

# POLITECNICO DI TORINO

**Corso di Laurea Magistrale  
In Ingegneria Biomedica**

Tesi di Laurea Magistrale

## **Modello di deep learning per la segmentazione automatica di lesioni alla mammella in immagini DCE-MRI**



**Politecnico  
di Torino**

### **Relatori**

Prof. Filippo Molinari  
Prof. Massimo Salvi

### **Candidato**

Federica Nitto

Anno Accademico 2023-2024



# Abstract

Il tumore alla mammella costituisce una delle neoplasie più diffuse, soprattutto nelle giovani donne, e la diagnosi precoce è essenziale per aumentare le probabilità di sopravvivenza. In questo contesto, la risonanza magnetica dinamica con mezzo di contrasto (DCE-MRI) fornisce immagini ad alta risoluzione con una sensibilità generalmente più alta rispetto ad altre modalità di imaging, come la mammografia, in particolare nelle donne con tessuto denso e con un alto rischio di tumore al seno.

Strumenti automatici per l'identificazione e la caratterizzazione delle lesioni possono rappresentare un valido supporto decisionale, migliorando l'efficienza e l'accuratezza nella diagnosi. Automatizzando la segmentazione delle lesioni, il modello può ridurre il tempo dedicato dai senologi all'analisi delle immagini, accelerando il flusso di lavoro e diminuendo la variabilità inter-operatore.

Il lavoro di tesi, in collaborazione con l'azienda Tecnologie Avanzate S.r.l. e l'ospedale Città della Salute e della Scienza di Torino, propone un modello di deep learning per la segmentazione automatica delle lesioni mammarie su immagini DCE-MRI. Il dataset utilizzato è composto da 151 volumi di risonanza magnetica acquisiti con due sistemi diversi, e la segmentazione manuale effettuata da un senologo esperto, utilizzata come ground truth.

L'analisi è stata effettuata sulla terza sequenza post-contrasto. Sono state dapprima applicate tecniche di pre-processing per uniformare il dataset con la medesima risoluzione e dimensione, seguite da una normalizzazione dei valori di intensità dei pixel.

Successivamente, è stato suddiviso in modo bilanciato il dataset in training set, validation set e test set, garantendo che parametri quali il sistema di acquisizione, la dimensione delle lesioni e la morfologia (massa/non massa) fossero equamente rappresentati in ciascun set. È stato allenato un modello di rete CNN di tipo UNet 2D, il cui design a strati permette di catturare dettagli anatomici sia a bassa che ad alta risoluzione.

Sono stati poi ricostruiti i volumi 3D delle lesioni segmentate. A questi sono state applicate operazioni di post processing, tramite operatori morfologici, e rimozione delle componenti non connesse, per ridurre il numero di falsi positivi e migliorare l'accuratezza del modello. Le maschere automatiche sono state infine confrontate con quelle manuali e valutate in termini di Dice Similarity Coefficient (DSC), precisione, accuratezza, recall, Intersect over Union (IOU), Volume Overlap Error (VOE) e True Negative Rate (TNR).

I risultati ottenuti evidenziano la presenza di overfitting, fenomeno per cui il modello di rete presenta ottime prestazioni sui dati di allenamento, ma limitata capacità di generalizzazione sui nuovi dati mai visti, attribuibile a uno sbilanciamento del dataset. Possibili sviluppi potrebbero includere l'ottimizzazione del modello di rete per migliorarne la capacità di generalizzazione, esplorando architetture più complesse e l'utilizzo di tecniche di regolarizzazione.



# Indice

<b>Abstract .....</b>	<b>3</b>
<b>Indice .....</b>	<b>5</b>
<b>Introduzione.....</b>	<b>8</b>
<b>Capitolo 1 .....</b>	<b>11</b>
<b>DCE-MRI: Dynamic contrast-enhanced magnetic resonance imaging.....</b>	<b>11</b>
<b>1.1 Caratteristiche .....</b>	<b>11</b>
<b>1.2 Vantaggi e limiti della sequenza DCE-MRI .....</b>	<b>13</b>
<b>Capitolo 2 .....</b>	<b>14</b>
<b>Stato dell'arte.....</b>	<b>14</b>
<b>2.1 Dal machine learning al deep learning: .....</b>	<b>14</b>
<b>2.2 CNN: Convolutional Neural Networks .....</b>	<b>16</b>
<b>2.2.1 Architettura e layer principali .....</b>	<b>16</b>
<b>2.2.2 Architettura U-Net .....</b>	<b>19</b>
<b>2.2.3 Vantaggi e limiti della rete U-Net .....</b>	<b>21</b>
<b>2.2.4 Apprendimento per trasferimento (Transfer Learning, TL) .....</b>	<b>22</b>
<b>Capitolo 3 .....</b>	<b>24</b>
<b>Materiali e metodi.....</b>	<b>24</b>
<b>3.1 Dataset a disposizione.....</b>	<b>24</b>
<b>3.1.1 Esempio di lesioni mass e non mass.....</b>	<b>25</b>
<b>3.2 Pipeline di lavoro .....</b>	<b>26</b>

<b>3.3 Pre processing</b> .....	<b>28</b>
<b>3.3.1 Allineamento e grayscale transformation</b> .....	<b>28</b>
<b>3.3.2 Ricampionamento e ridimensionamento</b> .....	<b>29</b>
<b>3.3.3 Normalizzazione</b> .....	<b>30</b>
<b>3.4 Divisione dataset</b> .....	<b>33</b>
<b>3.4.1 Esempio di lesioni secondo la dimensione</b> .....	<b>34</b>
<b>3.5 Configurazione modello U-Net</b> .....	<b>36</b>
<b>3.5.1 Backbone e funzioni di loss: prove effettuate</b> .....	<b>37</b>
<b>3.5.2 Ottimizzazione dei parametri</b> .....	<b>39</b>
<b>3.6 Post- processing</b> .....	<b>42</b>
<b>3.6.1 Operatori morfologici</b> .....	<b>43</b>
<b>3.7 Metriche di valutazione</b> .....	<b>44</b>
<b>Risultati</b> .....	<b>46</b>
<b>4.1 Valutazione delle tecniche di normalizzazione</b> .....	<b>47</b>
<b>4.1 Architettura della rete: backbone</b> .....	<b>48</b>
<b>4.2 Parametri di rete: funzioni di loss</b> .....	<b>51</b>
<b>4.3 Volumi 3D: scelta del post-processing migliore</b> .....	<b>53</b>
<b>4.4 Casi critici: Discussione</b> .....	<b>56</b>
<b>Capitolo 5</b> .....	<b>61</b>
<b>Conclusioni e sviluppi futuri</b> .....	<b>61</b>



