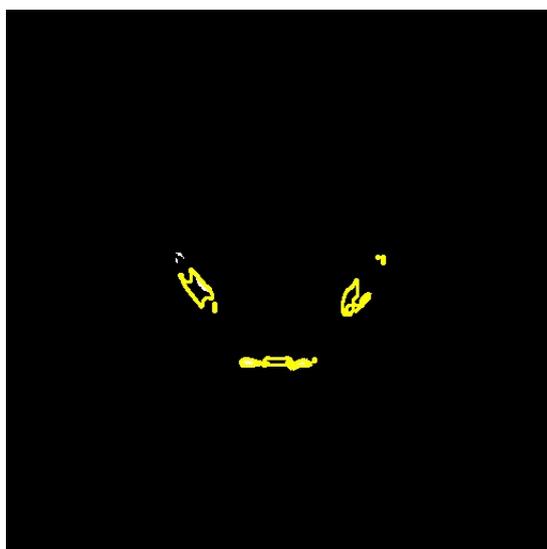


Figura 3.131: Slice n138 del paziente 2 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

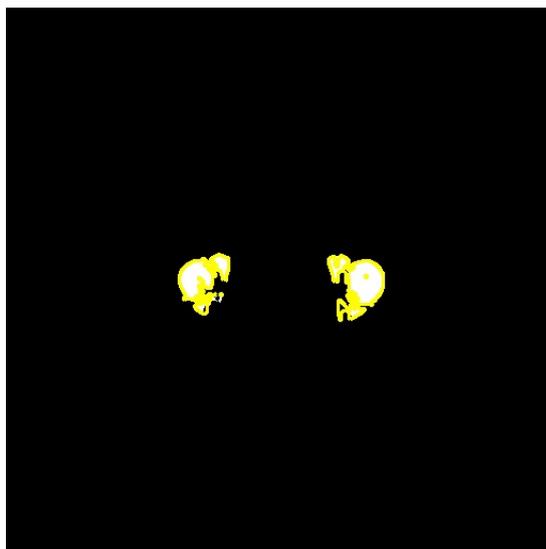


Figura 3.132: Slice n174 del paziente 2 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 3 - RM

Anche il paziente 3 ha comportamento uguale al paziente 2 sia nel Training Set che Validation Set. Infatti, come si vede nella slice n163 in Figura 3.135, in cui si ha il coccige, è presente un fenomeno di sovra-campionamento che non era presente col paziente 1, ciò comporta un alto tasso di FP e conseguentemente diminuiscono Pr e DSC, in particolare per questa slice, si hanno valori rispettivamente 85.7% e 77.3%. Analogamente nelle slice n211 e n217 (Figure 3.133 e 3.134) le zone delle teste femorali sono state predette come zone RM quando in realtà non lo sono.

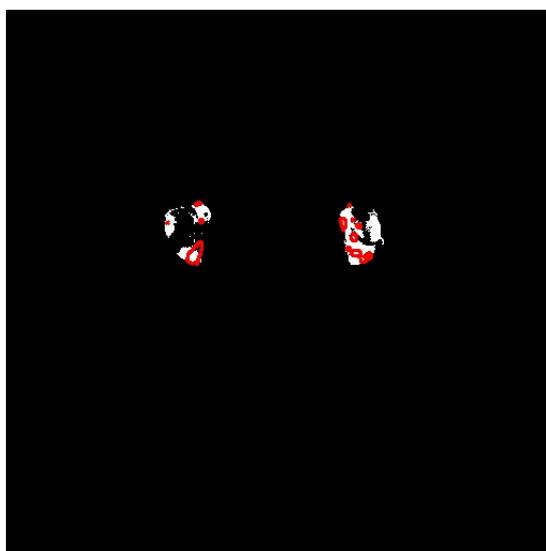


Figura 3.133: Slice n211 del paziente 3 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

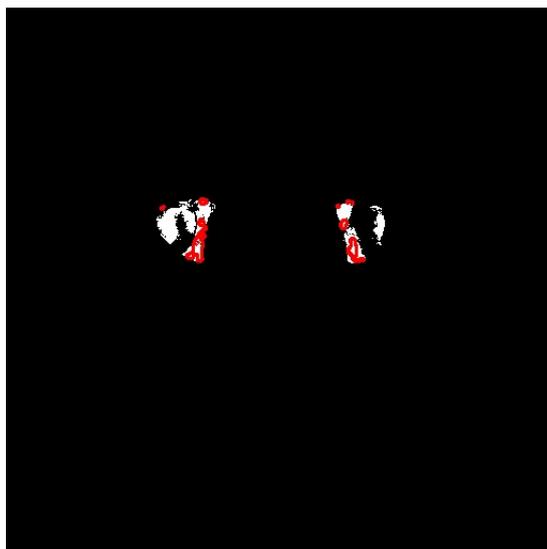


Figura 3.134: Slice n217 del paziente 3 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

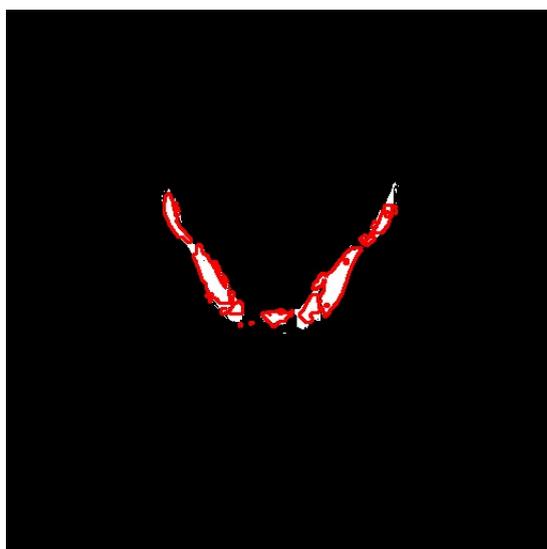


Figura 3.135: Slice n163 del paziente 3 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Le performance nel Validation Set non sono così basse come il Training Set perché il paziente 3 non presenta slice con le teste femorali, ma soltanto slice relative al coccige. Infatti, mediamente si ha un valore di DSC pari al 80%.

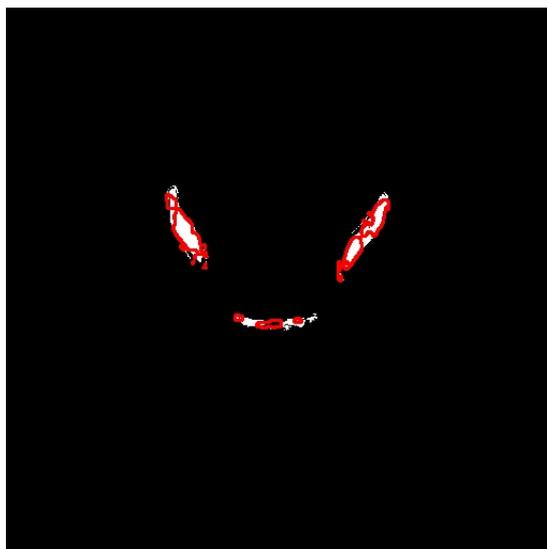


Figura 3.136: Slice n178 del paziente 3 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

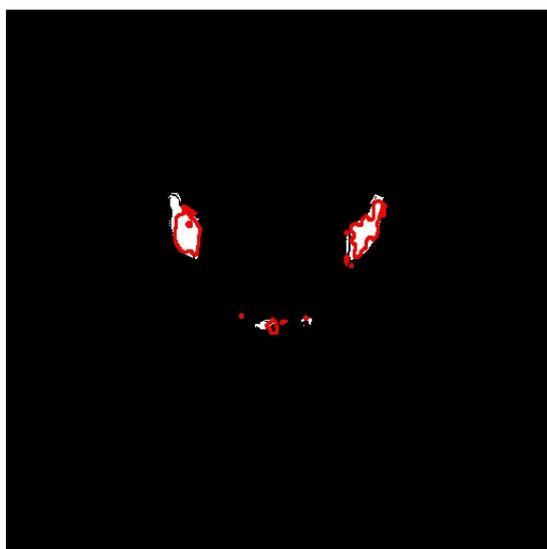


Figura 3.137: Slice n190 del paziente 3 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

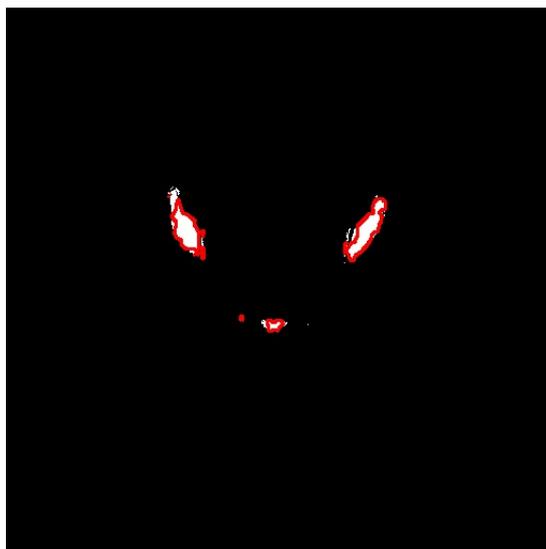


Figura 3.138: Slice n184 del paziente 3 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 3 - YM

La porzione del terzo paziente relative al Training Set YM hanno comportamento completamente opposto ai casi RM visti in precedenza e ciò è concorde con i precedenti pazienti. Infatti, nel momento in cui si deve predire la parte di midollo osseo giallo nel coccige, si hanno risultati non ottimali, invece, quando si devono predire le zone di YM presente nelle teste femorali si hanno risultati eccellenti, come nel caso delle slice nelle Figure 3.141 e 3.142 per il Training Set YM dove si raggiunge un DSC mediamente pari al 95%.

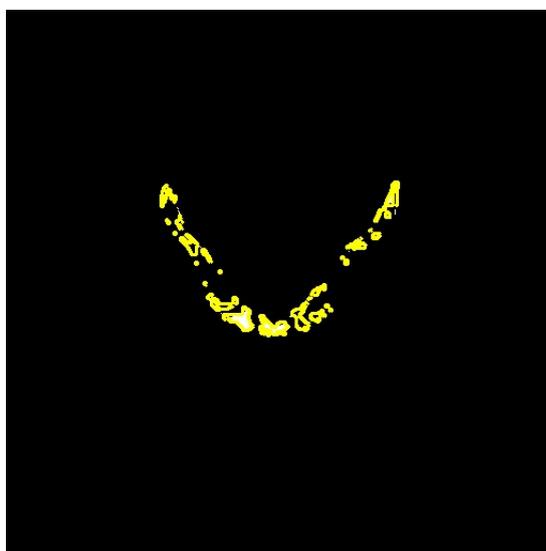


Figura 3.139: Slice n163 del paziente 3 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.140: Slice n199 del paziente 3 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.141: Slice n211 del paziente 3 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.142: Slice n217 del paziente 3 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

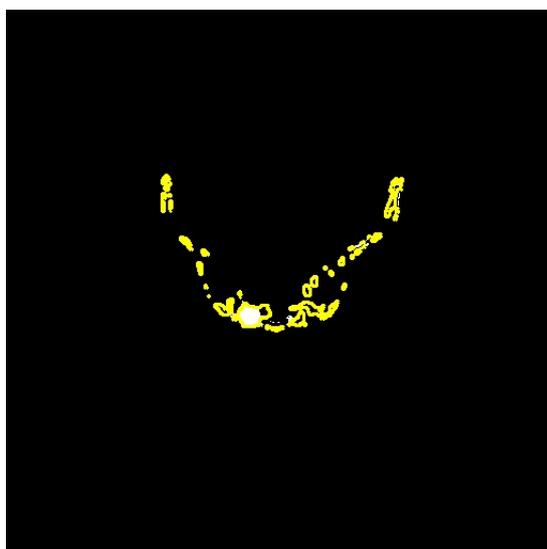


Figura 3.143: Slice n157 del paziente 3 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

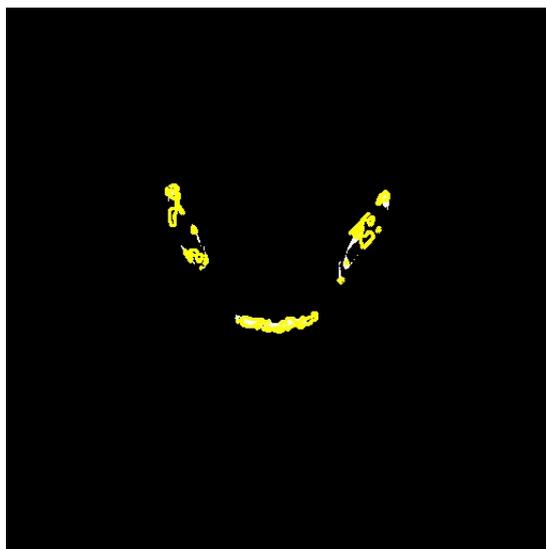


Figura 3.144: Slice n178 del paziente 3 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 4 - RM

Il paziente 4 ha comportamento coerente coi precedenti pazienti sia nel Training Set che Validation Set. Si vede che nelle slice n80.5 e n83.5 nelle Figure 3.147 e 3.148 in cui si ha il coccige, la predizione è molto promettente e si ottengono valori di DSC pari al 91% in ambedue le immagini. Analogamente nelle slice n128.5 e n131.5(Figure 3.145 e 3.146) le zone delle teste femorali sono state predette come zone RM quando in realtà non lo sono con conseguente aumento del tasso di FP e diminuzione di DSC e Pr.

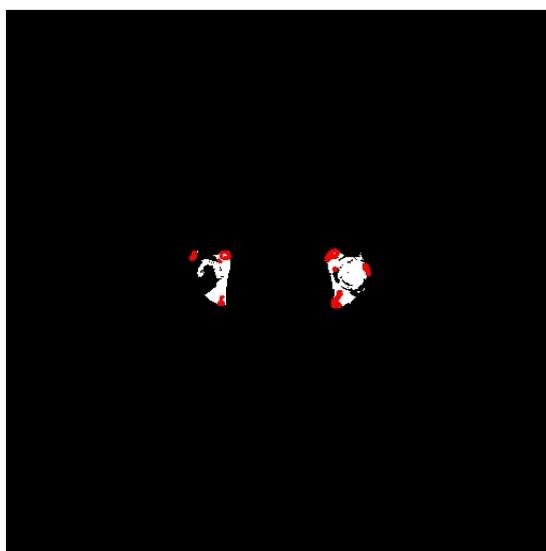


Figura 3.145: Slice n128.5 del paziente 4 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

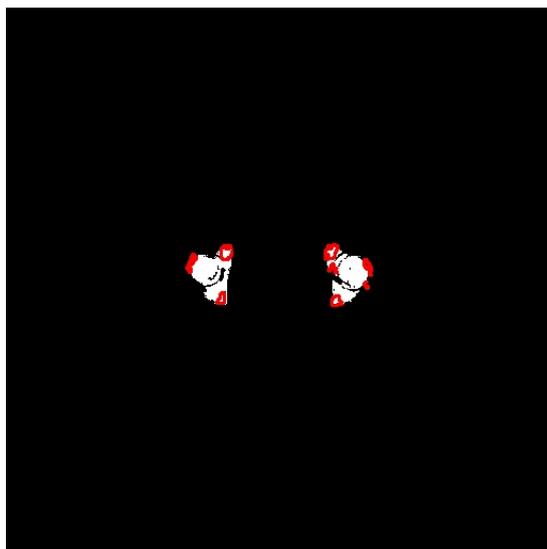


Figura 3.146: Slice n131.5 del paziente 4 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

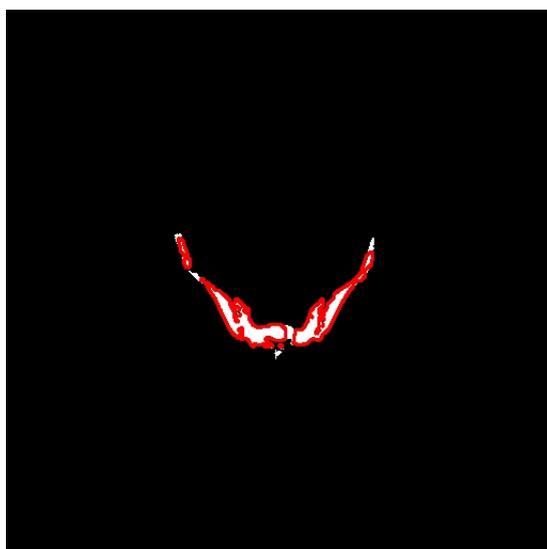


Figura 3.147: Slice n80.5 del paziente 4 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

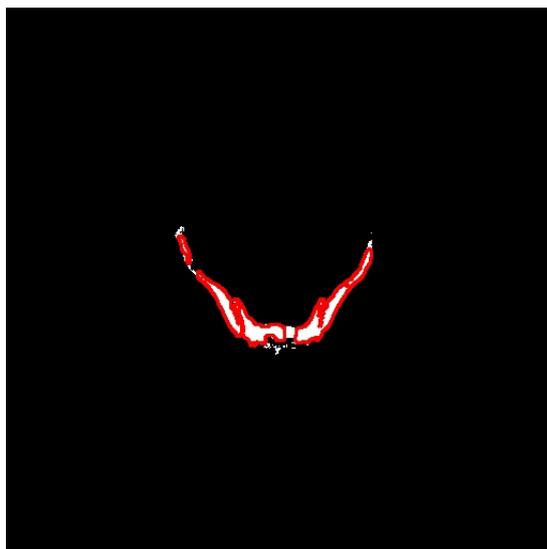


Figura 3.148: Slice n83.5 del paziente 4 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

La porzione di Validation Set relativa al paziente 4 è molto ridotta e conta soltanto 5 elementi. La slice n119.5 (Figura 3.149) era già stata analizzata nel caso di Training Set intero e aveva evidenziato delle criticità e conseguente DSC pari al 69.5%; anche a seguito della suddivisione non riesce a predire in maniera ottimale, ma in questo caso non sovra-campiona come avveniva precedentemente e il DSC aumenta del 0.5%. La slice n137.5 (Figura 3.150) ha lo stesso problema di predizione dovuto ai femori, ciò portano ad avere un DSC pari al 39.4%.

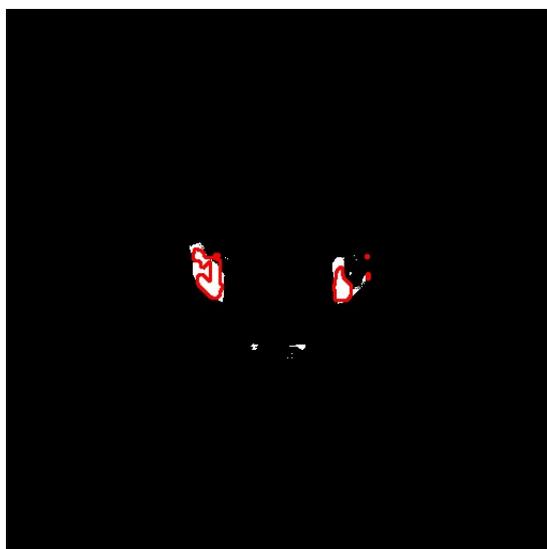


Figura 3.149: Slice n119.5 del paziente 4 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.



Figura 3.150: Slice n137.5 del paziente 4 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

La slice n110.5 (Figura 3.151) è l'unica che riesce ad ottenere un DSC superiore al 80%, più nello specifico arrivando ad 87.7%. Tuttavia, sono presenti fenomeni di sovracampionamento delle zone RM.

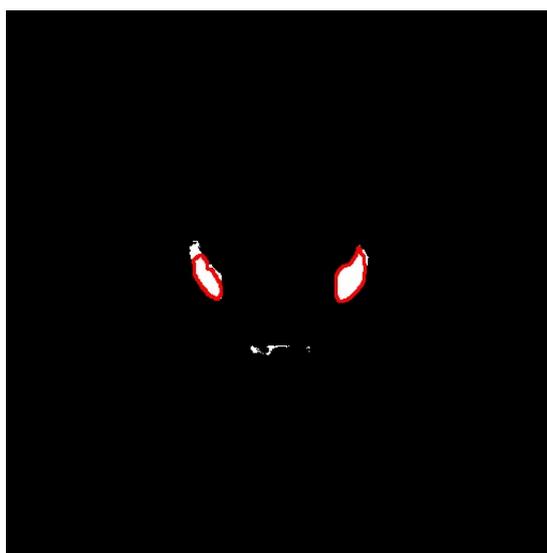


Figura 3.151: Slice n110.5 del paziente 4 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 4 - YM

La porzione del quarto paziente relative al Training Set YM hanno comportamento completamente opposto ai casi RM visti in precedenza e ciò è concorde con i precedenti pazienti. Infatti, nel momento in cui si deve predire la parte di midollo osseo giallo nel coccige, si hanno risultati non ottimali, degli esempi solo le slice n77.5 e n80.5 (Figure 3.152 e 3.153), quest'ultima trattata in precedenza e con performance promettenti mentre ora raggiunge un DSC pari al 69.8%. Invece, quando si devono predire le zone di YM presenti nelle teste femorali si hanno risultati eccellenti. Degli esempi sono le slice n128.5 e n131.5 nelle Figure 3.154 e 3.155 dove si raggiunge un DSC mediamente pari al 95% e una recall prossima al 100%.