# Introduzione

Il cancro al seno è una tra le principali cause di morte per tumore tra le donne a livello globale. Nel 2022 sono stati registrati più di 2 milioni di nuovi casi e 670.000 decessi, con un tasso di incidenza crescente al progredire dell'età. [1]

Questa patologia ha origine nei lobuli mammari e nei dotti galattofori con la crescita incontrollata di cellule che mutano e acquisiscono morfologie e funzioni atipiche. Tramite il sistema circolatorio o linfatico, tali cellule possono diffondersi e generare tumori secondari e metastasi. [1]

Diagnosticare precocemente i tumori al seno è determinante per migliorare gli esiti terapeutici e le probabilità di guarigione. Lo screening mammografico è attualmente il metodo principale di prevenzione che, attraverso l'utilizzo di raggi X a basso dosaggio, permette di individuare anche lesioni di dimensioni ridotte. [2] Tuttavia, nelle donne con tessuto mammario molto denso si riscontrano delle difficoltà diagnostiche che possono portare ad un aumento di casi falsi positivi, cioè di tessuti sani che vengono erroneamente identificati come tumorali.

La risonanza magnetica al seno, invece, è una tecnica di analisi di approfondimento di secondo livello che, tramite l'utilizzo di un mezzo di contrasto, fornisce immagini ad alta risoluzione con un'elevata sensibilità. E' molto efficace per identificare piccole lesioni non visibili con altre tecniche di imaging, permettendo anche di effettuare un'analisi più approfondita di aree ritenute sospette. Inoltre, la risonanza magnetica viene utilizzata anche per il follow-up di pazienti con carcinomi già diagnosticati, al fine di valutare il trattamento terapeutico più adatto e monitorare il decorso della malattia. La sequenza dinamica con contrasto (DCE-MRI), in particolare, fornisce informazioni sulla perfusione e vascolarizzazione dei tessuti, garantendo un'accurata distinzione tra tessuti sani e patologici. [3]

In questo contesto, l'applicazione di modelli automatici per l'individuazione e la segmentazione delle lesioni tumori sta emergendo come un ottimo strumento di supporto al radiologo. Avere una segmentazione accurata permette, difatti, di distinguere con precisione i tessuti mammari e analizzarne in dettaglio la morfologia.



Tuttavia, la segmentazione manuale effettuata dai senologi risulta essere un processo lungo e spesso soggetto alla variabilità inter e intra-operatore, cioè alla differenza di risultati ottenuti da due diversi operatori o dallo stesso operatore in momenti diversi della giornata. A questo proposito, tecniche automatiche di deep learning possono dare un contributo significativo riducendo i tempi di analisi e migliorando l'accuratezza diagnostica.

In questa tesi si è sviluppato un modello di rete neurale basato su tecniche di deep learning per la segmentazione automatica di lesioni tumorali presenti in immagini DCE-MRI della mammella. Il dataset è stato fornito dall'ospedale Città della Salute e della Scienza di Torino e il progetto è stato svolto in collaborazione con l'azienda Tecnologie Avanzate S.r.l.



# Capitolo 1

## DCE-MRI: Dynamic contrast-enhanced magnetic resonance imaging

### 1.1 Caratteristiche

La risonanza magnetica dinamica con contrasto (DCE-MRI) è una tecnica di imaging funzionale non invasiva, utilizzata per la sua elevata risoluzione spaziale e temporale che permette di identificare e caratterizzare le lesioni tumorali con elevata accuratezza. Questa tecnica si basa sull'acquisizione sequenziale di immagini sia prima che dopo l'iniezione endovenosa di un mezzo di contrasto paramagnetico, solitamente a base di gadolinio. Le molecole di gadolinio alterano le proprietà molecolari dei tessuti, rendendo le immagini più nitide, tali da evidenziare la presenza di eventuali tumori e infiammazioni, facilitando così la formulazione di una diagnosi più precisa. [4]

L'analisi dell'intensità del segnale di risonanza fornisce informazioni sul comportamento dinamico del mezzo di contrasto, ovvero sul suo assorbimento e rilascio all'interno dei tessuti. Questo genera le cosiddette curve dinamiche di wash-in e di wash-out, che riflettono la perfusione e la permeabilità vascolare della zona di interesse. [5] Dall'analisi di tali curve possono essere estratte informazioni sia morfologiche che funzionali, entrambe preziose per la caratterizzazione dei tessuti in esame.

Le immagini di risonanza magnetica possono essere pesate in T1 o in T2, in funzione dei tempi di rilassamenti dei protoni nei tessuti. Ciascuna modalità fornisce informazioni differenti ma integrative sulla morfologia e funzionalità dei tessuti, risultando così entrambe fondamentali per una corretta interpretazione delle immagini. Come illustrato nella figura 1, le sequenze pesate in T2 mettono in evidenza fluidi, cisti e linfonodi, distinguibili da un'elevata intensità del segnale grazie alla quale è possibile rilevare eventuali infiammazioni dei tessuti. Nelle immagini T1, invece, viene monitorata la distribuzione del mezzo di contrasto nel tempo per identificare tumori, edemi o cambiamenti morfologici del seno. [6]

La somministrazione del mezzo di contrasto provoca un aumento progressivo del segnale con un picco noto come "Maximum Enhancement" (ME), che rappresenta il momento in cui la concentrazione del contrasto nel tessuto è massima. Questo valore è importante per valutare la vascolarizzazione delle lesioni, poiché i tessuti cancerosi mostrano un enhancement rapido del segnale (wash-in) seguito da un wash-out moderato, coerentemente con la loro elevata perfusione.

Le immagini in T1, che presentano una sensibilità maggiore ai dettagli anatomici rispetto alle immagini T2, risultando quindi più valide per visualizzare la vascolarizzare nei tessuti.

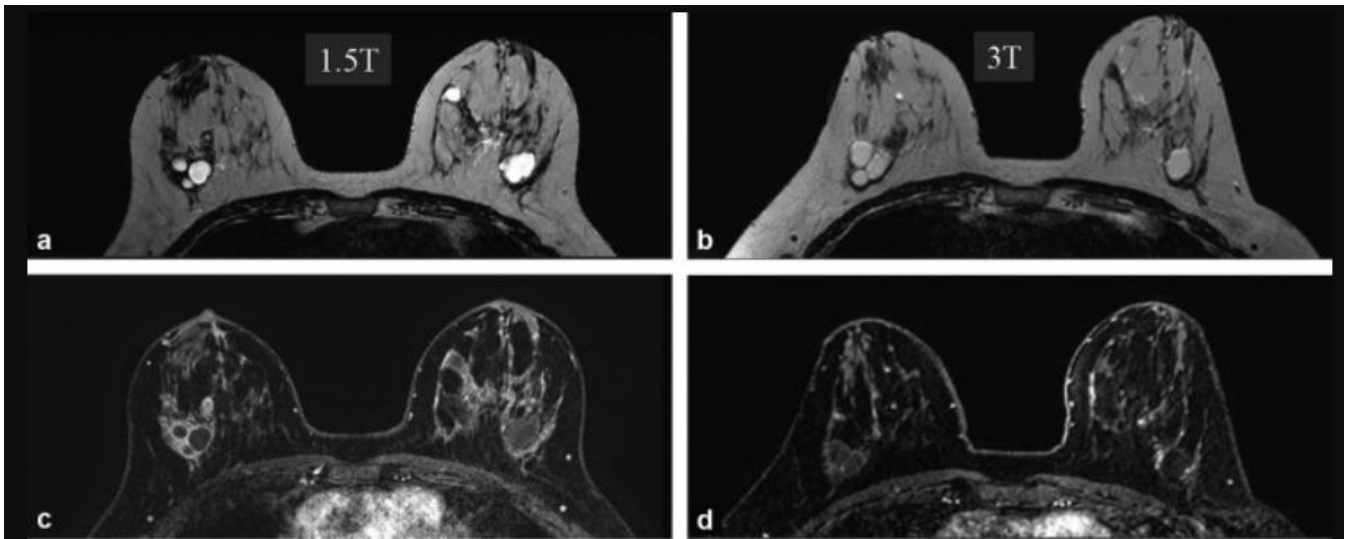


FIGURA 1 - ESEMPIO IMMAGINI MRI DEL SENO PESATE T1 (A E B) E T2 (C E D) [7]

Tra gli altri parametri utilizzati per valutare l'andamento dinamico del mezzo di contrasto, si trova l'area sotto la curva (AUC), che misura la quantità di contrasto che permane nel tessuto nel tempo. L'andamento della curva è illustrato in figura 2: un AUC elevato, in particolare, indica una maggior perfusione, tipica di lesioni altamente vascolarizzate come nei tumori maligni. Il tasso di wash-out (wash-out rate), invece, descrive quanto velocemente il mezzo di contrasto viene eliminato al tessuto; un rate elevato indica una vascolarizzazione instabile o irregolare.

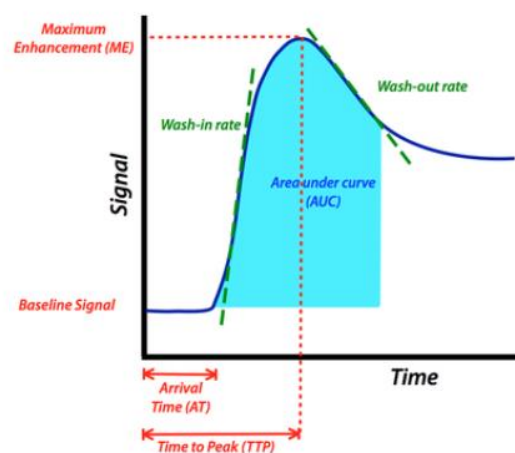


FIGURA 2 – CURVE DI WASH IN E WASH OUT [5]

## 1.2 Vantaggi e limiti della sequenza DCE-MRI

La risonanza magnetica dinamica è uno strumento diagnostico all'avanguardia diffusa per la sua efficacia nel fornire dettagli sul sistema vascolare dei tessuti. [8] Questa caratteristica è utile nella diagnosi precoce di neoplasie, in particolare per l'identificazione e caratterizzazione di lesioni tumorali maligne che, tramite il processo di neoangiogenesi, stimolano la loro proliferazione generando nuovi vasi sanguigni con struttura atipica ed elevata perfusione.

Un altro aspetto rilevante della sequenza DCE-MRI è il suo impiego nel monitoraggio della risposta ai trattamenti antitumorali e postchirurgici. Valutando i cambiamenti nella perfusione tumorale, è possibile interpretare una riduzione dell'assorbimento del contrasto come una risposta positiva al trattamento, sottolineando una riduzione dell'attività tumorale. Al contrario, un aumento della vascolarizzazione potrebbe suggerire una progressione della malattia, fornendo al radiologo dettagli utili per follow-up dei pazienti. [9]

Nonostante i vantaggi citati, la risonanza magnetica dinamica presenta anche delle criticità. Per garantirne un'interpretazione accurata, i dati acquisiti devono essere di alta qualità, poiché informazioni errate possono aumentare il rischio di falsi positivi. La sequenza dinamica di risonanza, infatti, presenta un'elevata sensibilità ma la sua specificità risulta inferiore, con un alto numero di tessuti sani classificati erroneamente come patologici.

L'uso della risonanza magnetica con contrasto al seno, in particolare, è vantaggioso per le donne con un tessuto mammario denso, ma rispetto ad altri metodi di imaging risulta essere più dispendioso e non sempre accessibile, con tempi di acquisizione ed analisi maggiori. [10].