Figura 3.152: Slice n77.5 del paziente 4 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.153: Slice n80.5 del paziente 4 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.154: Slice n128.5 del paziente 4 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.155: Slice n131.5 del paziente 4 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Come scritto prima, nel Validation Set sono presenti solo 5 slice del paziente 4. Le slice n92.5 e n95.5 nelle Figure 3.156 e 3.157 evidenziano le problematiche legate all'individuazione di YM nelle zone del coccige e le spine iliache con DSC rispettivamente pari al 61.4% e 69.6%.



Figura 3.156: Slice n92.5 del paziente 4 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

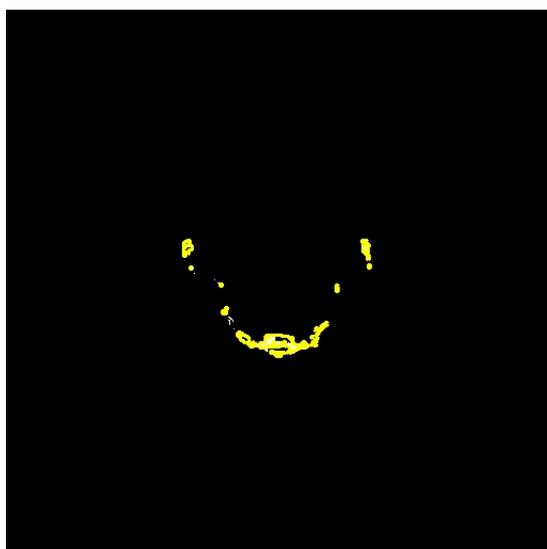


Figura 3.157: Slice n95.5 del paziente 4 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

La slice n137.5 (Figura 3.158), invece, conferma la capacità nel predire le zone di YM nelle teste femorali con un DSC pari al 96.4%. Invece, la slice n110.5 nel quale è stata predetta correttamente l'area YM del coccige ha ottenuto un DSC di 86.8%.

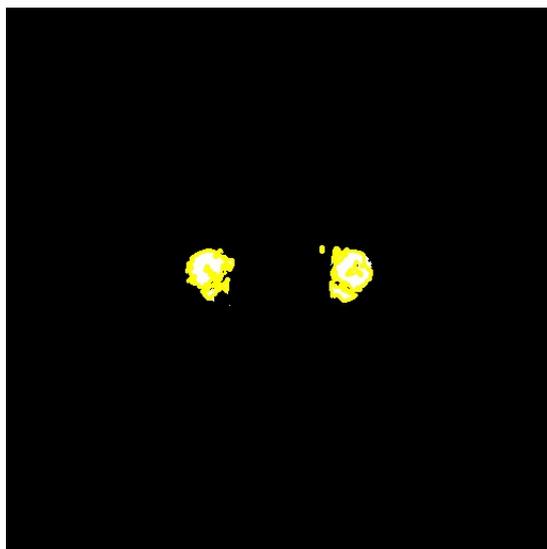


Figura 3.158: Slice n137.5 del paziente 4 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.159: Slice n110.5 del paziente 4 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 5 - RM

Il paziente 5 ha comportamento coerente coi precedenti pazienti sia nel Training Set che Validation Set. Nel primo si vede che nelle slice n117 e n120 nelle Figure 3.162 e 3.163 in cui si ha il coccige, la predizione è molto promettente e si ottengono valori di DSC pari al 91% in ambedue le immagini. Analogamente nelle slice n141 e n147 (Figure 3.160 e 3.161) le zone delle teste femorali sono state predette come zone RM quando in realtà non lo sono con conseguente aumento del tasso di FP e diminuzione di DSC e Pr; più nel dettaglio tali slice hanno valore di DSC appena sotto il 50%.

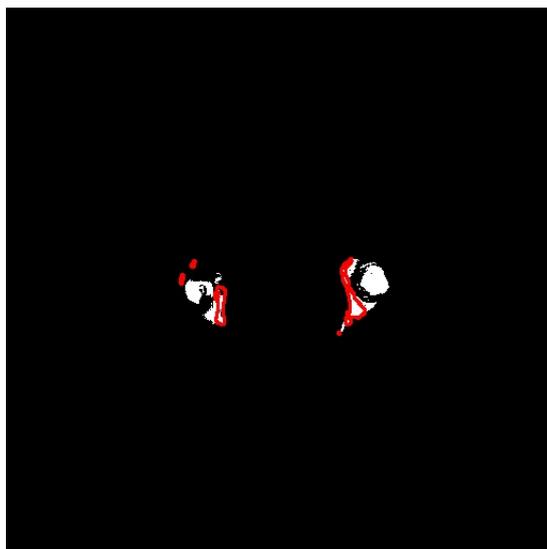


Figura 3.160: Slice n141 del paziente 5 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

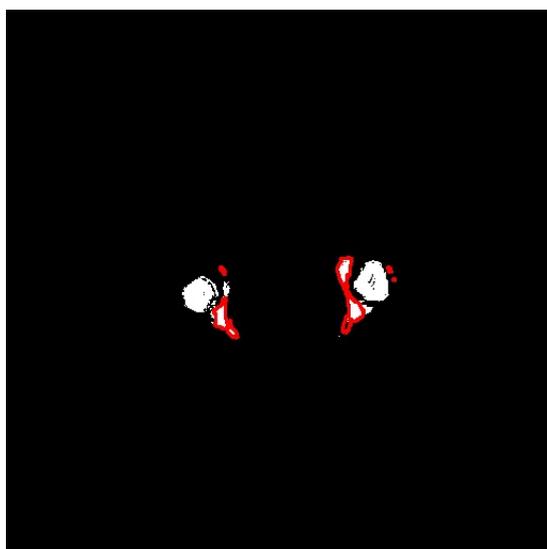


Figura 3.161: Slice n147 del paziente 5 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

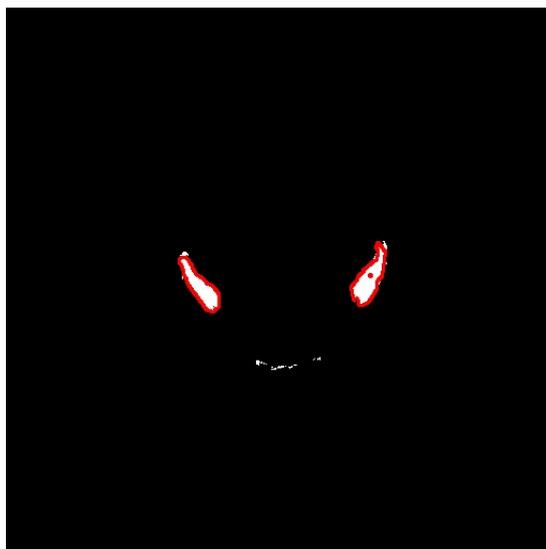


Figura 3.162: Slice n117 del paziente 5 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

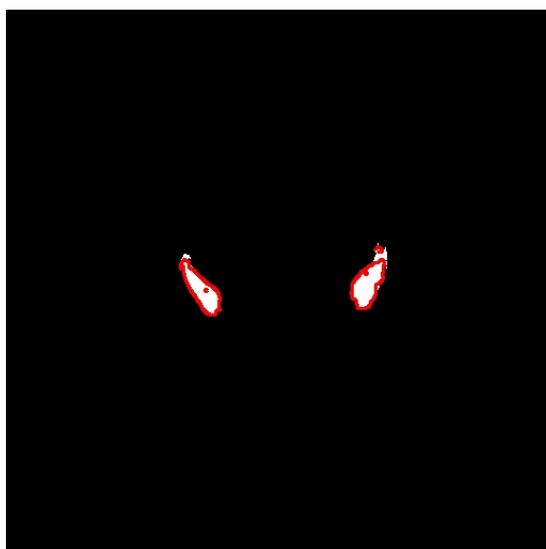


Figura 3.163: Slice n120 del paziente 5 relativa alla seconda suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Il Validation Set del paziente 5 conferma il fenomeno di sovra-campionamento. Infatti, le zone di RM presenti nelle slice n135 e n138 (Figure 3.164 e 3.165) sono predette efficacemente. Tuttavia, ci sono molti FP che portano le performance a diminuire sotto il 70%.

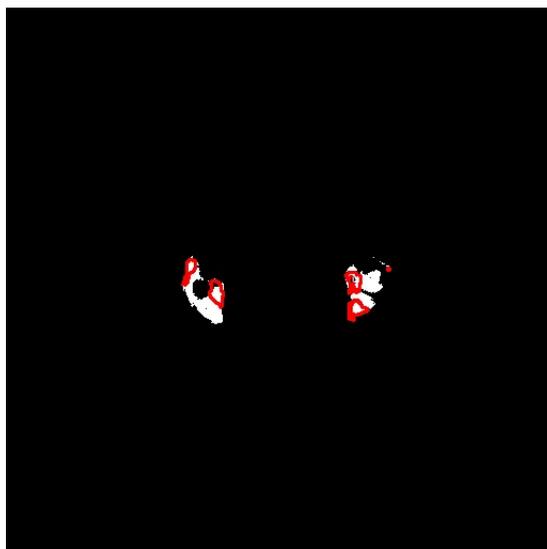


Figura 3.164: Slice n135 del paziente 5 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

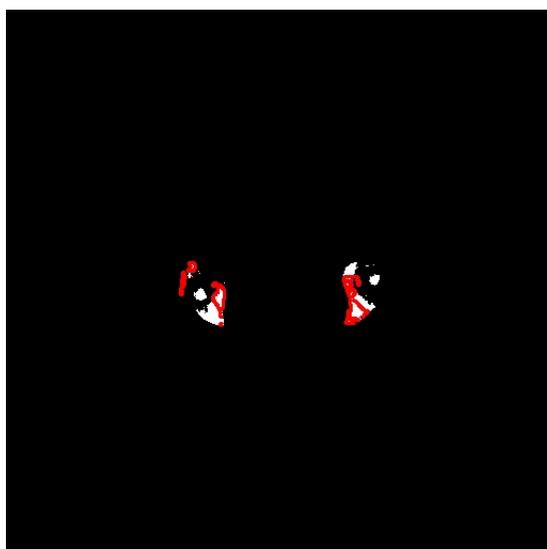


Figura 3.165: Slice n138 del paziente 5 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

Nel caso in cui si deve predire le zone di RM relative alla parte del coccige, si hanno dei lievi fenomeni di sovra-campionamenti. Tuttavia, come visibile nelle slice n111 e n114 (Figure 3.166 e 3.167), le punte iliache sono state individuate correttamente.

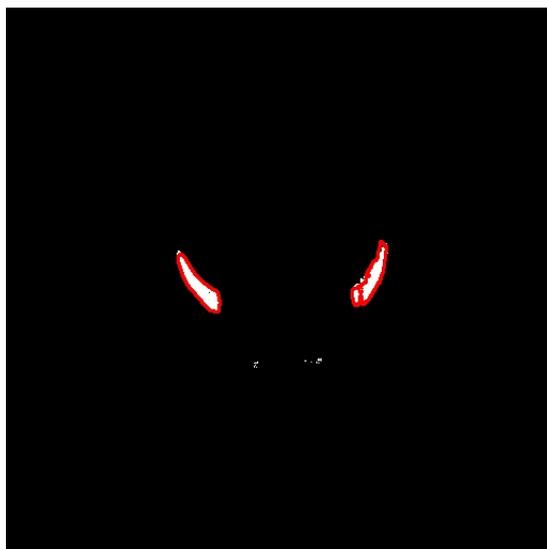


Figura 3.166: Slice n111 del paziente 5 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

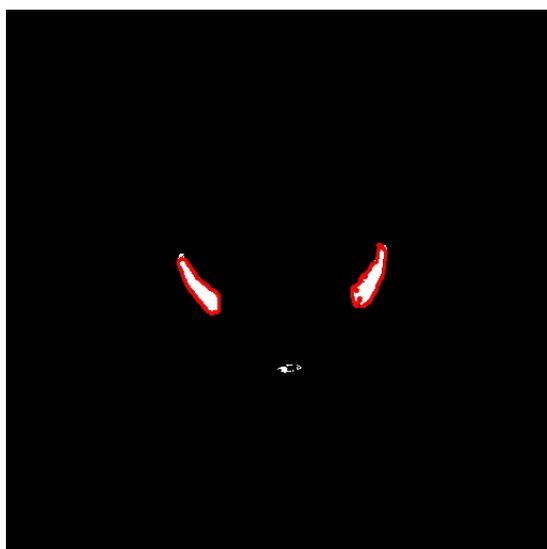


Figura 3.167: Slice n114 del paziente 5 relativa alla seconda suddivisione del Validation RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 5 - YM



Figura 3.168: Slice n123 del paziente 5 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

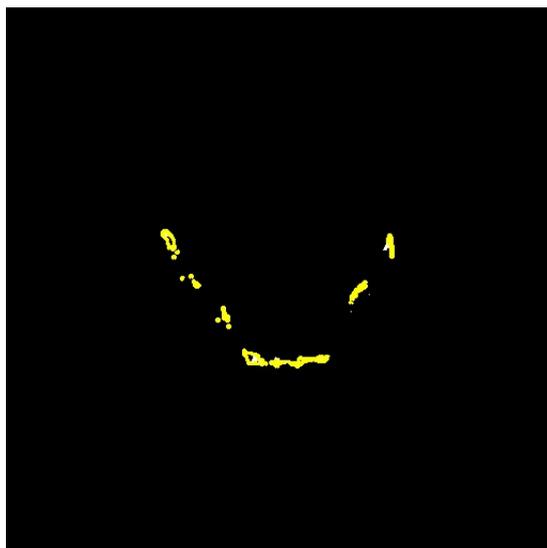


Figura 3.169: Slice n99 del paziente 5 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

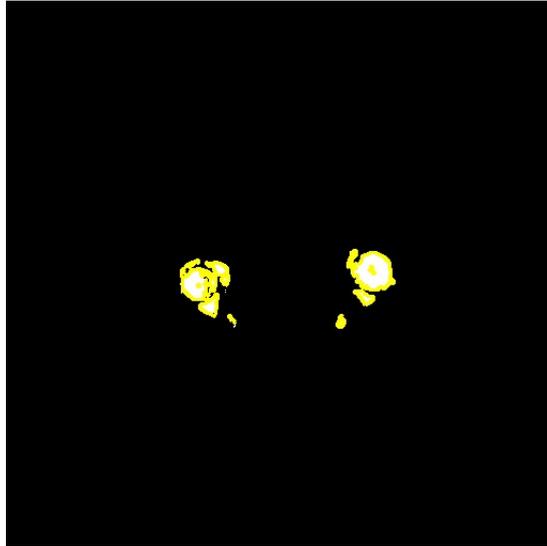


Figura 3.170: Slice n144 del paziente 5 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

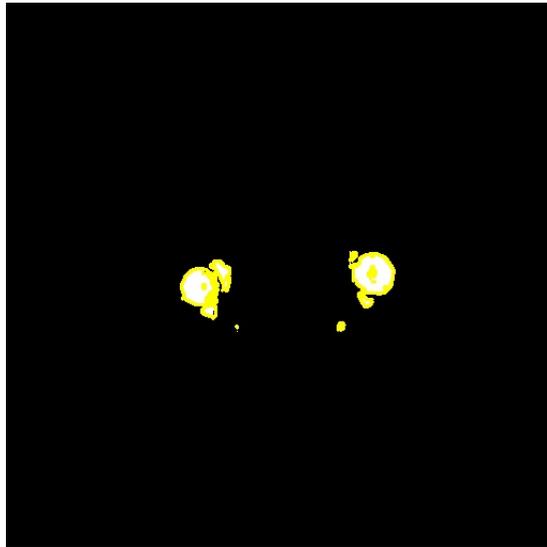


Figura 3.171: Slice n117 del paziente 5 relativa alla seconda suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.172: Slice n126 del paziente 5 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

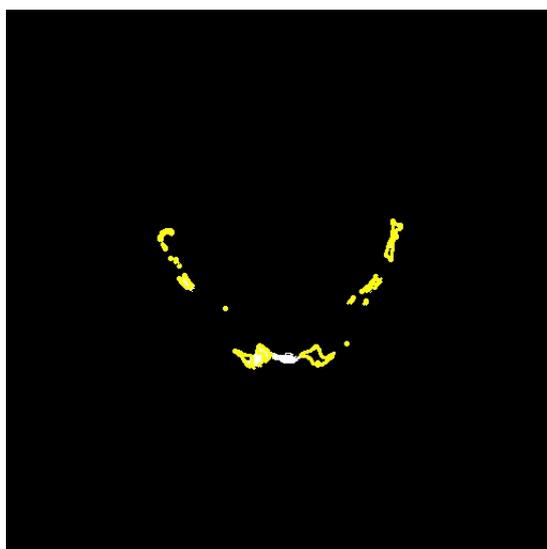


Figura 3.173: Slice n90 del paziente 5 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

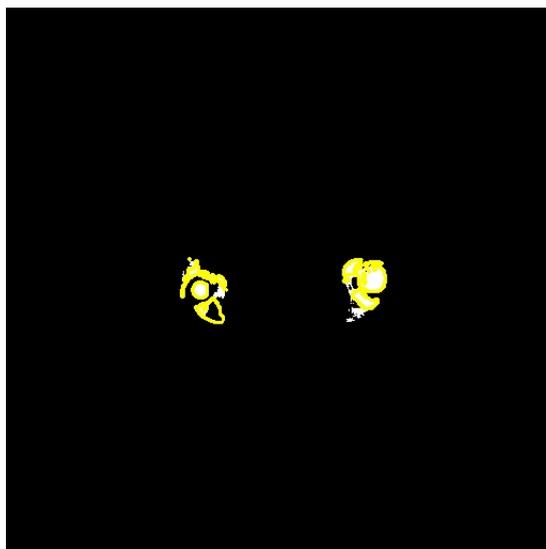


Figura 3.174: Slice n138 del paziente 5 relativa alla seconda suddivisione del Validation YM, i bordi gialli indicano la MaskYM effettiva.

Riassumendo, questa seconda suddivisione ha un comportamento abbastanza coerente tra tutti e 5 i pazienti. Nei vari casi analizzati, si ottengono sempre performance promettenti nel riconoscimento di zone RM presenti nel coccige e nelle spine iliache, raggiungendo valori di DSC molto alti e mediamente pari al 96%. Tuttavia, quando nelle slice originali iniziano ad essere predominanti le teste femorali, ciò porta la rete a sovra-campionare, aumentare il numero di FP e conseguente diminuire la percentuale di DSC e Pr. Quando, invece, si tenta di individuare la zona di YM, i risultati si invertono e le slice in cui è presente il femore sono quelle ad ottenere le performance migliori.

3.3.3 Terzo Data Set

Il terzo Data Set è quello più numeroso poiché i pazienti 1, 3 e 5 presentano un totale di 25 slice; il paziente 4 ha 24 slice ed infine, il paziente 2 solo 23 slice. Quest'ultimo blocco di slice ricopre, circa, la sezione anatomica della sinfisi pubica, le ossa basse del bacino.

La terza suddivisione di Training Set evidenzia delle ottime performance, soprattutto per i pazienti 1, 4 e 5. In questi pazienti, tutti i parametri sono superiori al 90%, ciò significa che non ci sono state evidenti criticità a riconoscere le aree di zone di midollo osseo rosso.

Nel paziente 2, i box plot evidenziano delle performance inferiori, ma comunque promettenti. Infatti, il 75esimo percentile del DSC si assesta intorno al 90%, il valore minimo è appena sotto il 70% ed infine, il primo percentile è appena sotto il 80%. La Pr, invece, ha mediana intorno al 90%; la Re ha maggiore variabilità, il valore minimo ha valore pari a 55%, tuttavia il terzo quartile ha valore superiore al 90%.

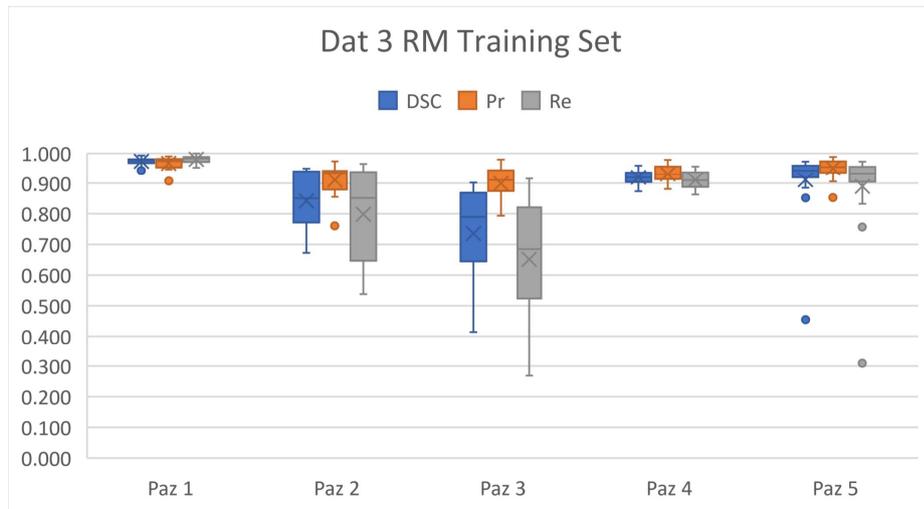
Anche in questa suddivisione il terzo paziente presenta performance inferiore e molte più criticità rispetto agli altri pazienti. Il valore di DSC massimo si assesta al 90%; tuttavia, il primo quartile è al 65% e il terzo quartile al 85%. Anche se la precisione resta abbastanza alta con mediana al 90%, la recall ha una grande variabilità e valori bassi. Infatti, il 75esimo percentile è poco sopra il 80% ed il 25esimo percentile ha valore poco sopra 50%.

La terza suddivisione di Validation Set conferma le ottime performance del paziente 1, infatti tutti i parametri sono superiori al 90%.

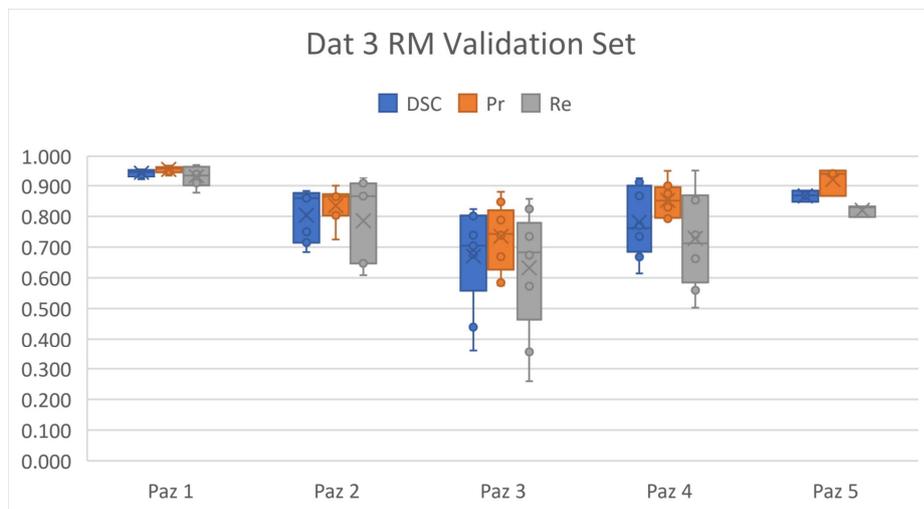
I pazienti 2 e 5 hanno performance molto coerenti con quelli ottenuti nel Training Set. Infatti i risultati sono quasi comparabili; per il paziente 2 si ha un DSC che nel terzo quartile è poco sotto il 90% e coincide con il massimo. Il paziente 5, invece, ha un DSC ed una Re inferiori del 5% rispetto alle performance del Training Set.

Il paziente 3, già nel Training Set, mostrava performance inferiori rispetto agli altri pazienti. Questa situazione si ripete anche nel Validation Set in quanto i tre parametri sono tutti inferiori al 85%.

Il paziente 4, rispetto alle performance ottenute nel Training Set, ha evidenziato un lieve calo nel Validation Set. Seppur ottiene il 90% nel terzo quartile di DSC e Pr, il primo quartile ha valore poco inferiore al 70% per il DSC e pari al 60% per la Re.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.175

Nel terzo Data Set utilizzando per addestrare la rete nell'individuamento di zone YM, si osserva che per quanto riguarda il paziente 1 nel Training Set YM si sono ottenuti dei risultati eccellenti, infatti tutti e tre i parametri sono superiori al 95%. Ciò si verifica anche con il paziente 2; il quale, a meno di un outlier che ha una Re intorno al 85%, presenta delle slice che sono state predette molto bene.

Il paziente 3 mostra performance lievemente inferiori rispetto ai precedenti pazienti. Il DSC presenta come valore massimo 95%, ha una mediana intorno al 90% e un 25esimo percentile a 85%. Anche questo paziente presenta un outlier con precisione appena sotto il 80%. Infine, la Re non è simmetrica e ha massimo in 95% e minimo in 85%.

Il paziente 4 ha comportamento simile al paziente 3 però le performance sono aumentate del 5%, infatti viene raggiunta una recall pari al 100%. Tuttavia, presenta valore minimo in 65% e vi sono 2 outlier che si riflettono pure sul box plot del DSC.

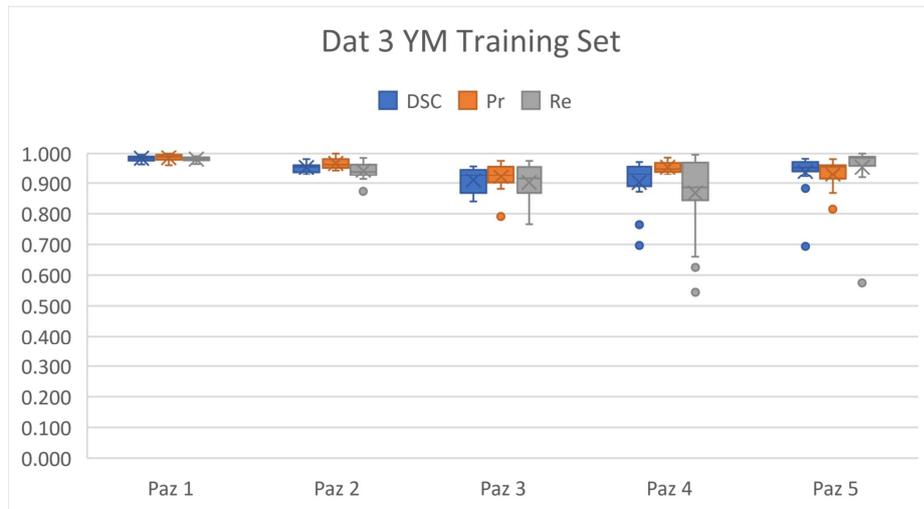
Il paziente 5 ha un andamento molto più simile ai pazienti 1 e 2; per questo paziente, tutti e tre i parametri hanno ottenuto generalmente valori superiori al 90%. Le uniche eccezioni sono i due outlier e il minimo della precisione che si assesta intorno al 85%.

Con la terza suddivisione di Validation Set YM si sono ottenuti valori che, seppur inferiori al Training Set, sono comunque promettenti. Soprattutto nel paziente 1, i tre parametri calcolati hanno dato tutti valori superiori al 100%; l'unica eccezione è una slice con la quale si è ottenuta una recall con valore di 85%.

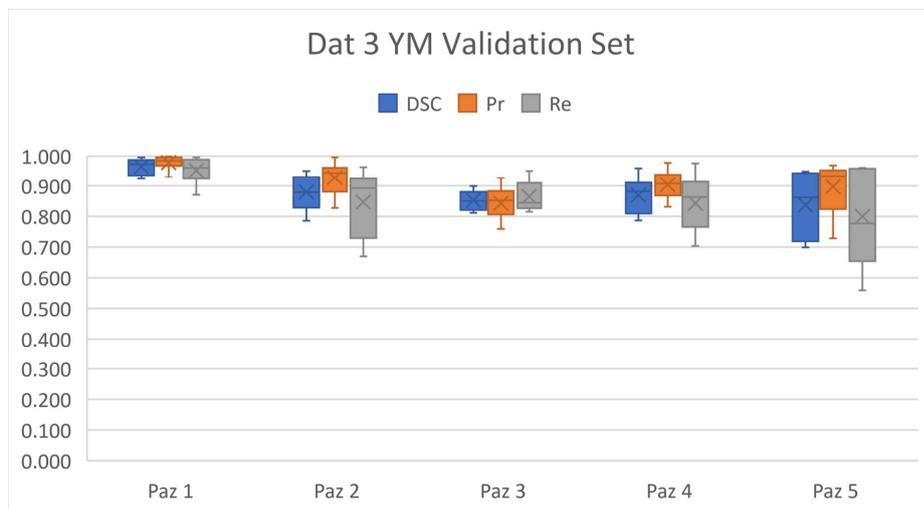
Le performance del paziente 2 sono inferiori nel Validation Set YM. I tre parametri sono scesi di circa il 5%. Tuttavia, il DSC ha una mediana di 85% e primo quartile pari ad 80%; la recall ha una maggiore variabilità di valori ed il 25esimo percentile ha valore circa il 75%.

Le performance del paziente 3 nel Validation Set YM restano coerenti con quelle ottenute nel Training Set YM. Nuovamente, si osserva un lieve decremento del 5%, tuttavia si ha un aumento della recall poiché non si ha più un minimo sotto il 80%. Ciò si verifica anche per il paziente 4, infatti le performance ottenute mediamente hanno valore intorno al 85%

Il paziente 5 è quello che mostra il peggioramento più evidente nel Validation Set YM; infatti, la distribuzione di valori aumenta e il DSC ha un 25esimo percentile appena sopra il 70%. Anche la recall ha una maggiore variabilità, il valore minimo è pari a 55% ed il 25esimo percentile è pari al 65%.



(a) Box Plot che riassume i valori di DSC, Precisione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precisione e Recall nel Validation Set.

Figura 3.176

Risultati Paziente 1 - RM

Il paziente 1, come scritto prima, ha ottenuto solo predizioni positive e superiori al 90%. Questo è visibile nelle slice n183.5 e n225.5 (Figure 3.177 e 3.178) del Training Set RM che nel Validation Set RM con le slice n177.5 e n201 (Figure 3.179 e 3.178).

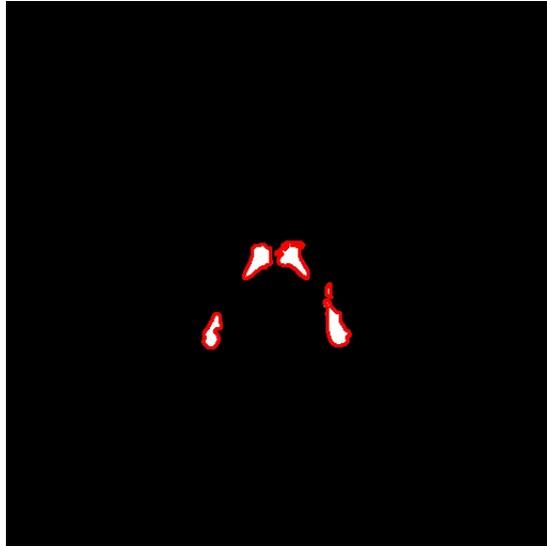


Figura 3.177: Slice n183.5 del paziente 1 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

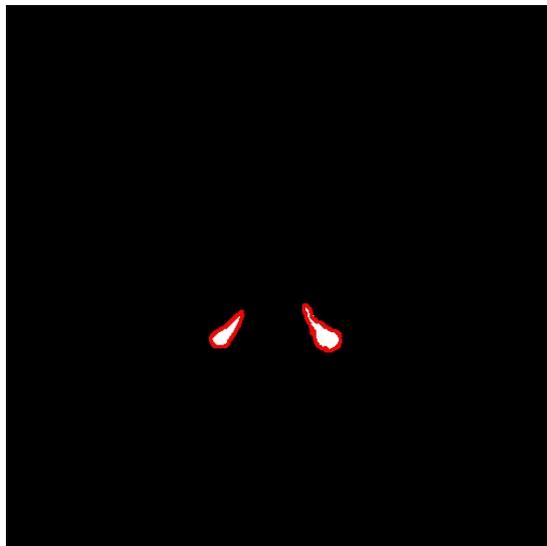


Figura 3.178: Slice n225.5 del paziente 1 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

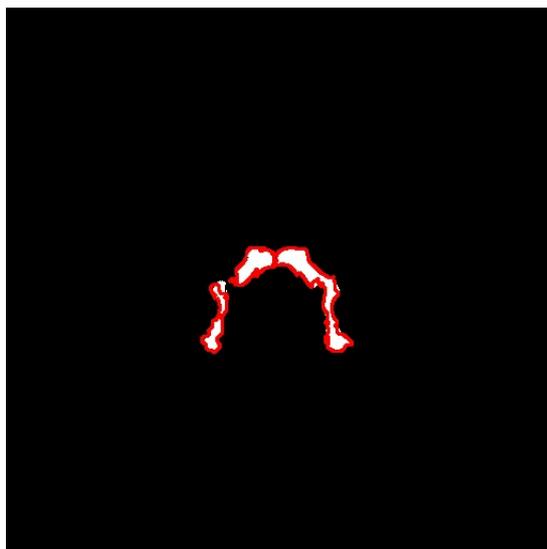


Figura 3.179: Slice n177.5 del paziente 1 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

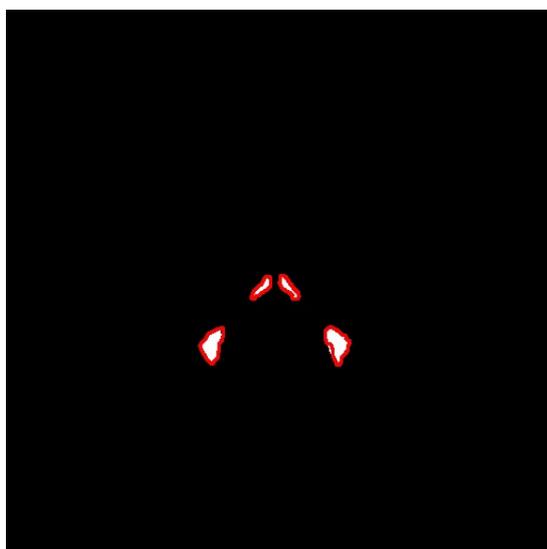


Figura 3.180: Slice n201.5 del paziente 1 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 1 - YM

Il paziente 1, ha ottenuto solo predizioni positive e superiori al 90% anche nel caso di riconoscimento di zone YM. Questo è visibile nelle slice n192.5 e n228.5 (Figure 3.181 e 3.182) del Training Set RM che nel Validation Set RM con le slice n189.5 e n234 (Figure 3.183 e 3.184).



Figura 3.181: Slice n192.5 del paziente 1 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

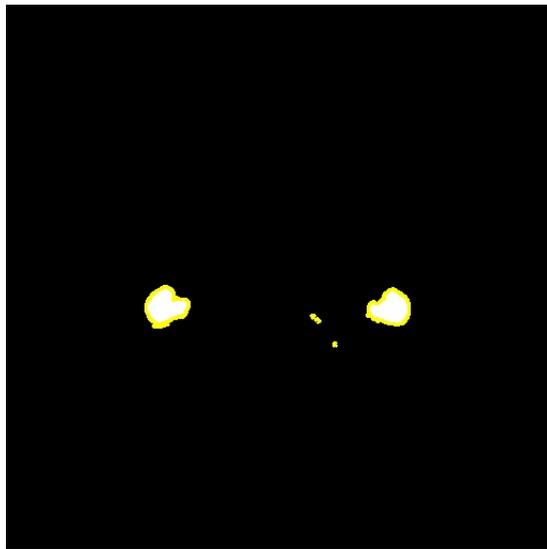


Figura 3.182: Slice n228.5 del paziente 1 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

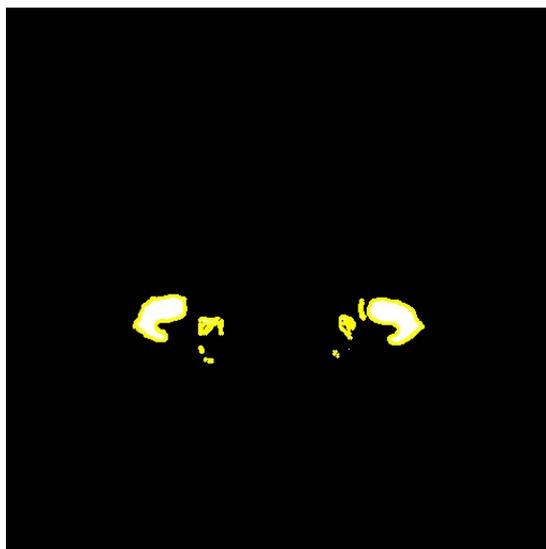


Figura 3.183: Slice n189.5 del paziente 1 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

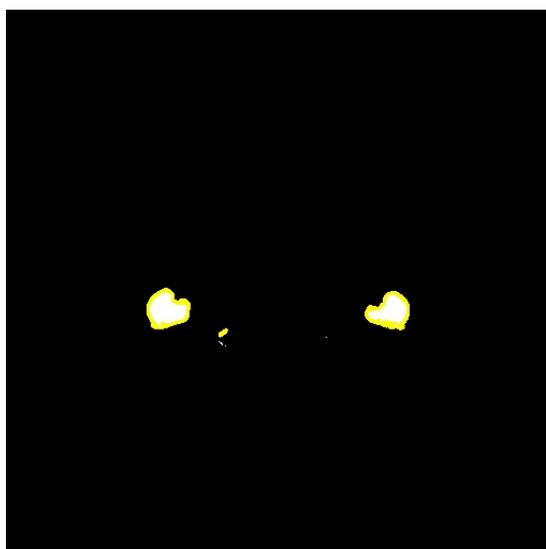


Figura 3.184: Slice n234.5 del paziente 1 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 2 - RM

Nel paziente 2 le criticità sono presenti quando le zone di midollo osseo rosso sono molte strette e sparpagliate. Tuttavia, questi problemi non portano a bassi valori di DSC nel Training Set. La slice n186 è l'unica ad avere un valore inferiore al 70% e si ha difficoltà nel riconoscere zone di RM che sono composte da 2-3 pixel al massimo.



Figura 3.185: Slice n186 del paziente 2 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Quando le zone di RM diventano molto più compatte si hanno predizione ottime. Questo è l'esempio delle slice n210 e n225 (Figure 3.186 e 3.187) che rappresentano la parte di sinfisi pubica.

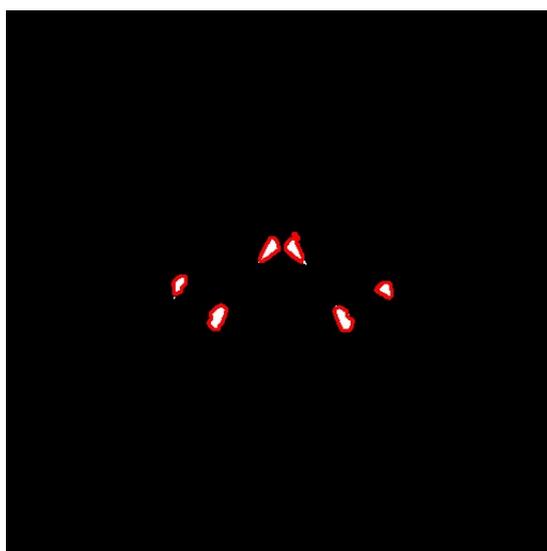


Figura 3.186: Slice n210 del paziente 2 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

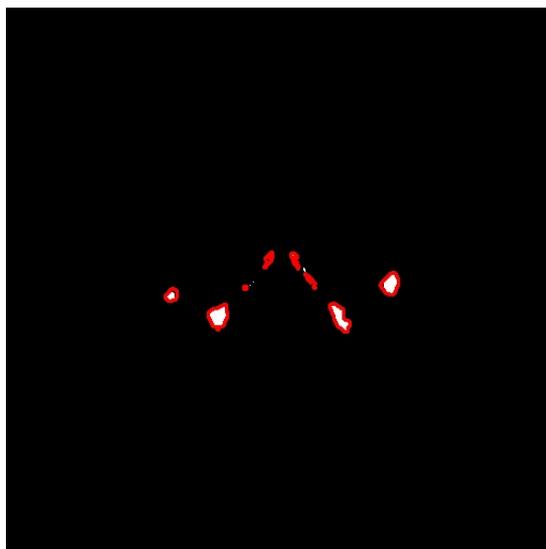


Figura 3.187: Slice n225 del paziente 2 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Nel Validation Set RM si hanno gli stessi problemi poiché, spesso, le zone di RM sono strette ed è difficile individuarle, come nella slice n183 (Figura 3.177). Quando invece si ha una sezione in cui la sinfisi pubica è molto più visibile, in quel caso le performance aumentano fino a superare anche il 90% per il DSC. Ciò accade nelle slice n234 e n198 (Figure 3.189 e 3.190).