Alla luce di ciò, i modelli d'intelligenza artificiale stanno emergendo come strumenti efficaci per identificare dettagli che spesso sfuggono all'occhio umano, garantendo una maggior specificità e diagnosi più accurate. [11]

# Capitolo 2

## Stato dell'arte

### 2.1 Dal machine learning al deep learning:

Negli ultimi decenni, il deep learning ha registrato una rapida crescita nel campo nella diagnostica per immagini, con notevoli risultati in applicazioni quali segmentazione, classificazione e caratterizzazione.[12]

Prima che queste tecniche avanzate prendessero piede, gli algoritmi di machine learning tradizionale costituivano la base per l'analisi dei dati, utilizzando modelli statistici e matematici per comprendere task specifici e formulare previsioni. Tali algoritmi possono essere suddivisi in tre categorie principali:

- **Supervisionati:** l'addestramento del modello avviene su dati etichettati, in cui ciascun input è associato ad un output già noto. A questi appartengono algoritmi come la regressione logistica, gli alberi decisionali e le macchine a vettori di supporto (SVM), adatti per problemi di classificazione e regressione.
- **Non supervisionati:** l'addestramento si basa su dati non etichettati, con l'obiettivo di identificare pattern specifici senza alcuna informazione predefinita sull'output. Questa categoria trova luogo in analisi volte a raggruppare dati o pazienti secondo criteri specifici e caratteristiche comuni, come nei casi di clustering.
- **Semi supervisionati:** combinano elementi delle prime due categorie, utilizzando un numero limitato di dati etichettati insieme ad un'ampia quantità di dati non etichettati. Questo approccio consente di ridurre i costi estrazione delle caratteristiche e ottimizzare l'utilizzo di informazioni disponibili.

Tuttavia, la dipendenza dall'estrazione manuale di feature costituisce uno dei limiti principali del machine learning. Questo processo richiede l'identificazione di caratteristiche rilevanti di dataset strutturati per ottenere performance elevate. In assenza di dati ben etichettati, questi algoritmi risultano poco efficaci nell'adattamento a compiti complessi e variazione di dati, compromettendo l'efficacia delle prestazioni.

Tali limiti sono stati superati grazie all'adozione di tecniche di deep learning come l'impiego di reti neurali profonde, le quali estraggono automaticamente feature di alto livello da dati grezzi. Differentemente dalle tecniche tradizionali, le reti neurali fanno uso di un'architettura gerarchica per apprendere i dati di input, con un'accuratezza che aumenta all'aumentare della profondità dei layer utilizzati.

Per le immagini mediche vengono utilizzate soprattutto reti convoluzionali (CNN) che imparano dai dati grezzi caratteristiche come bordi e contorni già dai primi strati. Progressivamente, tramite i layer successivi della rete, vengono identificati anche pattern più complessi utili per caratterizzare i tessuti e classificarli. [13]

## 2.2 CNN: Convolutional Neural Networks

### 2.2.1 Architettura e layer principali

Le Convolutional Neural Networks (CNN) sono una classe di reti neurali artificiali progettate per apprendere le principali caratteristiche contenute in grandi volumi di dati complessi, come nel caso delle immagini mediche. Ciò che caratterizza le CNN è l'operazione di convoluzione che applica dei filtri (o kernel) su piccole porzioni delle immagini in analisi, permettendo l'estrazione di diversi tipi di informazione a ciascun livello. In figura 3 viene riportata l'architettura di una rete CNN. Essa è costituita da layer di input, layer nascosti e layer di output, con la particolare aggiunta di tre blocchi: convoluzione, pooling e attivazione. Essi si alternano progressivamente portando la rete ad apprendere feature specifiche riducendo la dimensione dei dati di ingresso, ma mantenendo al contempo le informazioni importanti per il task assegnato.

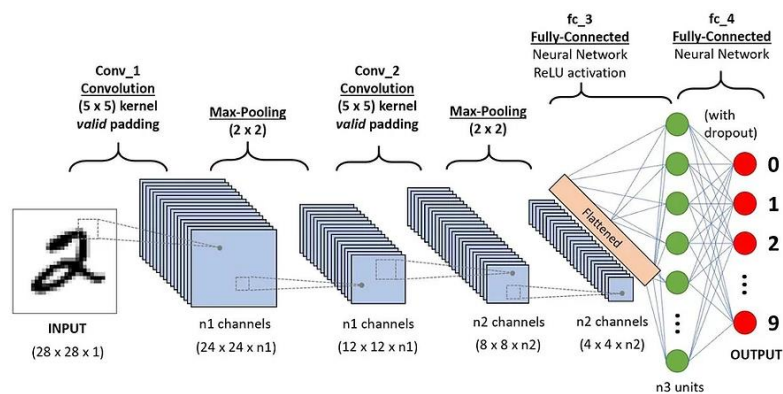


FIGURA 3 - ESEMPIO ARCHITETTURA DI UNA CNN

In particolare, il layer di input struttura le immagini come un tensore tridimensionale di dimensioni (Altezza, Larghezza, n° di canali). Il successivo layer di convoluzione estrae le feature delle immagini attraverso dei filtri (kernel) di dimensioni inferiori a quelle delle immagini originali, che scorrono (sliding window) fornendo in output una feature map.



Come mostrato in figura 4, ogni valore della matrice di origine viene moltiplicato per il corrispondente valore del kernel e il risultato viene sommato per ottenere un singolo valore che costituisce l'elemento centrale della matrice di destinazione.

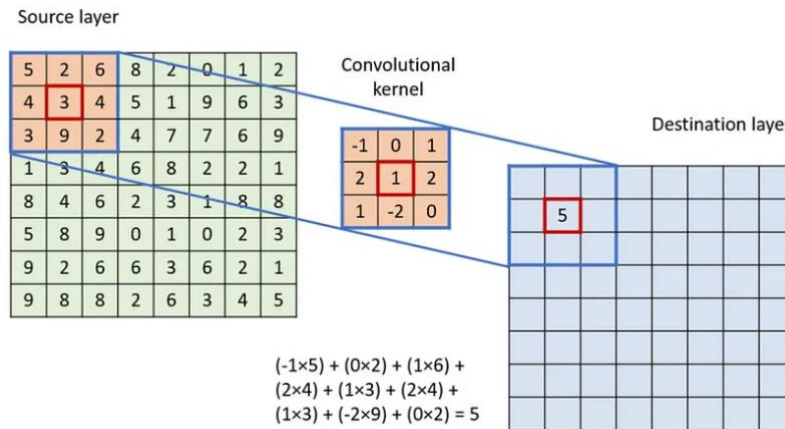


FIGURA 4 - APPLICAZIONE DI UN KERNEL AD UNA MATRICE

La dimensione dell'immagine di output è controllata dalla profondità, cioè dal numero di filtri utilizzati e dallo stride, che rappresenta il numero di pixel di cui il kernel si sposta sull'immagine. Nei casi in cui il kernel supera i bordi dell'immagine, si applica la tecnica dello zero padding: viene aggiunto un bordo di pixel con valore zero fino a quando non si raggiunge la dimensione target (figura 5).