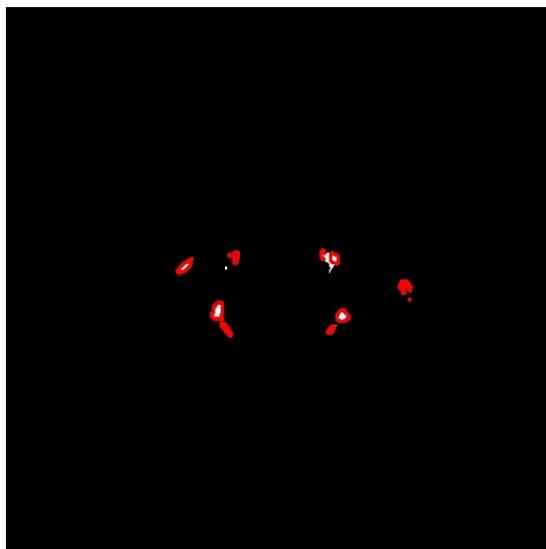


Figura 3.188: Slice n183 del paziente 2 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

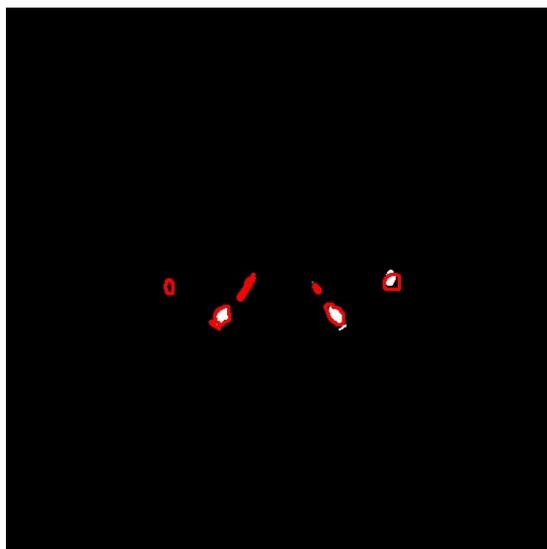


Figura 3.189: Slice n234 del paziente 2 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

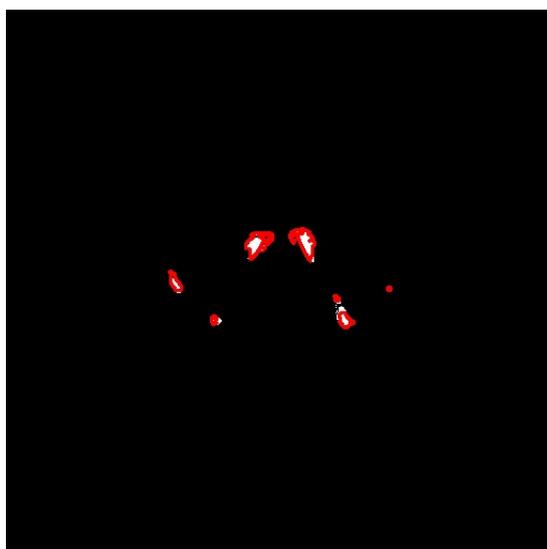


Figura 3.190: Slice n198 del paziente 2 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

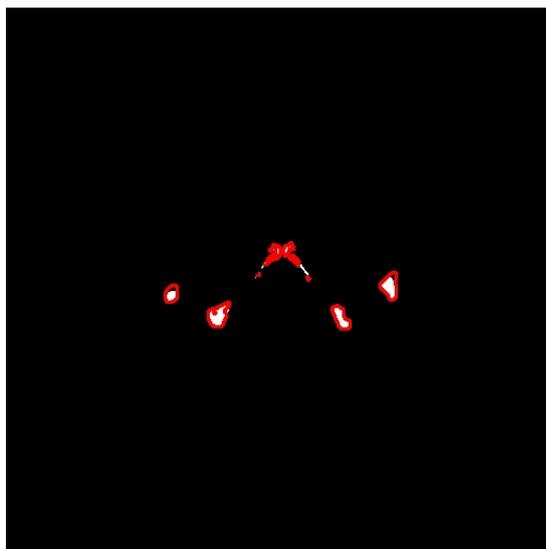


Figura 3.191: Slice n222 del paziente 2 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 2 - YM

La porzione del paziente 2 relativa al Training Set YM evidenzia, proprio come era successa nella seconda suddivisione, un'ottima predizione delle zone YM relative ai femori. Ciò è visibile nella slice n189 (Figura 3.192) nel quale si ottiene un DSC pari al 98%, ma in generale l'andamento è sempre superiore al 90%.

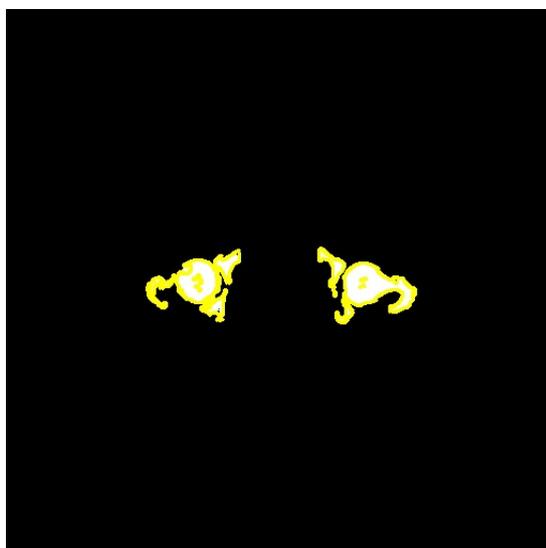


Figura 3.192: Slice n189 del paziente 2 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

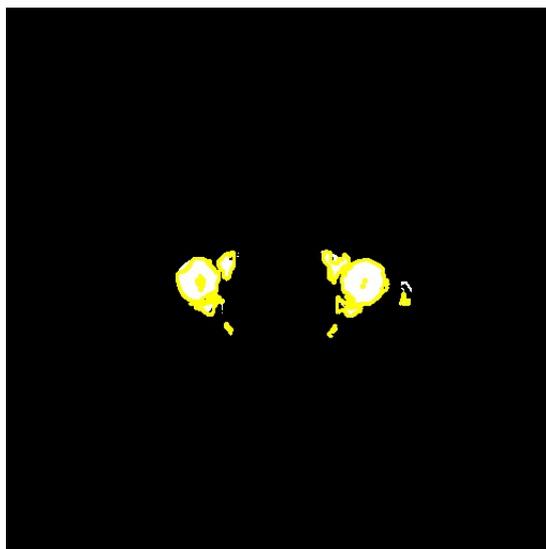


Figura 3.193: Slice n183 del paziente 2 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

L'unica slice ad avere una performance inferiore con DSC pari al 78% è la n201 (Figura 3.194); in essa si osserva la corretta predizione dei femori, ma la parte adiacente soffre un sotto-campionamento di zone YM.



Figura 3.194: Slice n201 del paziente 2 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 3 - RM

Il paziente 3 non ha ottenuto performance eccezionali come nei pazienti 1, 4 e 5. Il Training Set RM ha evidenziato alcune criticità date da slice, come la n226 e n286, visibili nelle Figure 3.195 e 3.196, in cui la zona di RM è troppo limitata rispetto anche lo sfondo nero.

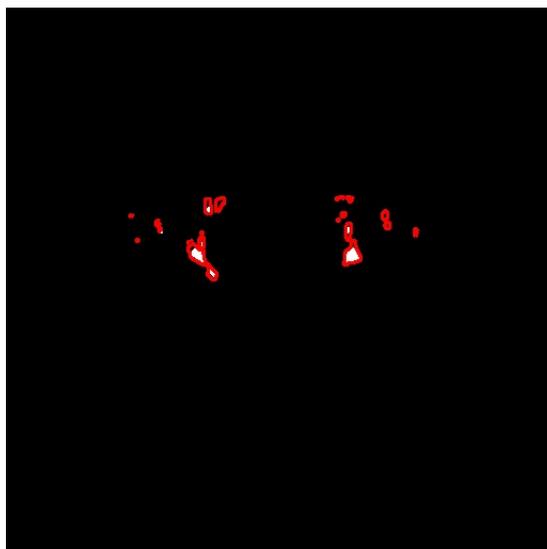


Figura 3.195: Slice n225 del paziente 3 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

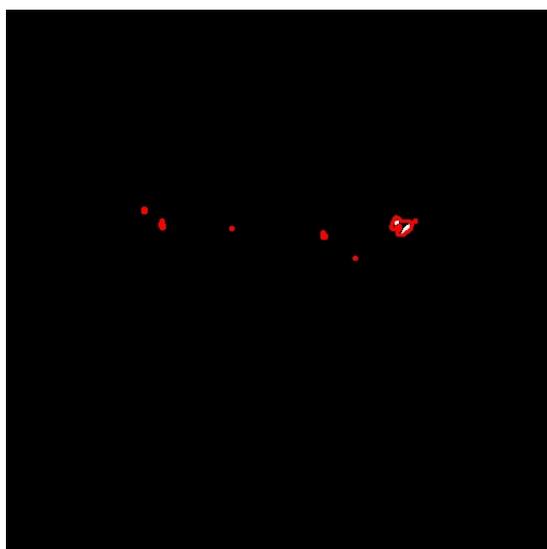


Figura 3.196: Slice n286 del paziente 3 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Le performance ottimali si sono, nuovamente, ottenute quando l'area di RM da individuare riguarda la sinfisi pubica. Le slice n235 e n253 nelle Figure 3.197 e 3.198 hanno valori di DSC rispettivamente pari a 87.2% e 90.2%.

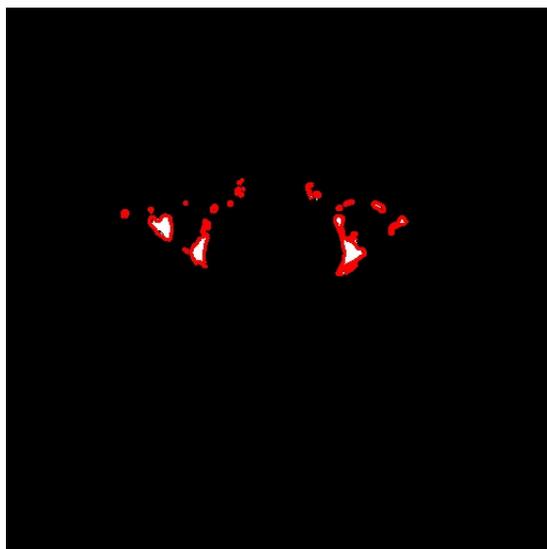


Figura 3.197: Slice n235 del paziente 3 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

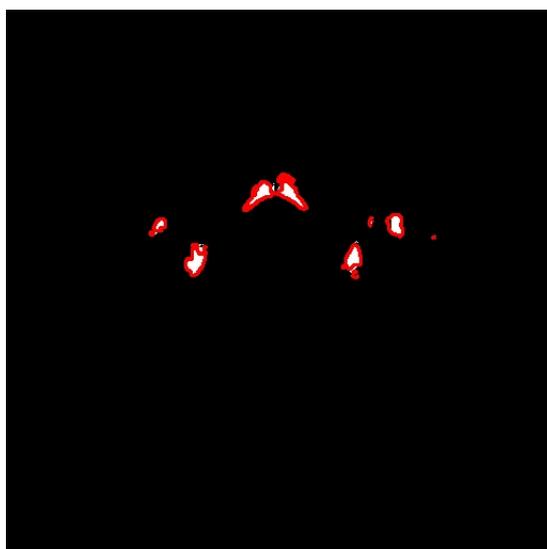


Figura 3.198: Slice n253 del paziente 3 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Il paziente 3 ottiene delle performance non ottime nel Validation Set RM, questo è dovuto alla presenza di molte slice in cui la zone di RM è frastagliata e molto irregolare come visibile nelle slice n223 e n277 (Figure 3.199 e 3.200).

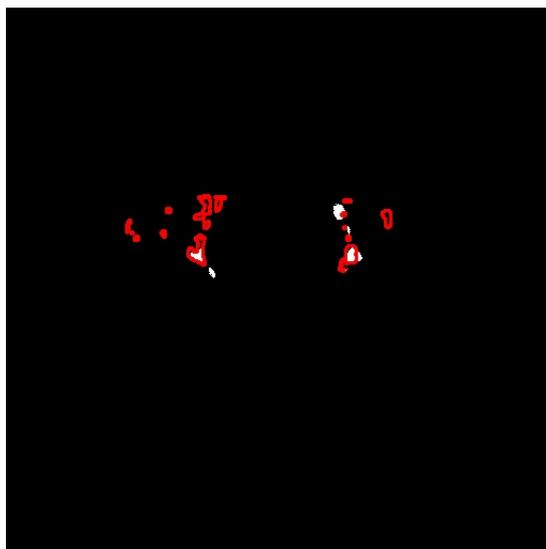


Figura 3.199: Slice n223 del paziente 3 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.



Figura 3.200: Slice n277 del paziente 3 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 3 - YM

Il paziente 3 ha mostrato dei buoni risultati nell'individuazione delle zone YM; in generale i valori di DSC sono tutti superiori al 90% e l'unica eccezione è la slice n271 con DSC pari al 84%. Come si vede in Figura 3.201, c'è stato un problema nell'individuazione della testa femorale.

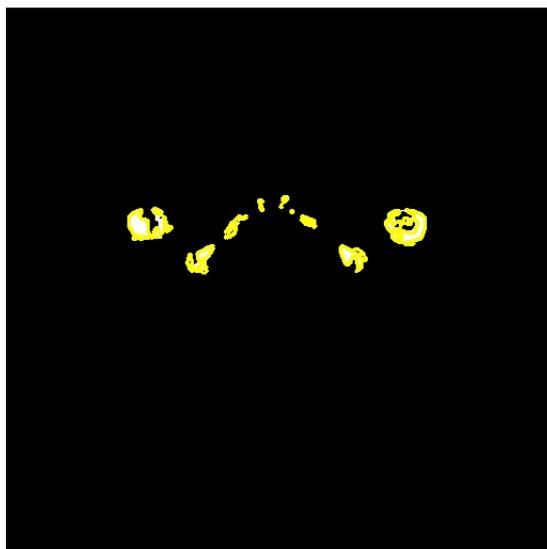


Figura 3.201: Slice n271 del paziente 2 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Le slice n226 e n283 nelle Figure 3.195 3.203 mostrano come la zona YM dei femori è stata correttamente predetta pur mostrando molte irregolarità, infatti hanno valore di DSC intorno al 95%.

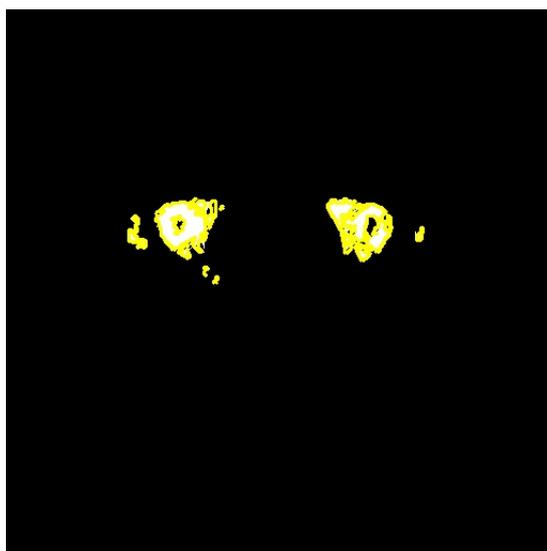


Figura 3.202: Slice n226 del paziente 2 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.203: Slice n283 del paziente 2 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Nel Validation Set YM vengono confermati gli andamenti presentanti nel Training Set YM sopra descritto. Nello specifico, si hanno problemi quando la zona di YM diventa troppo irregolare e frastagliata, anche se è relativa alle teste femorali che in generale permettevano delle buone predizioni; la slice n223 in Figura 3.204 è un esempio di queste criticità, osservando che la stessa slice aveva dato problemi anche nell'individuazione di zone RM.

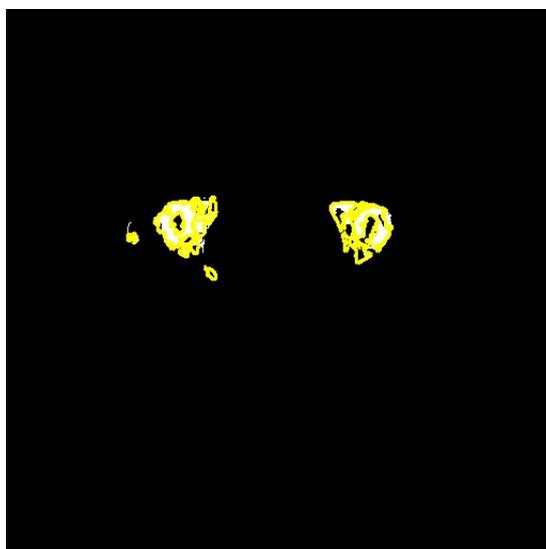


Figura 3.204: Slice n223 del paziente 2 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Le slice n232 e n256 nelle Figure 3.205 e 3.206 sono due esempi di immagini in cui la predizione ha ottenuto un DSC rispettivamente pari a 90.1% e 87.7%.

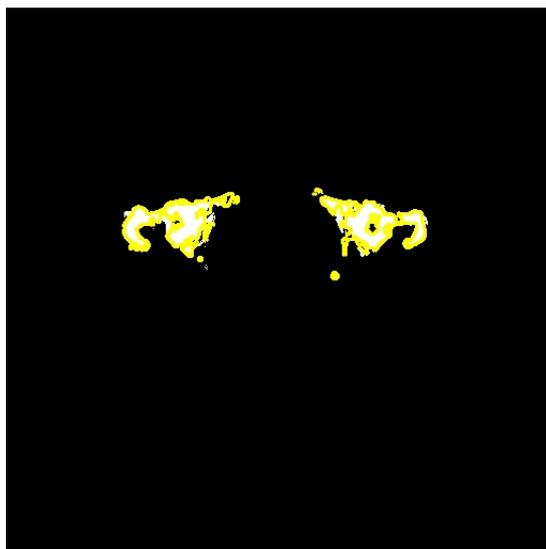


Figura 3.205: Slice n232 del paziente 2 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.206: Slice n256 del paziente 2 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 4 - RM

Il paziente 4 nel terza suddivisione del Training Set presenta tutte performance superiori al 90% ed un esempio è dato dalle due slice n155.5 e n194.5 nelle Figure 3.207 e 3.208

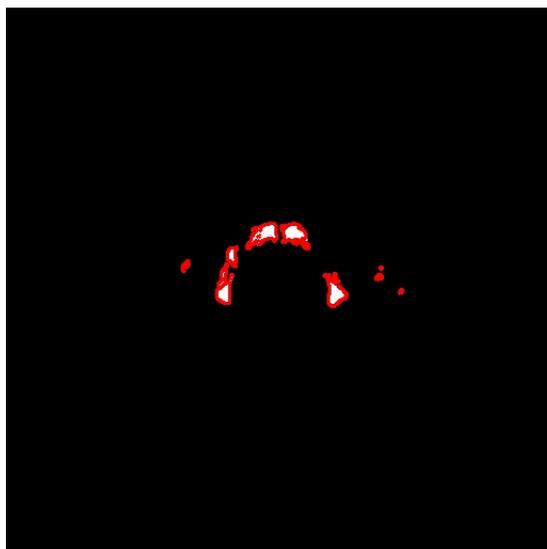


Figura 3.207: Slice n155.5 del paziente 4 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

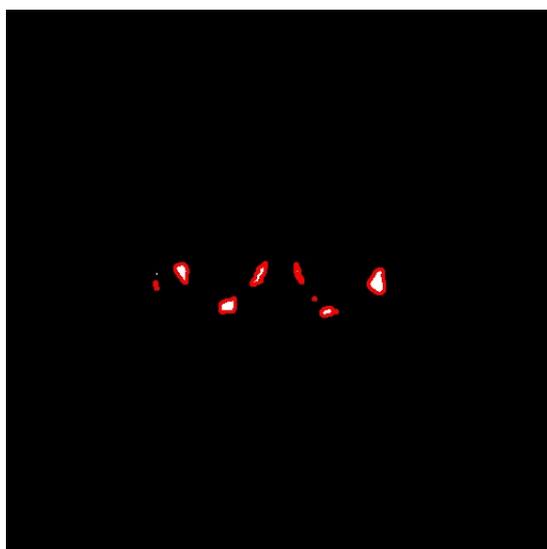


Figura 3.208: Slice n194.5 del paziente 4 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

A differenza del training set, nel Validation Set il modello ha difficoltà nel predire le zone di RM, poiché queste hanno una dimensione ridotta e presentano delle irregolarità evidenti nelle forme. Un esempio sono le slice n164.5 e 206.5 nelle Figure 3.209 e 3.210, in questo caso il DSC vale rispettivamente 74.5% e 61.4%.

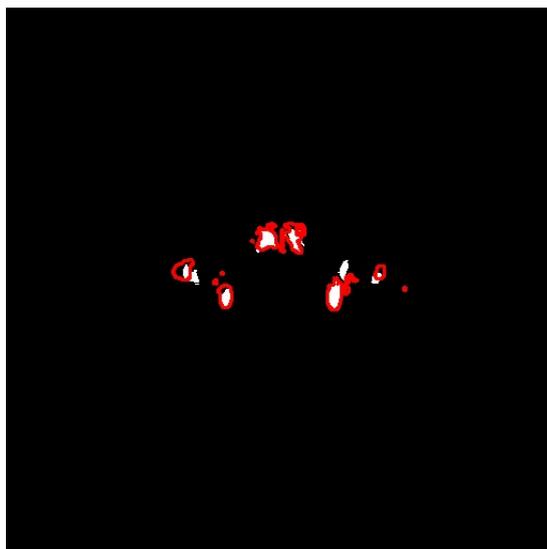


Figura 3.209: Slice n164 del paziente 4 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

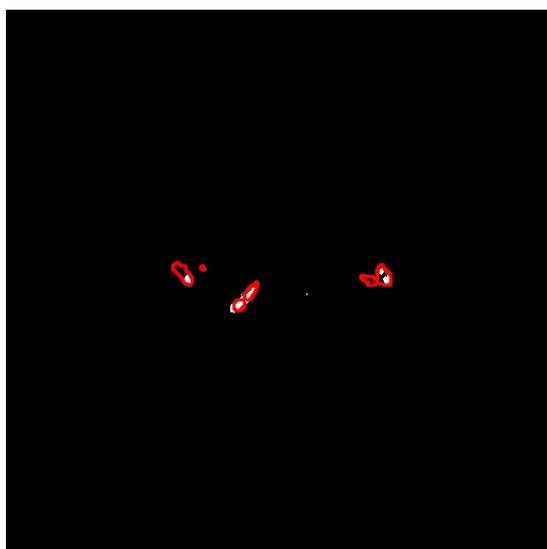


Figura 3.210: Slice n206 del paziente 4 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Tuttavia alcune slice relative alla parte della sinfisi pubica hanno mostrato dei buoni risultati, come ad esempio la n149.5 e 200.5 in Figura 3.211 e 3.212.

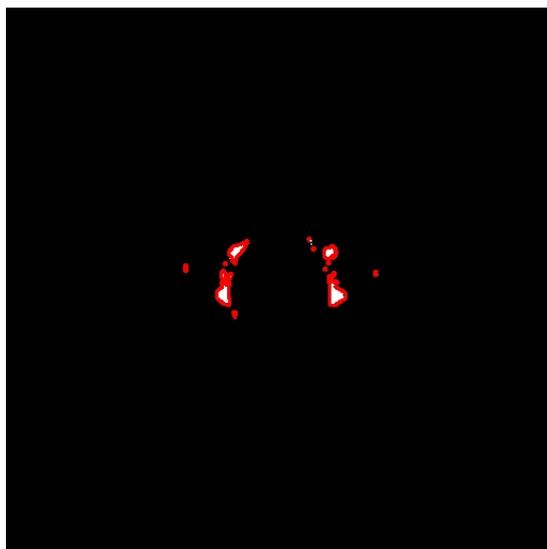


Figura 3.211: Slice n149.5 del paziente 4 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

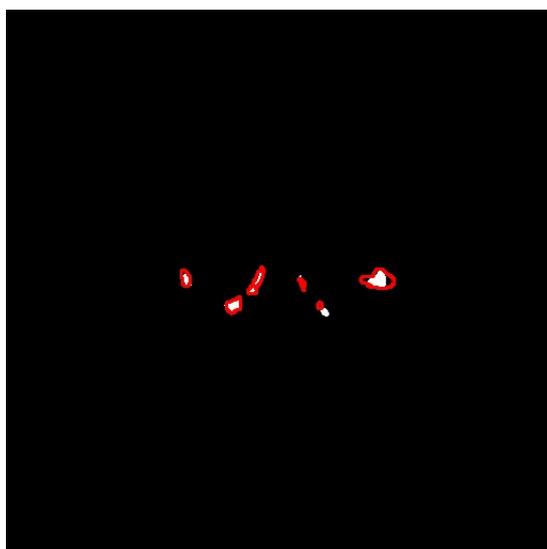


Figura 3.212: Slice n200.5 del paziente 2 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 4 - YM

In maniera analoga al caso RM, nel Training Set, la rete raggiunge dei buoni risultati nel paziente 4 anche nell'identificazione del YM, come ad esempio nelle slice n191.5 e n143.5, nelle Figure 3.213 e 3.214. L'unica criticità si può evidenziare nella slice n215.5 in Figura 3.213



Figura 3.213: Slice n215.5 del paziente 4 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.214: Slice n191.5 del paziente 4 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

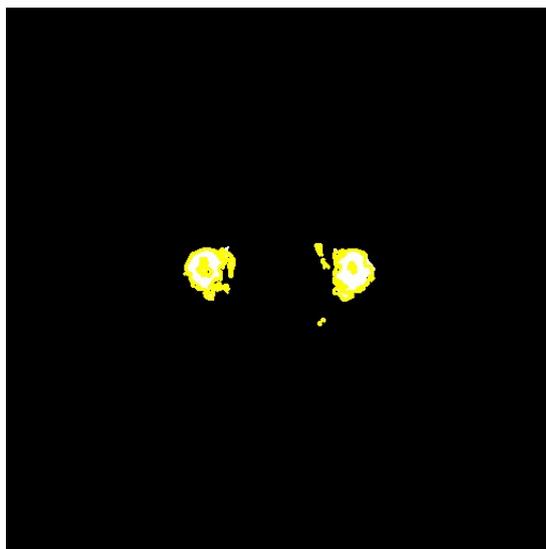


Figura 3.215: Slice n143.5 del paziente 4 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Nel Validation Set, rispetto al caso RM, osserviamo un notevole miglioramento delle performance. Nello specifico nelle slice 206.5 e 149.5 (Figure 3.216 e 3.217) si è riusciti a incrementare il DCF di circa 10%, mentre invece nei casi in cui la performance in RM era buona il risultato è rimasto pressoché invariato, come nella slice n149.5 in Figura 3.217.



Figura 3.216: Slice n206.5 del paziente 4 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

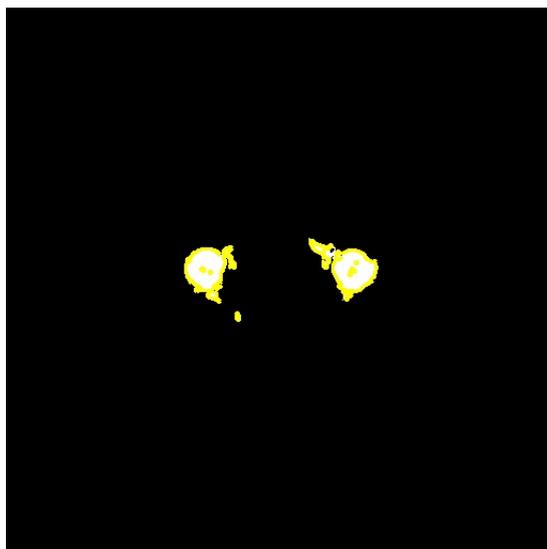


Figura 3.217: Slice n149.5 del paziente 4 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.



Figura 3.218: Slice n185.5 del paziente 4 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

Risultati Paziente 5 - RM

Nel training set del Paziente 5 l'identificazione del RM risulta accurata in tutte le slide considerate, ad eccezione della slice n216 in Figura 3.219 che ha ottenuto un DCF del 60%. I risultati migliori sono stati ottenuti nelle slice n150 e n189, nelle Figure 3.220 e 3.221.

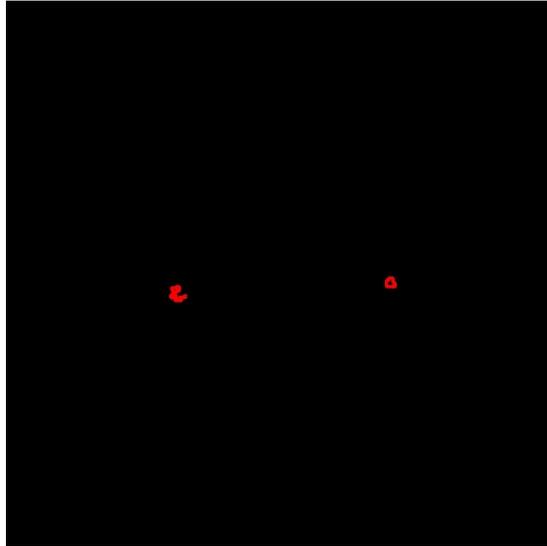


Figura 3.219: Slice n216 del paziente 5 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

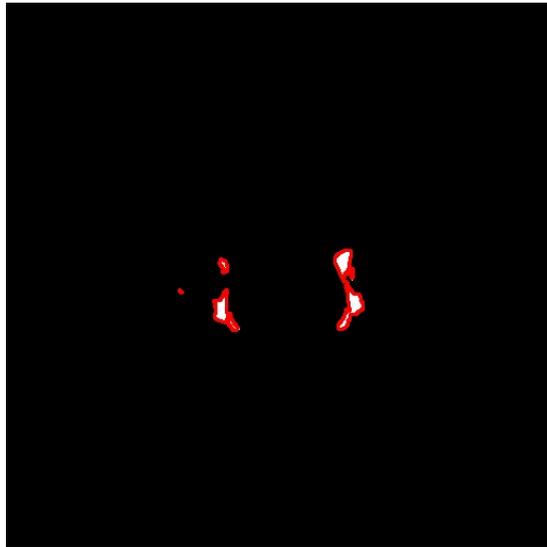


Figura 3.220: Slice n150 del paziente 5 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

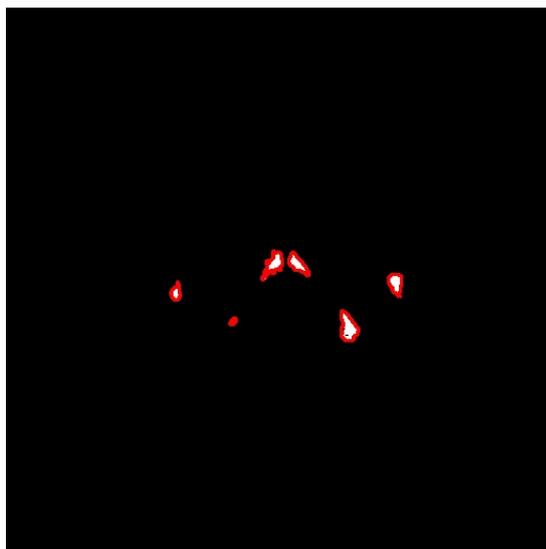


Figura 3.221: Slice n189 del paziente 5 relativa alla terza suddivisione del Training Set RM, i bordi rossi indicano la MaskRM effettiva.

Il validation set ha raggiunto delle buone performance, come ad esempio nella slice 153 in Figura 3.222. In altre slice più critiche, con aree di RM piccole o forme irregolari, le performance sono rimaste generalmente basse, come nella slice n213 in Figura 3.223 che ha ottenuto un DCF del 13.3%

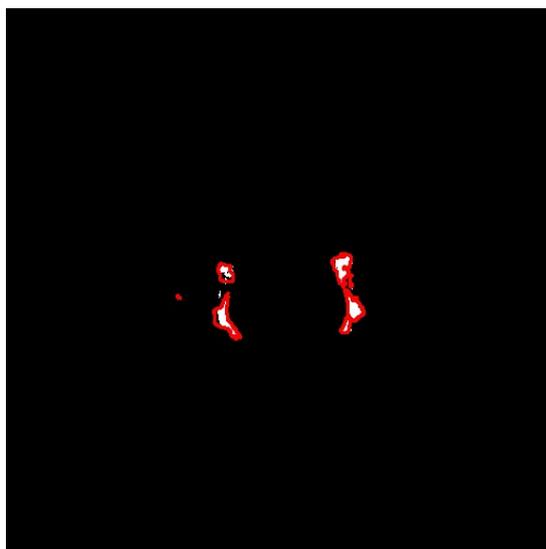


Figura 3.222: Slice n153 del paziente 5 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

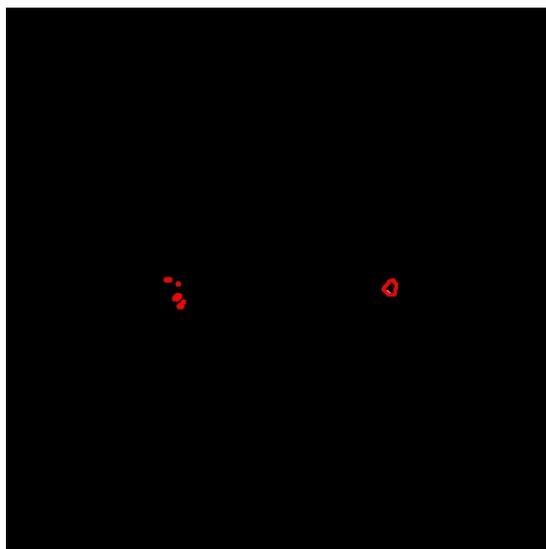


Figura 3.223: Slice n213 del paziente 5 relativa alla terza suddivisione del Validation Set RM, i bordi rossi indicano la MaskRM effettiva.

Risultati Paziente 5 - YM

I risultati nel YM sono rimasti pressoché invariati e si sono ottenute performance simili nei casi migliori e peggiori. Nelle Figure 3.224, 3.225 e 3.226 sono mostrate le slice prese in considerazione nel riconoscimento del RM con la maschera YM.

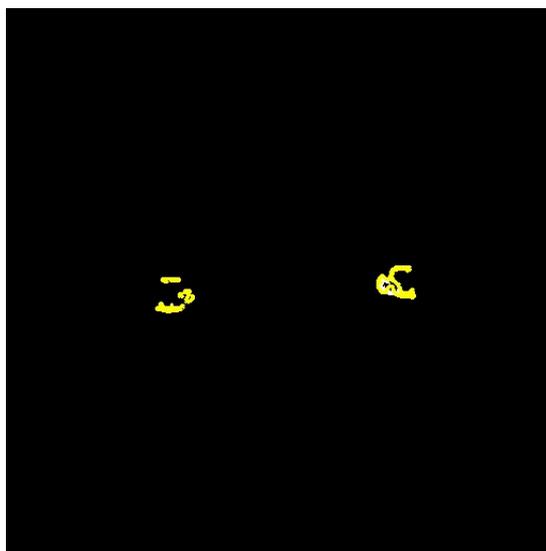


Figura 3.224: Slice n216 del paziente 5 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

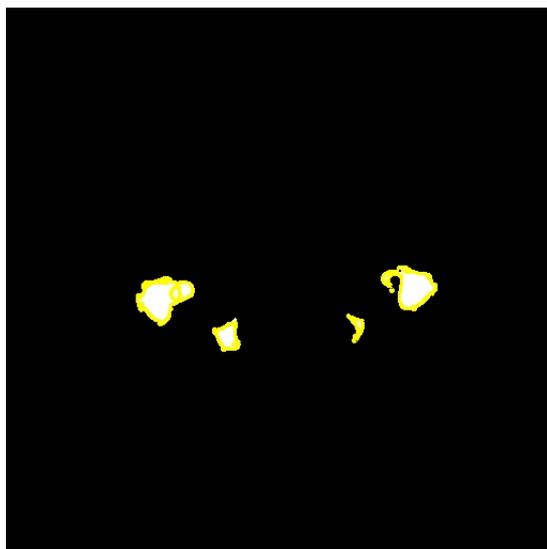


Figura 3.225: Slice n189 del paziente 5 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

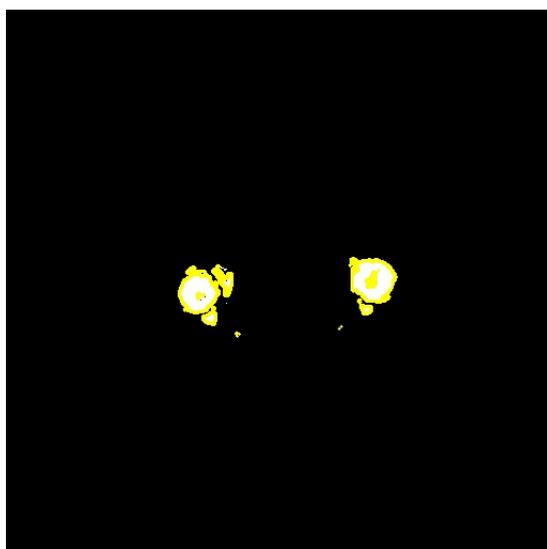


Figura 3.226: Slice n150 del paziente 5 relativa alla terza suddivisione del Training Set YM, i bordi gialli indicano la MaskYM effettiva.

Come per il training set, nelle Figure 3.227, 3.222 sono state presente in considerazione le slice utilizzate nella classificazione di RM, poiché presentavano risultati simili. Questo ad eccezione della slice n153 che nel caso di RM presentava un DCF di 0%, mentre nell'identificazione di RM si è raggiunto un valore del 70%.



Figura 3.227: Slice n219 del paziente 5 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

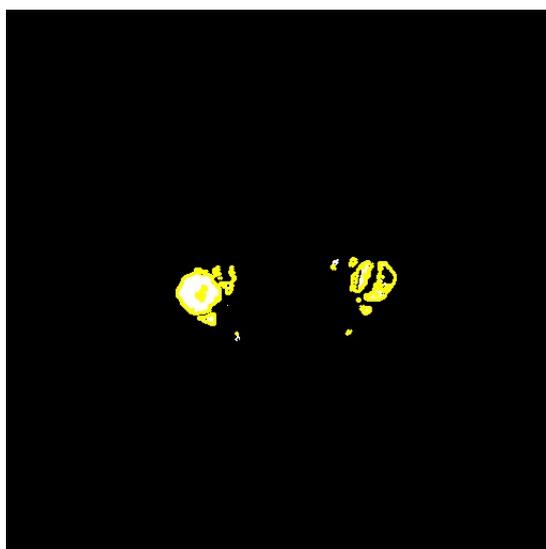


Figura 3.228: Slice n153 del paziente 5 relativa alla terza suddivisione del Validation Set YM, i bordi gialli indicano la MaskYM effettiva.