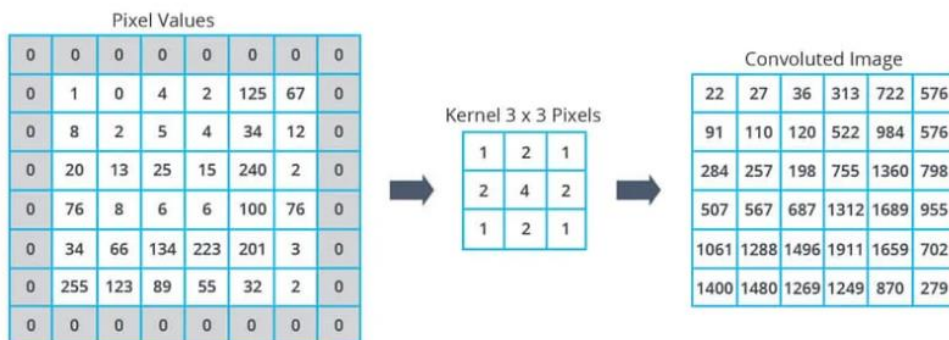


FIGURA 5 - ESEMPIO APPLICAZIONE DELLA TECNICA ZERO PADDING

Il layer di pooling, invece, ha il compito di ridurre le dimensioni spaziali delle immagini di input senza perdere le informazioni più importanti. Questo porta alla generazione di nuove feature map che tengono conto solo delle caratteristiche più rilevanti, eliminando tutto ciò che potrebbe generare errori. Come mostrato in figura 6, il pooling può essere applicato con tre diverse operazioni:

- **Max:** seleziona il valore massimo degli elementi dentro il kernel
- **Average:** calcola la radice quadrata della somma dei quadrati dei pixel dentro il kernel
- **Sum:** somma gli elementi interni al kernel



FIGURA 6 - OPERAZIONI DI POOLING: VALORE MASSIMO, MEDIA E SOMMA

Per l'apprendimento di dati più complessi, vengono utilizzati dei layer di attivazione che introducono una non linearità alla rete neurale. La più comune funzione di attivazione è la ReLu (Rectified Linear Unit) descritta dall'equazione  $f(x) = \max(0, x)$ . Si tratta di una funzione di trasferimento in cui sulle ascisse si ha l'input e sulle ordinate gli output: per valori negativi di input l'uscita sarà pari a zero, mentre per valori positivi l'uscita mantiene invariati i dati di input (figura 7).



FIGURA 7 – APPLICAZIONE DELLA FUNZIONE ReLU AD UNA MATRICE

Il layer di output, infine, prevede degli strati pienamente connessi (fully connected layers) con funzioni di attivazione che variano secondo il tipo di applicazione. Per la segmentazione di immagini mediche, l'output della rete CNN è costituito da una mappa 2D in cui ciascun pixel viene classificato secondo un'etichetta specifica. Nelle reti U-Net utilizzate per segmentare lesioni tumorali, ad esempio, al layer di output viene applicata la funzione di attivazione sigmoide o softmax per restituire per ciascun pixel la probabilità di appartenere o meno alla lesione.

### **2.2.2 Architettura U-Net**

La U-Net è una rete molto utilizzata per la segmentazioni di immagini, in particolare nell'analisi di immagini mediche. Fu nel 2012 con Ciresean et al [14] che si scoprì come poter migliorare la segmentazione di immagini mediche tramite operazioni di convoluzioni e pooling su vasti database ben etichettati. [15]

La U-Net, in particolare, è una Fully Convolutional Network (FCN), una tipologia di rete composta esclusivamente da strati convoluzionali, la cui caratteristica forma ad "U" è costituita da due parti principali : un percorso di contrazione (encoder) e uno di espansione (decoder). [16]

Il path di contrazione è formato da blocchi convoluzionali e operazioni di pooling che riducono progressivamente le dimensioni delle immagini di input estraendone le caratteristiche globali. Nel decoder, invece, sono presenti operazioni di upsampling (deconvoluzioni) che vanno a ripristinare le dimensioni e le risoluzioni delle immagini originali , generando una feature map dettagliata.

In particolare, secondo l'architettura riportata da Ronneberger e Fisher , ogni layer dell'encoder è costituito da due convoluzioni 3x3, una funzione di attivazione ReLu e un'operazione di max-pooling 2x2, che dimezza la risoluzione spaziale dell'immagine ponendo il focus sulle feature più rilevanti. [16]. Tra i due path è presente uno strato intermedio, il bottleneck, un blocco di convoluzione che permette alla rete di apprendere le caratteristiche globali. Nel decoder, infine, le feature map generate vengono sovra-campionate tramite convoluzioni trasposte 2x2 per aumentare la risoluzione spaziale. A queste susseguono le concatenazioni con le feature dell'encoder e due ulteriori convoluzioni 3x3, seguite ciascuna da una funzione ReLu.

Encoder e decoder, inoltre, sono collegati tramite connessioni di tipo skip che copiano le feature maps di alta risoluzione del primo e le concatenano alle feature di bassa risoluzione del decoder. Questo metodo aiuta ad evitare la perdita di informazioni causata dal sotto campionamento, preservando i dettagli rilevanti delle immagini necessari per segmentare lesioni piccole e complesse.

L'ultimo layer della U-Net , infine, è costituito da una convoluzione 1x1 che riduce il numero di canali della mappa di feature finale al numero di classi da segmentare o classificare, producendo così una feature map finale. Nei casi di segmentazione dove si vuole ottenere una maschera binaria (0: sfondo, 1: lesione), il layer di output produce un'immagine con un unico canale. [16]

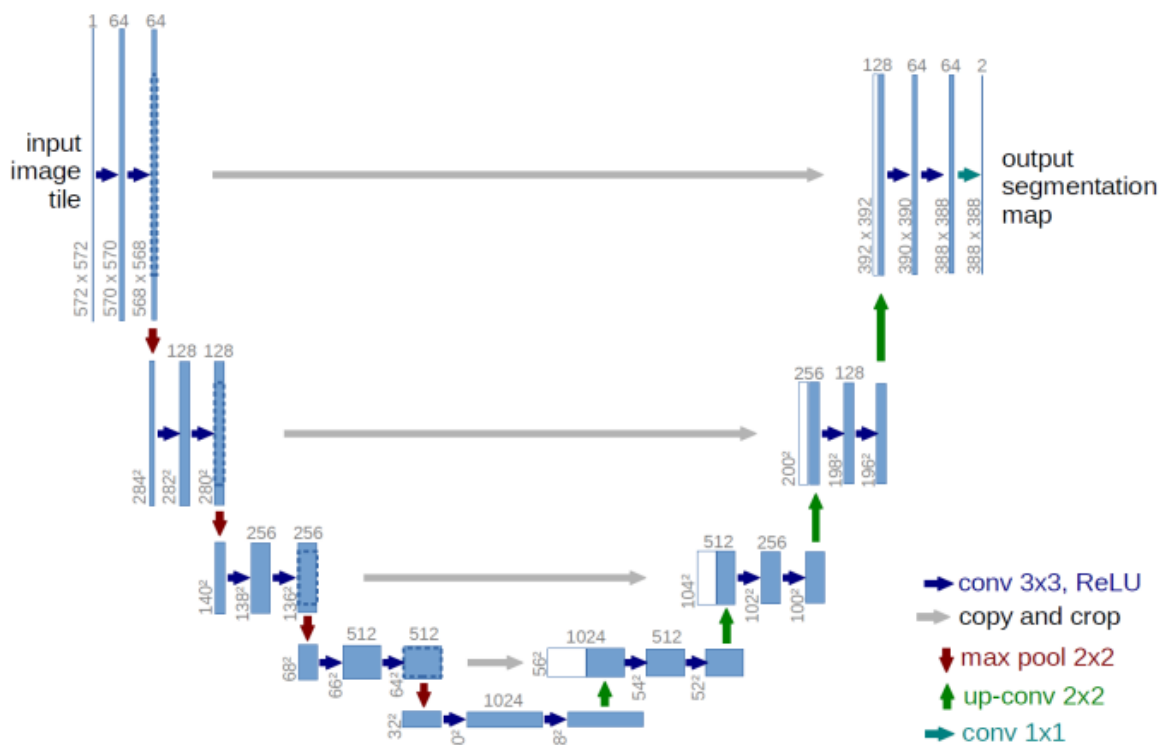


FIGURA 8 - ARCHITETTURA U-NET [16]

### 2.2.3 Vantaggi e limiti della rete U-Net

Uno dei principali vantaggi della rete U-Net risiede nella sua architettura ad U, la quale permette di effettuare una segmentazione dettagliata e precisa anche su un set di dati limitati. Le skip connections utilizzate sono essenziali per ottenere delle prestazioni efficienti, poiché collegano i path di contrazione ed espansione trasferendo feature locali ad alta risoluzione. Ciò è utile per preservare dettagli importanti che permettono di distinguere con precisione i tessuti sani da quelli patologici.

Rispetto ad altre architetture, la U-Net può presentare ottime performance anche con poche epoche di addestramento, risultando così più veloce e prestante. Questa caratteristica rappresenta un vantaggio importante, giacché un numero ridotto di epoche comporta un minor consumo di risorse computazionali.

Un altro aspetto positivo della rete U-Net è la sua robustezza anche in presenza di rumore ed artefatti generati durante l'acquisizione delle immagini mediche. Nonostante ciò, se la qualità dei dati di addestramento è bassa, le prestazioni della rete possono comunque risultare compromesse, rendendo difficile l'individuazione e la segmentazione di lesioni tumorali con morfologie differenti tra loro. Ciò sottolinea l'importanza di avere un dataset di addestramento ben bilanciato.

Infine, limite che si può riscontrare nell'utilizzare delle reti convoluzionali è l'overfitting. La rete, potrebbe adattarsi molto rapidamente ai dati di allenamento a tal punto da non riuscire a generalizzare accuratamente su nuovi casi mai visti. Per evitare l'overfitting, possono essere utilizzate tecniche avanzate come l'applicazione di un early stopping per terminare precocemente l'allenamento o il dropout, il quale rimuove in maniera casuale dei campioni dalla rete durante l'addestramento. [17]

## 2.2.4 Apprendimento per trasferimento (Transfer Learning, TL)

L'apprendimento per trasferimento è una tecnica di machine learning in cui le features apprese precedentemente dal modello di rete su un ampio dataset, vengono sfruttate per affrontare nuovi problemi con dati differenti ma correlati. Questa tecnica è utile nell'imaging medico, per cui non sempre le immagini presentano annotazioni abbastanza complete da poter allenare al meglio reti neurali profonde. [18] Il Transfer Learning, inoltre, riduce i tempi di addestramento ed ottimizza le risorse computazionali, raggiungendo la convergenza con un numero inferiore di epoche rispetto ad una rete neurale che utilizza dei pesi non pre-addestrati. Le principali tecniche sono:

- ***Fine tuning***: una rete pre-addestrata su un dataset generico viene riaddestrata su un dataset specifico di immagini mediche. E' ampiamente utilizzato per task di classificazione e segmentazione.
- ***Feature extraction***: i layer iniziali di una rete pre-addestrata vengono congelati per sfruttare le feature generali estratte, mentre i layer finali del modello vengono riaddestrati su nuovi dati. E' utile per riutilizzare le feature di alto livello estratte dai primi layer.
- ***Domain adaptation***: un modello di rete addestrato su un dataset, viene trasferito su altre immagini con caratteristiche differenti ma correlate. Un esempio è dato dall'adattamento di una rete addestrata su scansioni di risonanza al cervello per segmentare immagini di risonanza al fegato.
- ***Self-Supervised Transfer Learning***: l'apprendimento supervisionato su un vasto dataset non etichettato, viene adattato a un task target con apprendimento supervisionato.
- ***Multitask Transfer Learning***: la rete apprende più task correlati contemporaneamente, aumentando la capacità di generalizzazione. Un esempio sono i modelli per la segmentazione e la classificazione di lesioni tumorali in immagini mammografiche, i quali sfruttano le caratteristiche comuni dei due task.
- ***Generative Transfer Learning***: i modelli generativi (GAN) generano immagini sintetiche con caratteristiche simili a quelle del dataset target, aumentando la quantità di dati disponibili per l'allenamento della rete. E' utile in ambiti in cui si hanno pochi dati a disposizione.