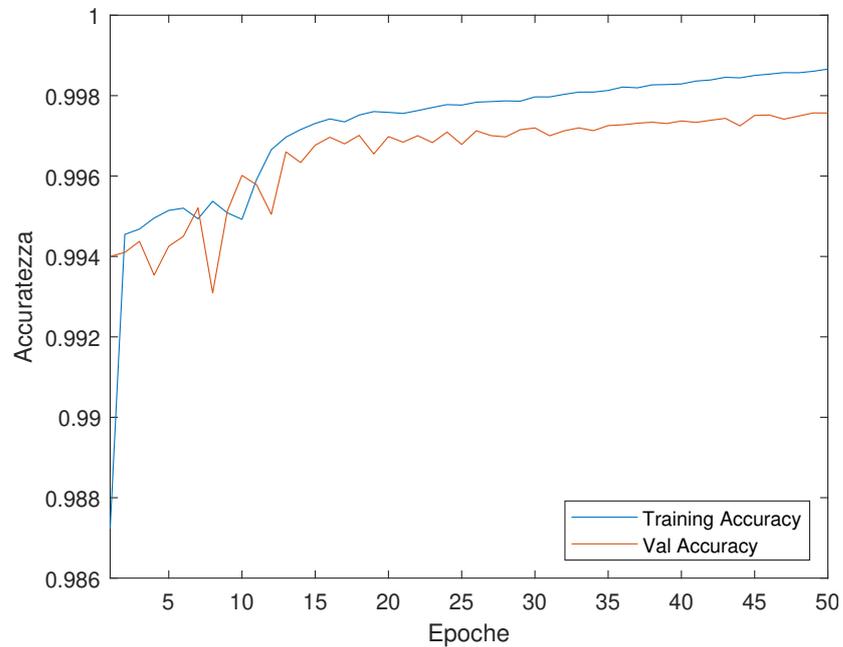
Riassumendo, questa terza ed ultima suddivisione del Data Set ha mostrato nuovamente la facilità con cui le zone di YM presenti nelle teste femorali sono facilmente individuabili, a meno di qualche sporadica criticità. La sezione del pube in cui è visibile la sinfisi pubica che presenta aree di midollo osseo rosso, non ha mostrato grandi difficoltà nell'essere individuata. Anche in questo caso ci sono state delle criticità non rilevanti poiché le performance non sono mai scese sotto il 80% nella maggior parte dei casi.

3.4 Applicazione U-Net con 6 pazienti

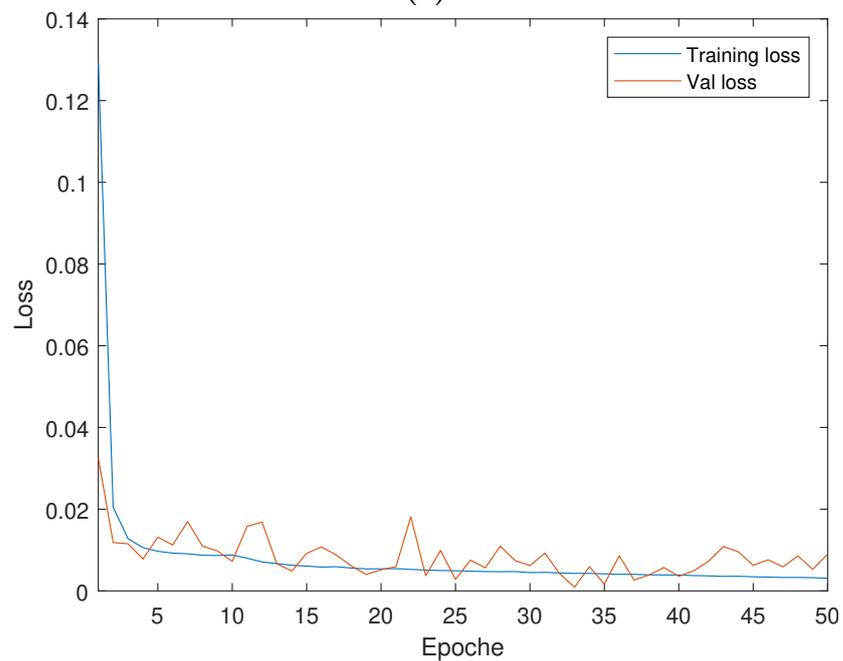
Per ultimare gli studi condotti su questo Data Set, è stato inserito un sesto paziente, mai utilizzato prima e sempre elaborato nei precedenti studi [14]. In questo modo si è potuto analizzare se, l'aggiunta di nuove informazioni, possa beneficiare la rete. Ovviamente,

anche il paziente 6 come tutti gli altri, ha affrontato una fase pre-processing per la normalizzazione e successivamente è stato aggiunto ai Training Set e Validation Set pre-esistenti con i 5 pazienti (senza suddivisione, quindi quello trattato nel Capitolo 3.2) in maniera totalmente casuale.

Di seguito l'andamento di loss e accuratezza attraverso le epoche (sempre 50).

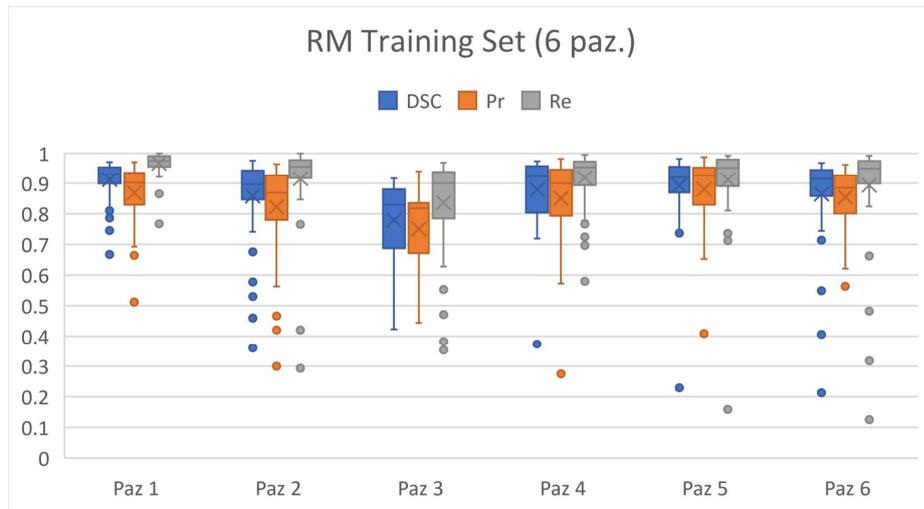


(a)

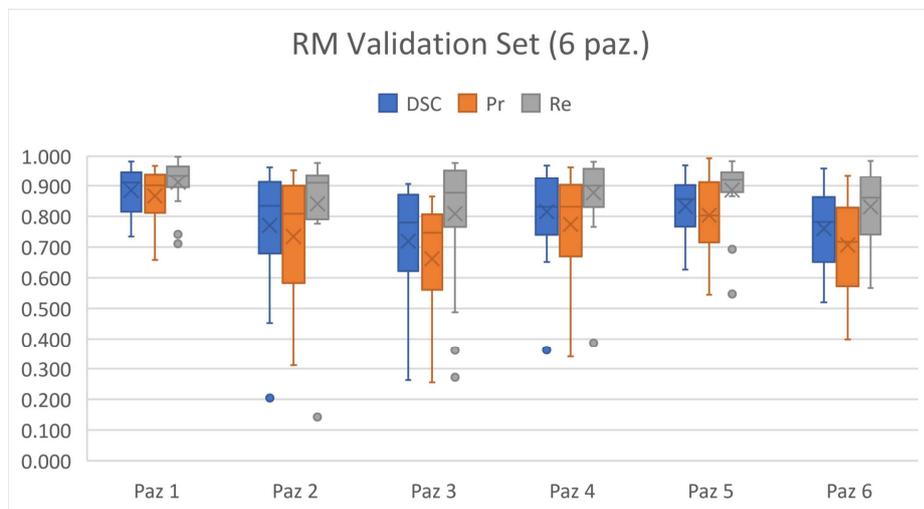


(b)

Figura 3.229



(a) Box Plot che riassume i valori di DSC, Precione e Recall nel Training Set.



(b) Box Plot che riassume i valori di DSC, Precione e Recall nel Validation Set.

Figura 3.230

Prendendo in riferimento i box plot presenti nelle Figure 3.230a e 3.230b e confrontandoli con i box plot in Figure 3.4a e 3.4b è possibile fare una descrizione delle variazioni ottenute in seguito all'aggiunta di un sesto paziente.

Approfondendo prima il Training Set, si osserva che per quanto riguarda il paziente 1, quest'ultimo con l'aggiunta del sesto paziente presenta meno outlier. I valori di DSC e Re non sono cambiati e sono sempre superiori al 90%, soltanto la Pr è scesa lievemente in quanto ora il primo quartile è inferiore al 85%.

Nel secondo paziente, la recall è decisamente migliorata poiché con l'aggiunta di un sesto paziente, la sua variabilità si è ridotta ed ora già dal primo quartile supera il 90%. Il DSC, invece, è rimasto pressoché uguale, mentre il primo quartile della precisione è scesa appena sotto il 80%. Inoltre, sono comparsi molti più outlier.

Le performance del terzo paziente non hanno beneficiato dell'aggiunta del sesto paziente. DSC, Pr e Re sono rimaste praticamente uguali e l'unica differenza è la presenza di più outlier.

In generale, anche le performance del quarto paziente non sono cambiate e sono rimaste comunque ottimali.

Il quinto paziente, invece, ha evidenziato un lieve calo poiché nel caso con solo 5 pazienti, il primo quartile per tutti e tre i parametri si aggirava intorno al 90%; DSC, Pr e Re, in seguito all'aggiunta del sesto paziente, sono calati lievemente di circa il 5%. Tuttavia, sono presenti meno outlier.

Il nuovo paziente, invece, presenta delle ottime performance. La mediana del DSC si assesta intorno al 90% e comunque il minimo non scende sotto il 70%. La recall mostra un primo quartile pari al 90%.

Approfondendo le performance ottenute nel Validation Set, nel paziente 1 non vengono riscontrate alcune differenze se non nella diminuzione di outlier e l'abbassamento del 25esimo percentile della precisione da 85% a 80%. Ciò avviene anche nel secondo paziente dove però il calo è più netto e fino al 60%.

Anche il terzo paziente ha un calo della precisione del 10% ed una diminuzione degli outlier. Tuttavia, DSC e Re restano praticamente uguali, ma sono presenti nuovi minimi rispettivamente a 25% e 50%.

Nel quarto paziente si osserva che la recall ha una distribuzione di valori più ristretta, infatti il 25esimo percentile è aumentato del 5%. DSC e Pr, invece, sono diminuiti del 5%; inoltre, con l'aggiunta del sesto paziente, una slice ha raggiunto il valore del 100% nella precisione.

Il quinto paziente è quello che beneficia di meno dell'aggiunta del sesto paziente, in quanto le performance nel Validation Set calano. Più nello specifico, la precisione ha un primo quartile che scende dal 70% al 55%; ciò si verifica anche con il DSC che scende da 75% a 65%.

Il paziente 6 non presenta outlier nel Validation Set. Il DSC è minore rispetto al Training Set, soprattutto si osserva che il 25esimo percentile arriva al 65% mentre prima si era ottenuto un 85%. Anche la Re è minore rispetto al Training Set, prima era sempre superiore al 90%, tranne che per un minimo ad 85%, mentre ora il primo percentile si trova al 75%.

3.4.1 Risultati Paziente 1

Le slice n147 e n36.5 viste in precedenza, e ora visibili nella nuova predizione nelle Figure 3.231 e 3.195 non hanno beneficiato in maniera evidente dell'aggiunta del sesto

paziente, infatti si osserva ancora una tendenza a sovra-campionare e il DSC è inferiore al 80% come in precedenza.

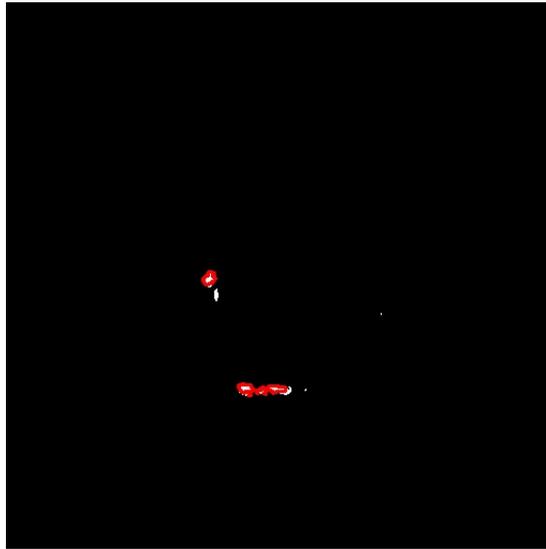


Figura 3.231: Slice n147.5 relativa al paziente 1 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

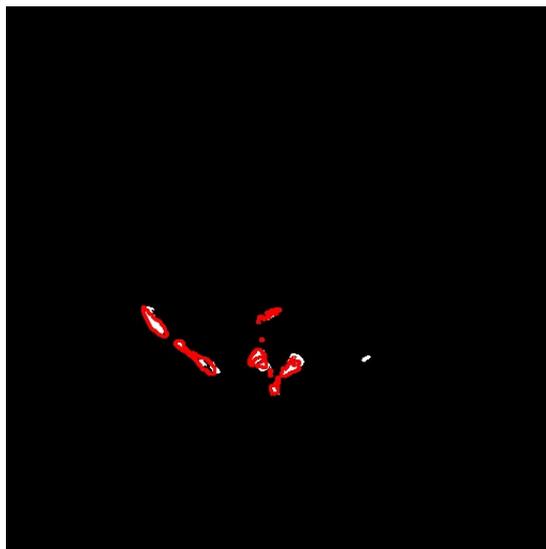


Figura 3.232: Slice n36.5 relativa al paziente 1 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

Le slice n222.5 e n183.5, visibili nelle Figure 3.233 e 3.234, rispetto allo studio fatto in precedenza pur mantenendo un valore molto alto di DSC, la comparsa di alcuni FP ha portato ad una diminuzione delle performance che rimangono comunque sopra il 90%.

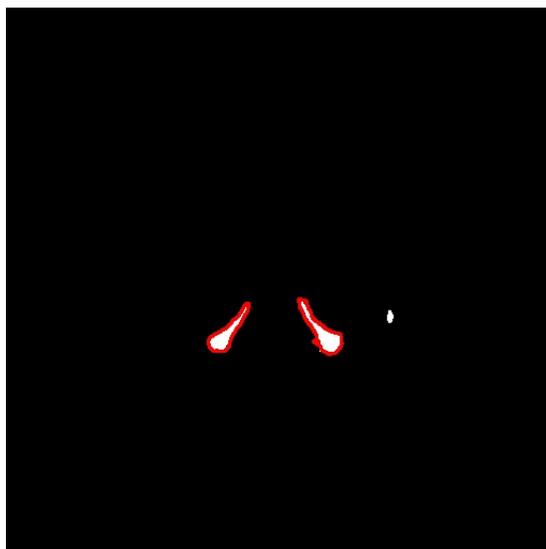


Figura 3.233: Slice n222.5 relativa al paziente 1 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

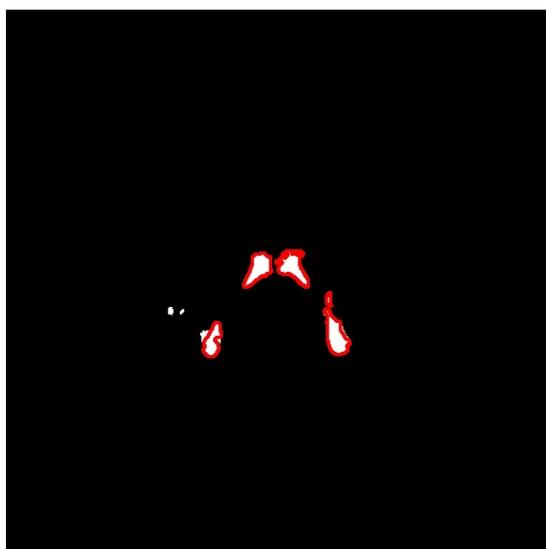


Figura 3.234: Slice n183.5 relativa al paziente 1 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

La slice n21.5 del paziente 1 ha beneficiato in maniera evidente dell'aggiunta del sesto paziente. Si osserva infatti che, quando era stata inizialmente analizzata, si erano presentati dei problemi di predizione nella parte centrale. Più nello specifico vi era un alto numero di FN. In questa nuova predizione, la parte centrale è perfettamente predetta (come visibile in Figura 3.235) e si passa da un DSC pari a 66.8% ad un ottimo 98.3%.

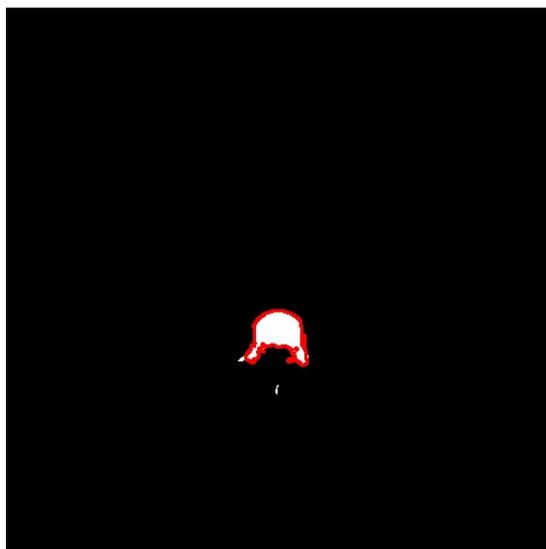


Figura 3.235: Slice n21.5 relativa al paziente 1 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

Le slice n228.5 e n195.5 nelle Figure 3.236 e 3.237 non sono state influenzate dalla presenza del sesto paziente e la loro predizione ha mantenuto un alto valore delle performance.

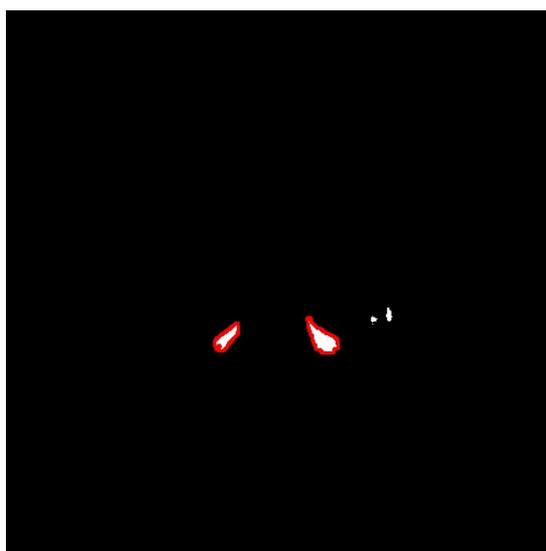


Figura 3.236: Slice n228.5 relativa al paziente 1 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

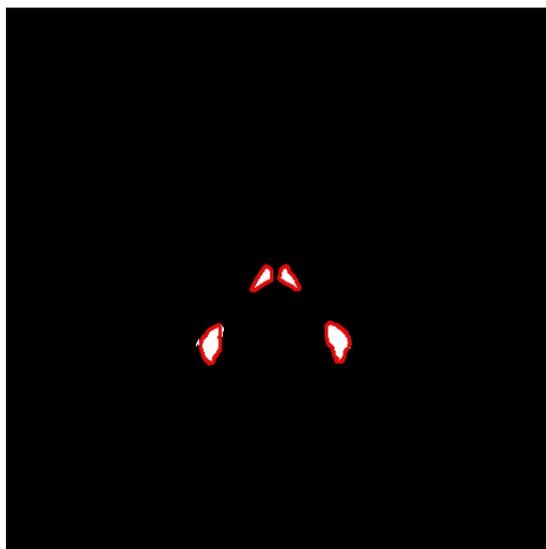


Figura 3.237: Slice n195.5 relativa al paziente 1 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

3.4.2 Risultati Paziente 2

L'aggiunta del sesto paziente non ha beneficiato alle criticità riscontrate nelle slice n51 e n189, visibile nelle Figure 3.238 e 3.239. Queste ultime continuano ad essere un problema per la rete ed inoltre, la slice n189 evidenzia un alto tasso di FP in relazione alle vere zone di RM.