L'approccio di apprendimento trasferito per task di segmentazione può essere impiegato utilizzando il backbone di una rete pre-addestrata su ImageNet, come ResNet o VGG, integrandolo all'encoder di un modello di rete. ImageNet è una vasta base di dati contenente oltre 14 milioni di immagini annotate manualmente ad alta risoluzione [19]. L'encoder sfrutta le conoscenze acquisite per riconoscere le feature di base, mentre il decoder della rete viene addestrato da zero con un numero ridotto di pesi pre-addestrati per adattarsi alle immagini MRI. Successivamente viene applicato il fine tuning, durante il quale si riaddestrano alcuni livelli affinché la rete neurale comprenda le caratteristiche principali delle immagini di risonanza.

# Capitolo 3

## Materiali e metodi

### 3.1 Dataset a disposizione

Il dataset utilizzato è costituito da immagini volumetriche di risonanza magnetica dinamica con mezzo di contrasto di 151 pazienti, fornite dal reparto di senologia dell'ospedale Città della Salute e della Scienza di Torino. Di questi, 97 sono stati esaminati con lo scanner Philips Achieva 1.5 T, mentre i restanti 54 provengono dal Centro Traumatologico Ortopedico (CTO) di Torino dove le immagini sono state acquisite tramite lo scanner GE Signa HDxt 1.5 T.

Insieme alle immagini DCE-MRI sono state fornite le rispettive maschere di segmentazione manuale create da un radiologo senologo sulla terza sequenza post-contrasto. Tali maschere, che presentano solo la lesione principale di maggiori dimensioni, sono state archiviate in formato Nifti (Neuroimaging Informatics Technology Initiative), diversamente dalle immagini originali che sono state archiviate in formato DICOM (Digital Imaging and Communications in Medicine).

Tutte le immagini sono state precedentemente anonimizzate per rispettare le normative sulla protezione dei dati personali e garantire la privacy dei pazienti.

In aggiunta, sono state fornite le annotazioni relative al tipo di lesione, distinguendole in lesioni mass e non-mass. Nel complesso, il dataset si presentava molto eterogeneo visto l'utilizzo di due scanner MRI con differenti parametri di acquisizione, oltre che per la variabilità nella tipologia e morfologia delle lesioni.

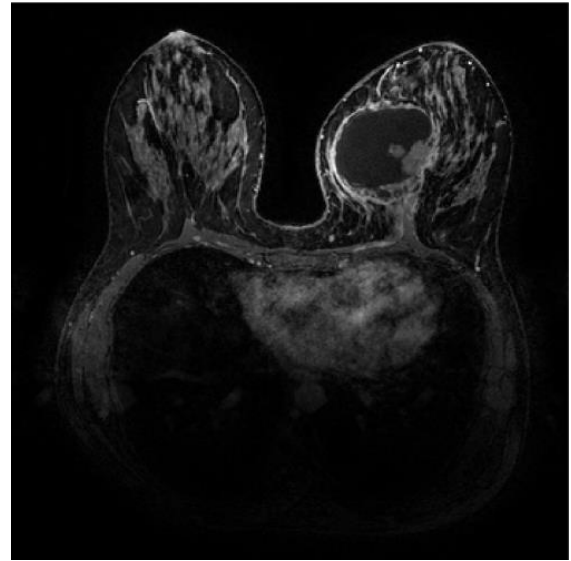


### 3.1.1 Esempio di lesioni mass e non mass

Le lesioni di tipo mass sono caratterizzate da confini e forme definiti, generalmente tondeggianti, in cui la distribuzione del contrasto risulta localizzata e uniforme. Al contrario, le lesioni tumorali di tipo non-mass, si presentano con volumi poco definiti, contorni irregolari e frammentati e mostrano un enhancement del contrasto diffuso. Di seguito nella figura 10 si riportano esempi di lesioni di tipo mass e non-mass nel dataset.



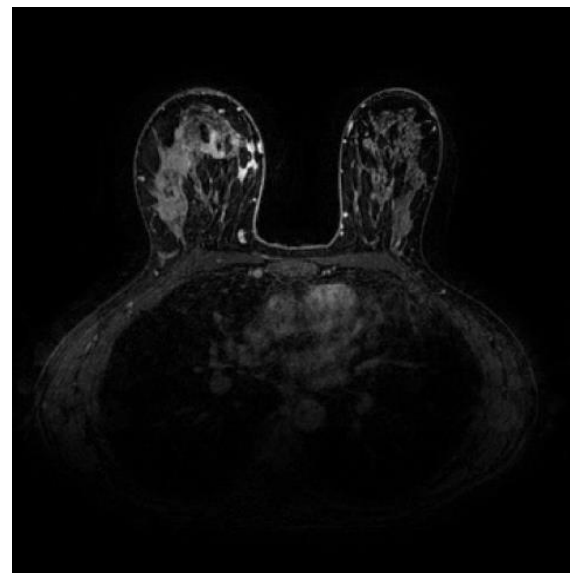
(A)



(B)



(C)



(D)

FIGURA 10 - ESEMPIO IMMAGINI LESIONI DI TIPO MASS (A E B) E LESIONI DI TIPO NON MASS (C E D)

## 3.2 Pipeline di lavoro

L'algoritmo di segmentazione automatico è stato interamente sviluppato in Python utilizzando l'ambiente di sviluppo integrato Spyder ed è costituito da cinque fasi principali:

- Pre-processing
- Preparazione del dataset
- Addestramento e validazione della rete
- Ricostruzione dei volumi 3D e post-processing delle maschere automatiche
- Valutazione dei risultati ottenuti

Dopo aver anonimizzato le immagini di ciascun paziente, è stata fatta un'analisi preliminare per determinare le operazioni di pre-processing necessarie ad uniformare i dati e migliorarne la qualità, al fine di massimizzare l'efficacia dell'addestramento della rete neurale che verrà applicata.

Successivamente, sono state prese in considerazione le informazioni relative all'area delle lesioni, il macchinario MRI utilizzato e la tipologia lesione (mass e non-mass) per suddividere il dataset in training set, validation set e test set.