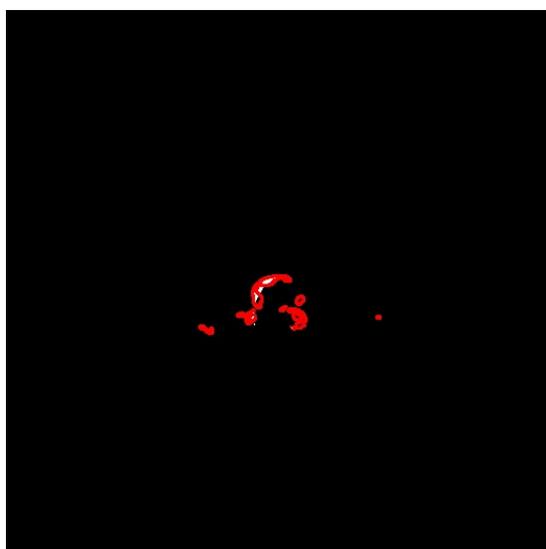


Figura 3.238: Slice n51 relativa al paziente 2 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

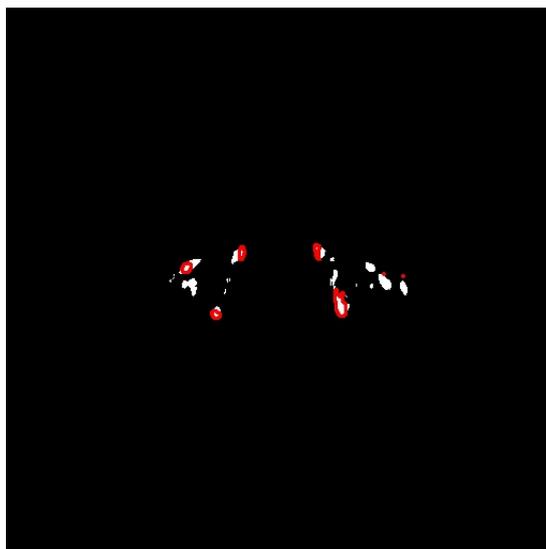


Figura 3.239: Slice n189 relativa al paziente 2 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

Anche le slice n99 e n210 (visibili nelle Figure 3.240 e 3.241) sono rimaste invariate e non sono state influenzate negativamente dalla presenza del sesto paziente. Anche la slice n192 presente nel Validation Set non ha riscontrato alcun beneficio ed è rimasta critica.

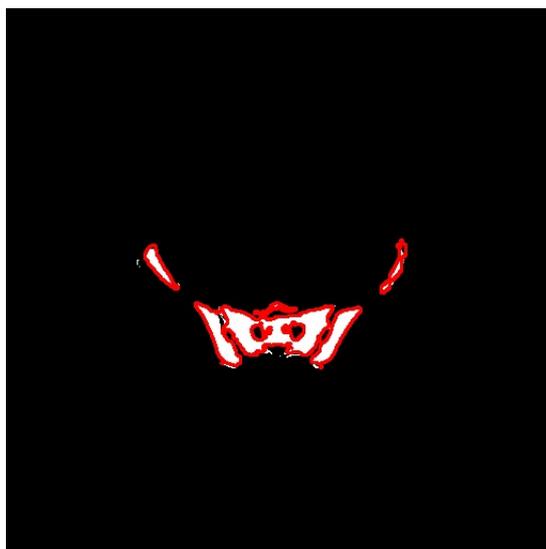


Figura 3.240: Slice n99 relativa al paziente 2 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

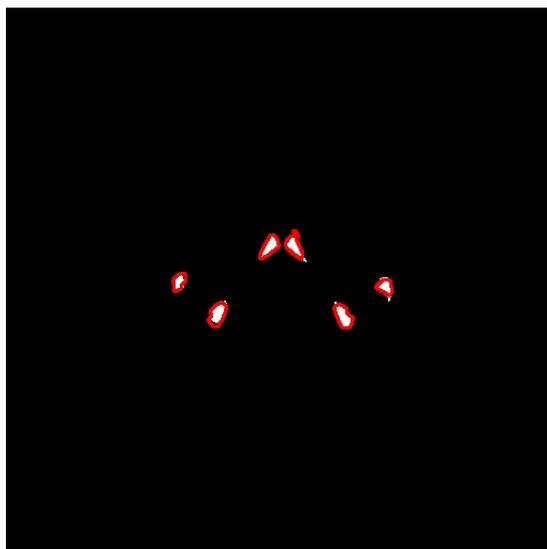


Figura 3.241: Slice n210 relativa al paziente 2 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

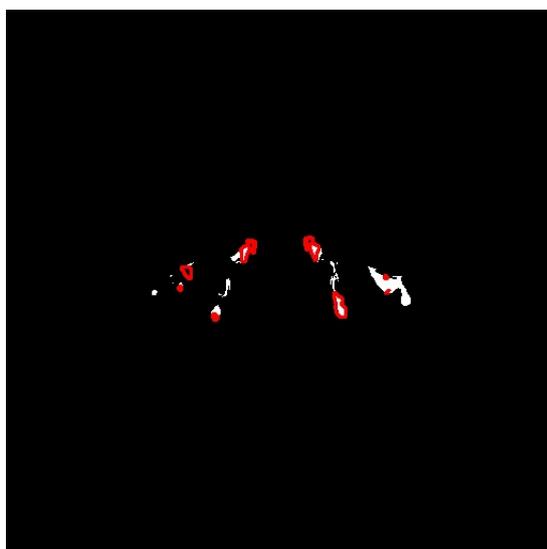


Figura 3.242: Slice n192 relativa al paziente 2 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

La slice n159 in Figura 3.243 è stata precedentemente trattata e aveva riscontrato alcune criticità nelle performance a causa del sotto-campionamento. L'aggiunta del nuovo paziente, questo fenomeno è stato ridotto e il DSC è aumentato da 67.9% a 83%.

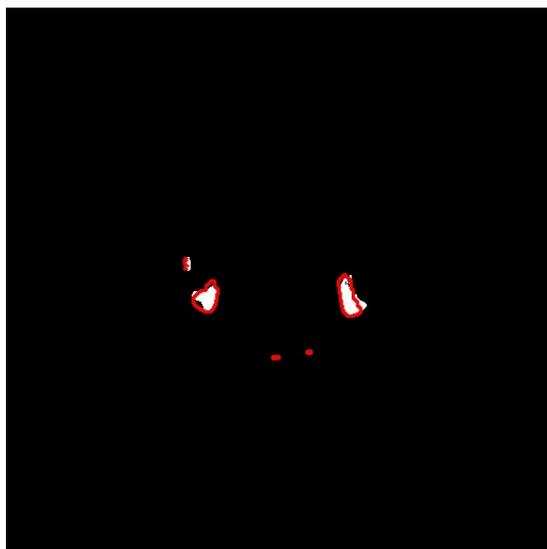


Figura 3.243: Slice n159 relativa al paziente 1 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

Le slice n108 e n63 visibili nelle Figure 3.244 e 3.245 non sono state influenzate negativamente dal sesto paziente ed hanno mantenuto delle performance ottimali e superiori al 90%.

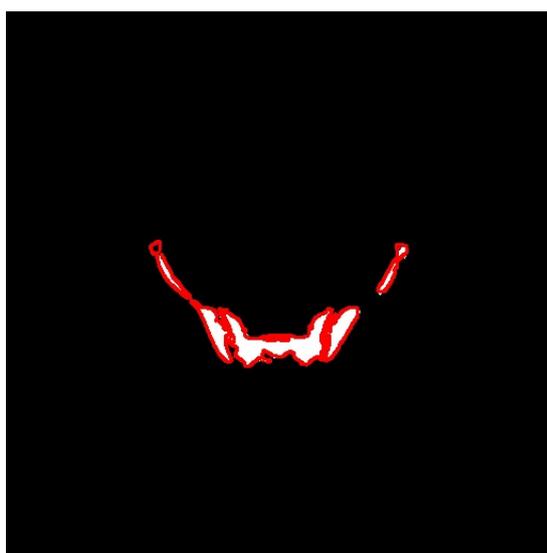


Figura 3.244: Slice n108 relativa al paziente 2 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

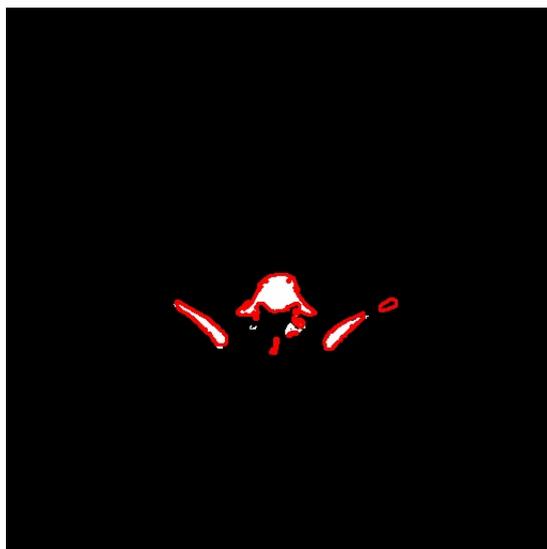


Figura 3.245: Slice n63 relativa al paziente 2 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

3.4.3 Risultati Paziente 3

La slice n202 relativa al paziente 3 e visibile in Figura 3.246, rispetto al caso in cui erano presenti solo 5 pazienti, ha mostrato un evidente miglioramento a seguito dell'aggiunta del sesto paziente. Infatti il DSC è aumentato dal 46% al 86.7% grazie all'attenuamento del fenomeno di sotto-campionamento.

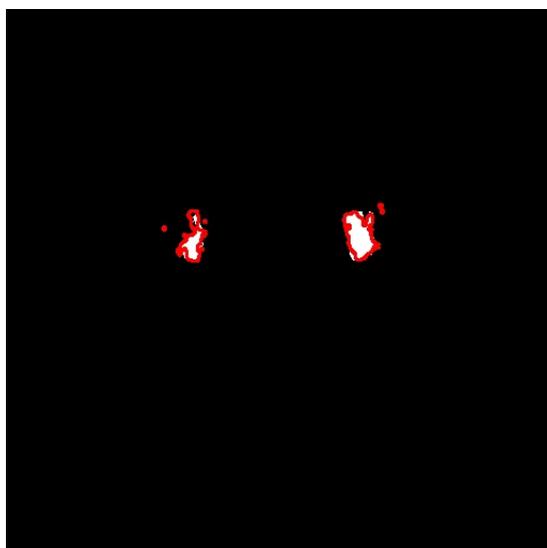


Figura 3.246: Slice n202 relativa al paziente 3 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

Le slice n289 e n286 (Figure 3.247 e 3.248) pur facendo parte rispettivamente di Training Set e Validation Set mostrano che questo tipo di criticità, in cui le zone di RM sono piccole e sparse, non è stato risolto con l'aggiunta del sesto paziente; si ha un lieve aumento del DSC che resta comunque sotto il 50%.

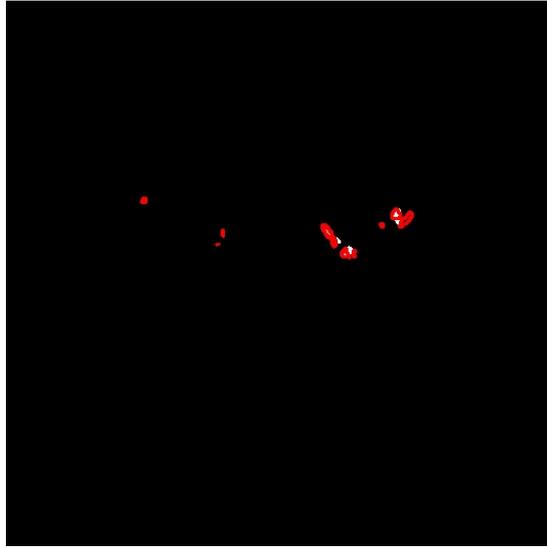


Figura 3.247: Slice n289 relativa al paziente 3 nel Training Set, i contorni rossi indicano la MaskRM effettiva.



Figura 3.248: Slice n286 relativa al paziente 3 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

La slice n85 mostra ancora un sovra-campionamento di RM della parte centrale (Figura 3.249). Infatti, il DSC rimane pressoché uguale ed inferiore al 45%.

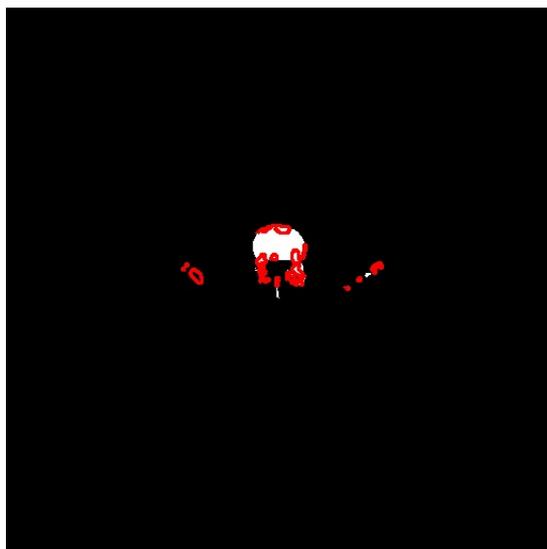


Figura 3.249: Slice n85 relativa al paziente 3 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

Le slice n127 e n136 (Figure 3.250 e 3.251) sono immagini relative alla parte centrale del bacino. Questa sezione è rimasta invariata rispetto l'aggiunta del sesto paziente; le performance rimangono comunque alte e superiori al 85%.

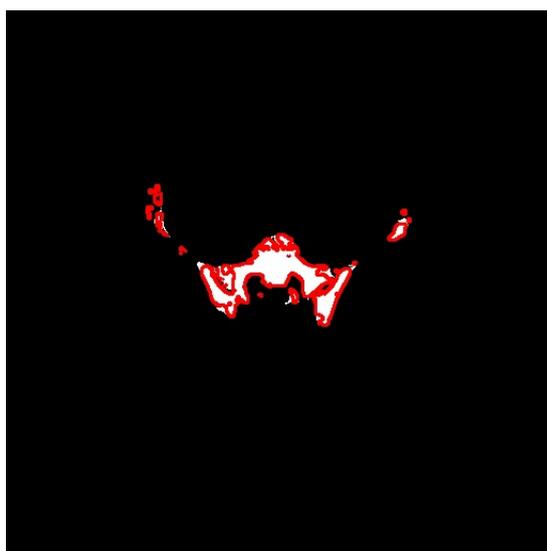


Figura 3.250: Slice n127 relativa al paziente 3 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

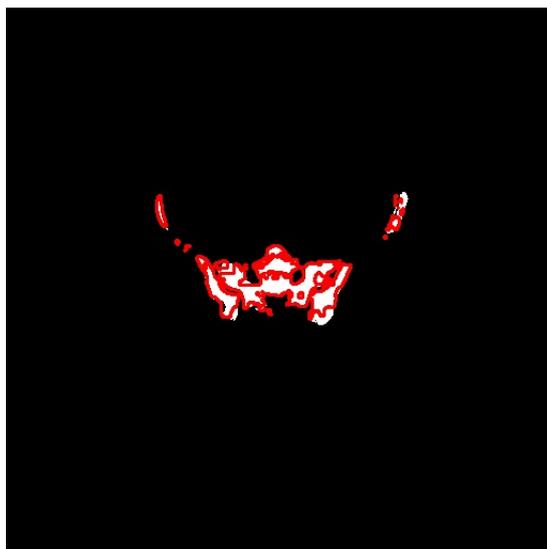


Figura 3.251: Slice n136 relativa al paziente 3 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

3.4.4 Risultati Paziente 4

La slice 128.5 relativa al paziente 4 non ha ottenuto alcun evidente beneficio in seguito all'aggiunta del sesto paziente. Come visibile nella Figura 3.252 a differenza del caso precedente, è presente un lieve fenomeno di sovra-campionamento di zone RM.



Figura 3.252: Slice n128.5 relativa al paziente 4 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

La slice n119.5 (Figura 3.253), invece, mostra un miglioramento da parte della rete nella predizione delle aree di midollo osseo rosso poiché viene attenuato il fenomeno di sotto-campionamento. Infatti, il DSC aumenta da 69.4% a 85.1%.

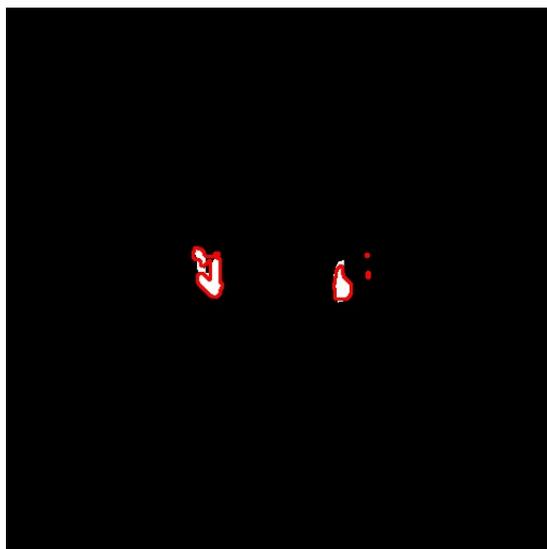


Figura 3.253: Slice n119.5 relativa al paziente 4 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

Sia le slice n8.5 e n83.5 (Figure 3.254 e 3.255) appartenenti al Training Set che le slice n74.5 e n101.5 (Figure 3.257 e 3.258) appartenenti al Validation Set, nel caso con soli 5 paziente aveva già ottenuto ottime performance con DSC superiore al 90%. L'aggiunta del sesto paziente non ha influenzato negativamente queste sezioni anatomiche che continuano ad essere predette in maniera ottimale dalla rete.

La slice n116.5, invece, riesce a beneficiare dell'aggiunta del sesto paziente, infatti la predizione diminuisce il fattore di sotto-campionamento permettendo di ottenere un DSC pari a 82.3%.

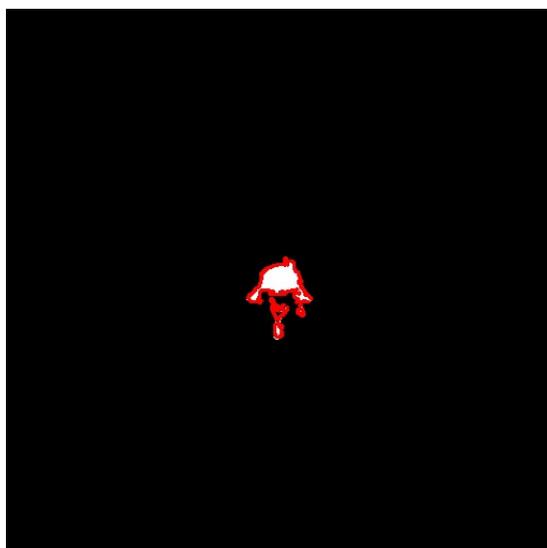


Figura 3.254: Slice n8.5 relativa al paziente 4 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

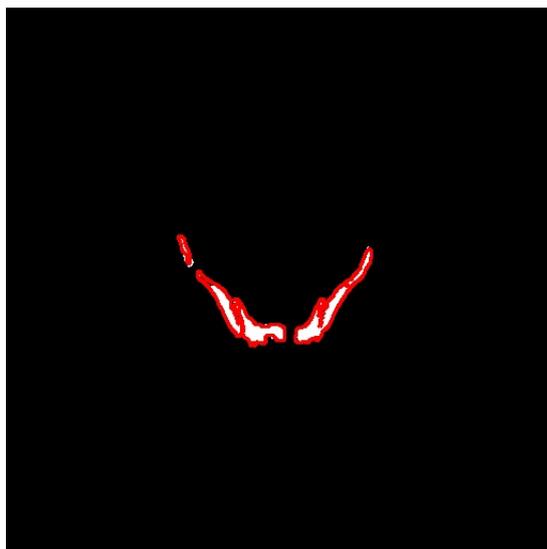


Figura 3.255: Slice n83.5 relativa al paziente 4 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

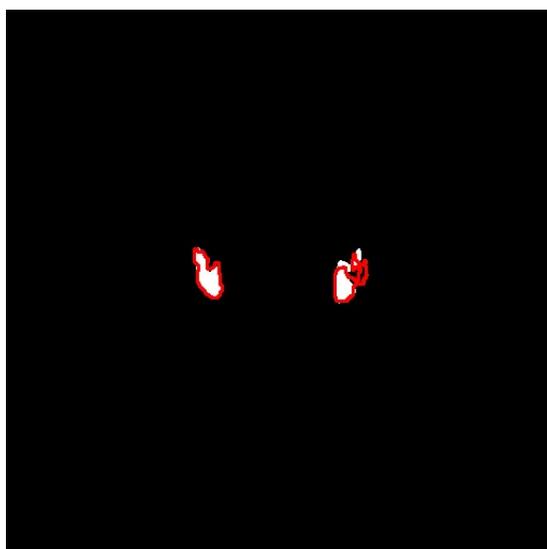


Figura 3.256: Slice n116.5 relativa al paziente 4 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

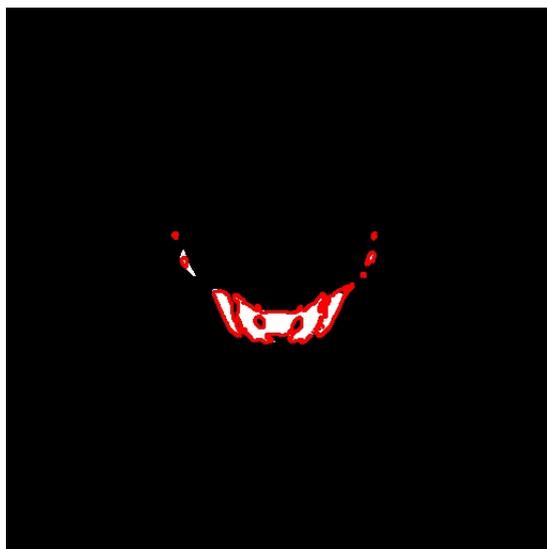


Figura 3.257: Slice n74.5 relativa al paziente 4 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

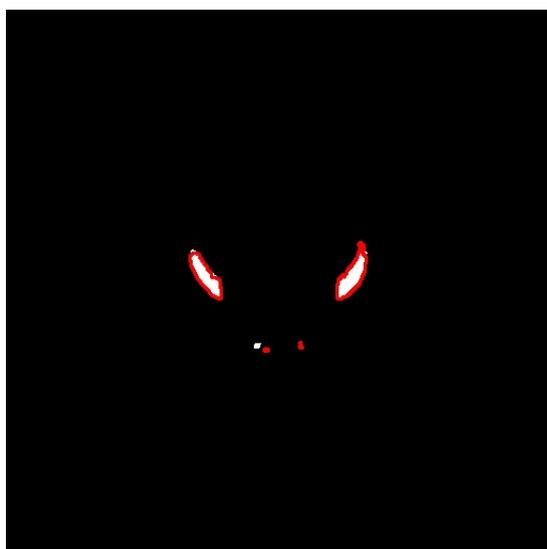


Figura 3.258: Slice n101.5 relativa al paziente 4 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

3.4.5 Risultati Paziente 5

Le maschere relative alle slice n6 e n132 (Figure 3.259 e 3.260, rispetto a quando sono state predette con 5 pazienti, presentano una predizione da parte della rete molto più performante, infatti i valori di DSC sono raddoppiati, arrivando anche a 80% per la slice n132.

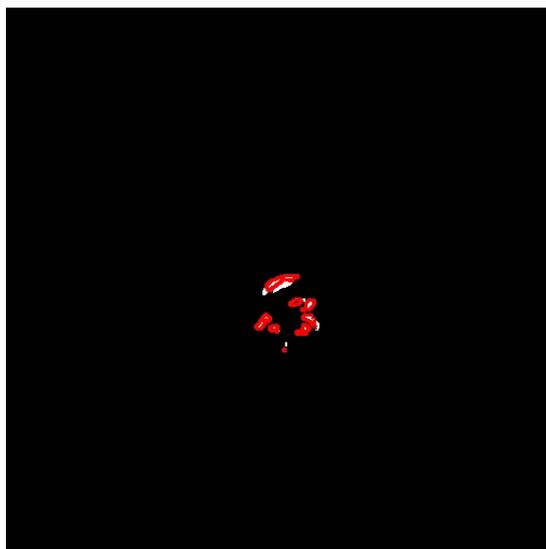


Figura 3.259: Slice n6 relativa al paziente 5 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

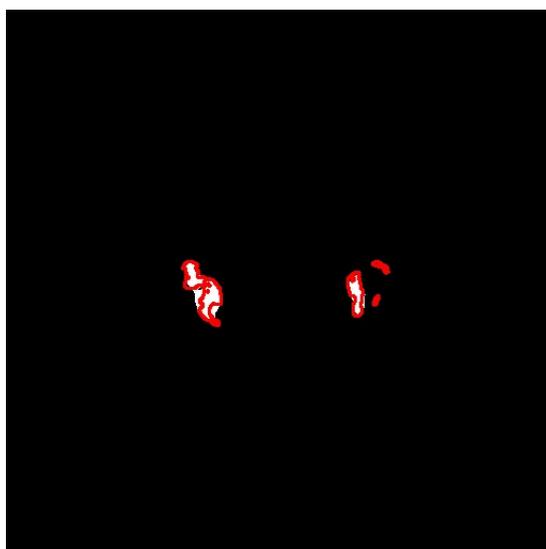


Figura 3.260: Slice n132 relativa al paziente 5 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

Per quanto riguarda le slice n15 e n114 (Figure 3.261 e 3.262), l'aggiunta di un nuovo paziente ha portato ad un leggero miglioramento.

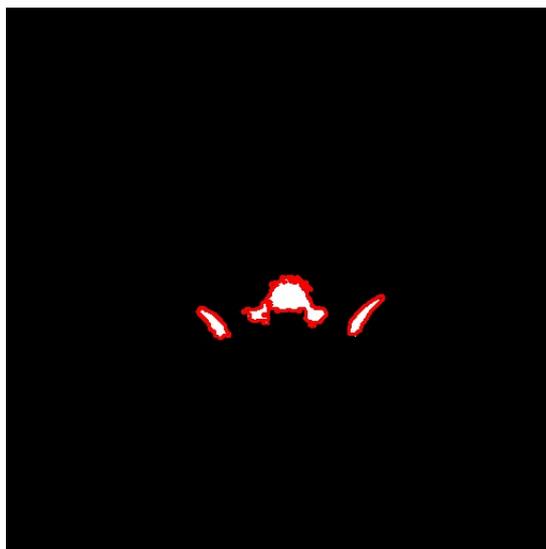


Figura 3.261: Slice n15 relativa al paziente 5 nel Training Set, i contorni rossi indicano la MaskRM effettiva.



Figura 3.262: Slice n114 relativa al paziente 5 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

L'aggiunta di un sesto paziente non ha risolto le criticità nell'identificazione del RM presenti nelle slice n9 e n168 (Figure 3.263 e 3.264), oltretutto è aumentato il fenomeno di sovra-campionamento di RM e conseguente aumento di FP.

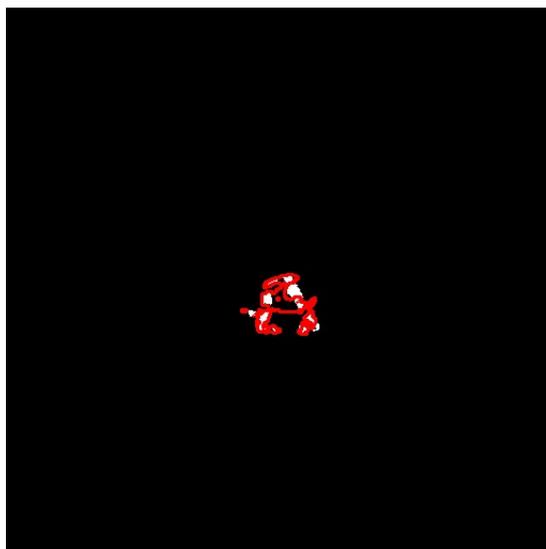


Figura 3.263: Slice n9 relativa al paziente 5 nel Training Set, i contorni rossi indicano la MaskRM effettiva.

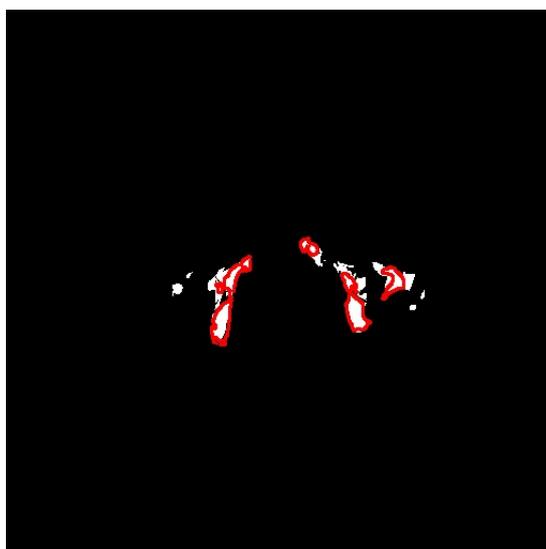


Figura 3.264: Slice n168 relativa al paziente 5 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

Per quanto riguarda le slice n168 e n45 (Figure 3.266 e 3.265), l'aggiunta di un nuovo paziente non ha influito sulle performance ottenute col precedente modello, mantenendo un DSC del 90%.

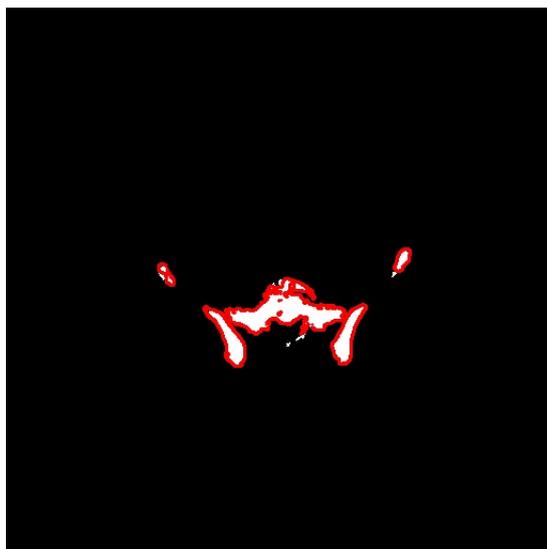


Figura 3.265: Slice n45 relativa al paziente 5 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

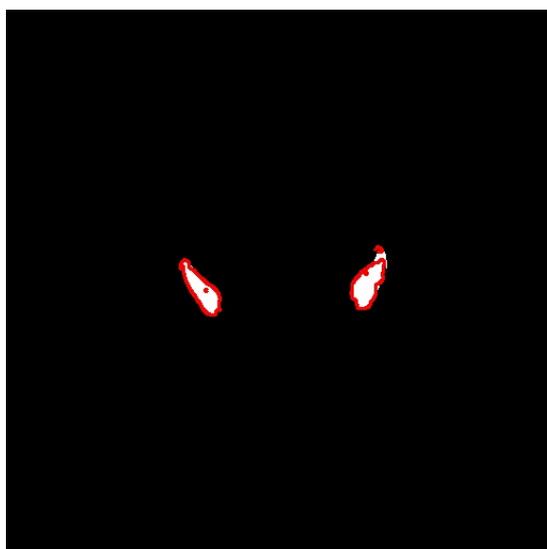


Figura 3.266: Slice n120 relativa al paziente 5 nel Validation Set, i contorni rossi indicano la MaskRM effettiva.

In conclusione, l'aggiunta di un sesto paziente al Data Set non ha influenzato negativamente l'abilità della rete di ottenere predizioni ottimali in specifiche sezioni anatomiche, come ad esempio la parte centrale del bacino e le creste iliache. Infatti, le performance sono sempre rimaste sopra il 90% o invariate. Tuttavia, questo approccio, ha permesso di ridimensionare il fenomeno di sotto-campionamento delle zone RM che affliggeva alcune maschere predette, e conseguentemente aumentando il DSC. Dall'altro lato, sono aumentati i FP nelle maschere con conseguente fenomeno di sovra-campionamento più accentuato in alcune slice.

4. Conclusioni

4.1 Conclusioni

L'obiettivo di questa tesi è l'implementazione di un algoritmo di Deep Learning per l'individuazione di midollo osseo attivo direttamente da immagini TC. Si è deciso di operare sfruttando le reti neurali convoluzionali, più nello specifico seguendo due approcci:

1. Implementazione di un modello di rete neurale convoluzionale semplice;
2. Implementazione del modello U-Net;

Il primo approccio ha mostrato difficoltà nel raggiungimento di questo scopo. Come mostrato nel Capitolo 2.4.1, nel caso in cui il modello prevedeva un solo layer di convoluzione, anche se la rete è stata addestrata per 200 epoche, le performance sono risultate basse rispetto ai precedenti studi [14]; ciò sia nel Training Set che nel Validation Set ed in ambedue le normalizzazioni delle ROI implementate. Infatti, i valori di accuratezza, sensibilità, specificità e precisione nel Training Set hanno una media intorno al 70%, mentre nel Validation Set hanno valore medio intorno al 55%. L'aggiunta di un secondo layer di convoluzione non ha portato ad alcun miglioramento; infatti, seppur le performance nel Training Set possano sembrare promettenti, con valori di accuratezza oltre il 90%, queste non si rispecchiano nei valori reali calcolati sul Validation Set, con una media vicina al 65%. Questo risultato indica come la rete tenda all'over-fitting, come evidenziato nel Capitolo 2.4.2.

L'ultima prova, effettuata utilizzando tre Convolutional Layer, ha mostrato nuovamente questo problema di over-fitting e, inoltre, le performance peggiorano nel Validation Set rispetto al caso con due layer di convoluzione non riuscendo a raggiungere il 60% nel caso in cui la normalizzazione avviene slice per slice. Questo, come già detto nel Capitolo 2.4.3, fa sorgere l'ipotesi che l'aumento della complessità della rete vada ad influenzare negativamente l'apprendimento. Infine, si può concludere che l'implementazione di una semplice rete neurale convoluzionale, ha mostrato difficoltà nell'apprendere dove effettivamente si trovi il midollo osseo ematopoietico.

Il secondo approccio ha visto l'implementazione di una rete neurale convoluzionale più articolata: U-Net. In questo caso si è lavorato direttamente con l'intera slice normalizzata tramite algoritmo di min-max scaling. La rete, è riuscita ad ottenere risultati molto promettenti, permettendo di ottenere valori di DSC, Precisione e Recall superiori al 90% in gran parte delle slice. Le maggiori criticità sono state riscontrate nelle slice in cui si ha una netta predominanza di sfondo nero e l'area di RM è molto ridotta e frastagliata; in questi casi si raggiungeva un valore di DSC molto basso ed intorno al 40%. Inoltre, non sono state considerate le slice che presentavano al più 3-4 pixel di RM.

Dividendo il Data Set in tre sotto-Data Set, si è osservato che la rete non riesce a riconoscere la parte centrale della colonna vertebrale ed inoltre, si ha difficoltà nel riconoscimento delle aree di RM dove la presenza dei femori nelle slice è predominante; ciò si è verificato in alcune slice portando i valori di DSC e Pr a non superare il 60%, mentre la Recall si è mantenuta alta a causa del basso tasso di FN poiché nella classe 0 rientrava pure lo sfondo nero. Uno studio dell'individuazione di aree YM (le cui maschere per allenare la rete sono state ottenute come differenza tra le MaskBM e MaskRM) ha mostrato che queste zone presenti nelle teste femorali sono facilmente individuabili dalla rete, riuscendo a raggiungere valori di DSC, Pr e Re intorno al 90%. La rete lavora molto bene nelle zone di RM molto estese, ad esempio dove è presente il bacino per intero oppure nelle creste iliache, in queste slice, la predizione della MaskRM ha portato a valori molto alti, anche superiori al 95%. Infine, è stato aggiunto un nuovo paziente al Data Set. Confrontando le performance ottenute nel primo Data Set con 5 pazienti con questo nuovo Data Set, si è notata una lieve variazione delle performance. Più nello specifico, come evidenziato nel Capitolo 3.4, ci sono stati casi in cui il sotto-campionamento presente nelle predizioni ottenute con 5 pazienti è stato risolto, permettendo di raggiungere DSC, Pr e Re pari al 80%.

4.2 Sviluppi Futuri

Visti i promettenti risultati della rete convoluzionale U-Net, gli algoritmi di Deep Learning possono essere sicuramente pensati per implementare l'individuazione di midollo osseo attivo direttamente da immagini CT. Tuttavia, è necessario riuscire ad ottimizzare efficacemente la struttura in modo da rendere tali modelli utilizzabili in un workflow clinico a supporto di medici e radiologi.

Possibili sviluppi futuri relativi possono essere ipotizzati. Innanzitutto, allenare la rete per più di 50 epoche; infatti, U-Net estrapola circa otto milioni di parametri e tale numero epoche, risulta essere troppo basso per riuscire ad elaborarli tutti in maniera corretta. In aggiunta, si può pensare all'eliminazione delle teste femorali dalle slice dei pazienti tramite una fase di pre-processing. Queste influenzano troppo negativamente la predizione, inducendo la rete a sovra-campionare poiché questi elementi sono troppo predominanti. Un'ultima ipotesi è quella di considerare separatamente il midollo osseo e lo sfondo. Quest'ultimo, in questo tipo di immagini, è predominante e rischia di mettere in difficoltà la rete. Esistono tool e librerie su Python che permettono di impostare lo sfondo in modo tale da essere ignorato durante l'addestramento; ciò per permettere alla rete di focalizzarsi soltanto sull'oggetto preso in considerazione. In alternativa, si potrebbe anche considerare lo sfondo come una classe a sé stante, cambiando, di conseguenza, l'impostazione del problema poiché non saranno più presenti solo due classi (0 e 1), bensì tre classi: midollo osseo rosso, midollo osso giallo e sfondo.

Bibliografia

- [1] Matthew D. Zeiler, Rob Fergus. «Visualizing and Understanding Convolutional Networks.» In: (2014).
- [2] Vogler 3rd JB, Murphy WA. «Bone marrow imaging.» In: (1988).
- [3] 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.» In: (1999).
- [4] Rose PG, Bundy BN, Watkins EB, et al. «Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer». In: (1999).
- [5] 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.» In: (2000).
- [6] Green JA, Kirwan JM, Tierney JF, et al. «Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix.» In: (2001).
- [7] 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.» In: (2008).
- [8] Mell LK, Tiriyaki 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.» In: (2008).
- [9] J. S. Blebea et al. «Structural and Functional Imaging of Normal Bone Marrow and Evaluation of its Age-Related Changes.» In: (2007).
- [10] Datz FL, Taylor A. «The clinical use of radionuclide bone marrow imaging.» In: (1985).
- [11] John C. Roeske, Anthony Lujan, Richard C. Reba, Bill C. Penney, S. Diane Yamada, Arno J. Mundt. «Incorporation of SPECT bone marrow imaging into intensity modulated whole-pelvic radiation therapy treatment planning for gynecologic malignancies.» In: (2005).
- [12] Everitt SJ, Ball DI., Hicks RJ, et al. «Differential ^{18}F -FDG and ^{18}F -FLT uptake on serial PET/CT imaging before and during definitive chemoradiation for small cell lung cancer.» In: (2014).

- [13] S. Rosati, G. Balestra, P. Franco, C. Fiandra, F. Arcadipane, P. Silveti, U. Ricardi, E. Gallio. «Radiomics for identification of active bone marrow from CT: an exploratory study.» In: (2018).
- [14] Armando De Chiara. «Radiomica per Individuazione di Midollo Osseo Emopoietico da Tomografia Computerizzata». In: (2019).
- [15] G. Sambuceti. «Estimating the whole bone-marrow asset in humans by a computational approach to integrated PET/CT imaging.» In: (2012).
- [16] N.E. Makris et al. «An automatic delineation method for bone marrow absorbed dose estimation in 89-Zr PET/CT studies.» In: (2016).
- [17] P. Leydon et al. «Automatic bone marrow segmentation for PETCT imaging in multiple myeloma.» In: (2016).
- [18] Jurgen Schmidhuber. «Deep Learning in neural networks: An overview». In: (2014).
- [19] Charu C. Aggarwal. «Neural Networks and Deep Learning.» In: (2018).
- [20] Lee, Gobert, Fujita, Hiroshi. «Deep Learning in Medical Image Analysis: Challenges and Applications.» In: (2020).
- [21] P. Meyer, V. Noblet, C. Mazzara and A. Lallement. «Survey on deep learning for radiotherapy.» In: (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.» In: (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.» In: (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.» In: (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.» In: (2017).
- [26] Krizhevsky, A., Sutskever, I., Hinton. «Imagenet classification with deep convolutional neural networks.» In: (2012).
- [27] Girshick, R., Donahue, J., Darrell, T., Malik, J. «Rich feature hierarchies for accurate object detection and semantic segmentation.» In: (2014).
- [28] MurrayS. «An exploratory analysis of multi-class uncertainty approximation in Bayesian convolutional neural networks.» In: (2018).
- [29] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, Thomas Brox, Olaf Ronneberger. «U-Net – Deep Learning for Cell Counting, Detection, and Morphometry.» In: (2019).
- [30] Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. «A Probabilistic U-Net for Segmentation of Ambiguous Images.» In: (2018).
- [31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation.» In: (2015).