Per lo sviluppo di questo progetto, sono state utilizzate immagini bidimensionali DCE-MRI. La scelta di un approccio 2D al posto di uno 3D deriva dalle differenze tra le immagini provenienti da due differenti scanner MRI, in particolare per le discrepanze trovate tra i valori di slice thickness e spacing between slices. In particolare, le immagini GE presentavano caratteristiche comuni quali:

- **Slice Thickness** (spessore singola slice): 2.6 mm
- **Spacing Between Slices** (distanza tra le slice): 2.6 mm
- **Rows x Columns** (dimensione dell'immagine): 512x512 pixel

Al contrario, le immagini Philips mostravano una vasta variabilità in tali parametri con combinazioni diverse di risoluzioni e dimensioni. Nonostante i valori comuni di slice thickness di 1.8mm e spacing between slices di 0.9mm, le dimensioni delle immagini variavano da un minimo di 288x288 pixel ad un massimo di 480x480 pixel.

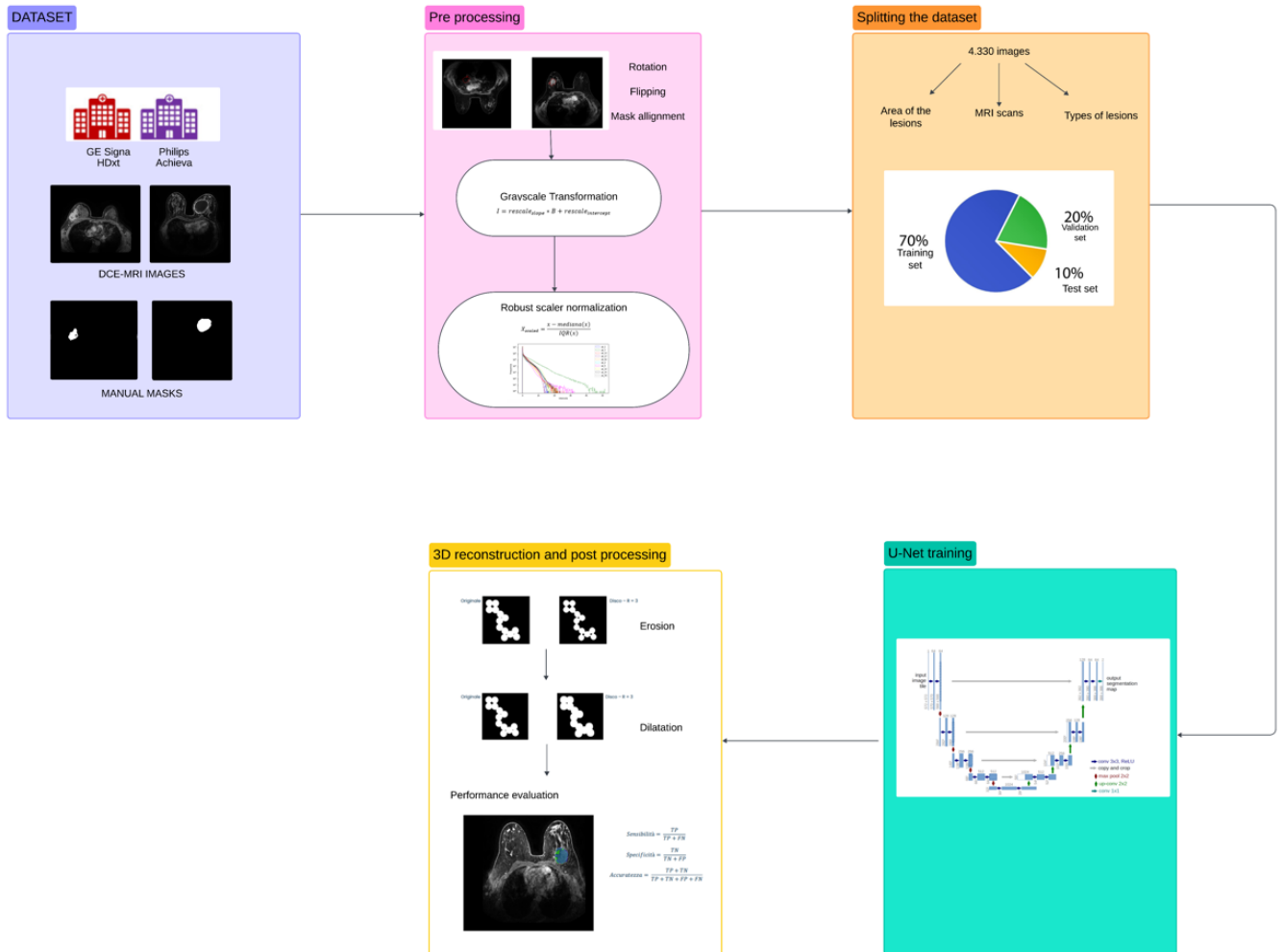


FIGURA 11 - FLOW-CHART PROGETTO

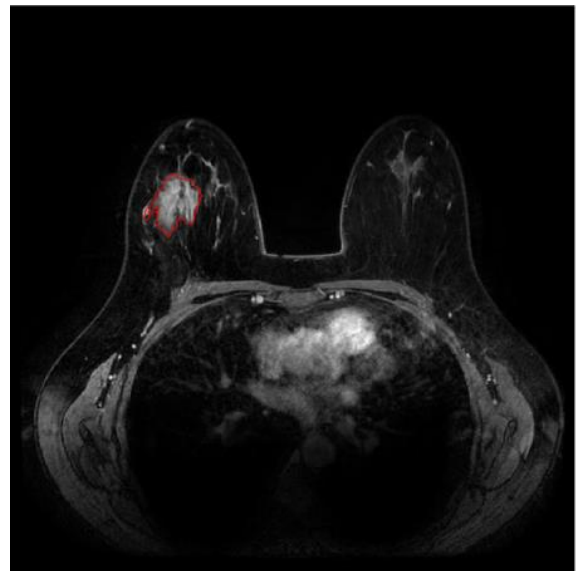
## 3.3 Pre processing

### 3.3.1 Allineamento e grayscale transformation

Nella prima fase di pre-elaborazione, sono state ruotate di  $180^\circ$  le immagini GE che presentavano il seno verso il basso, contrariamente alle immagini Philips. Inoltre, sono state importate le maschere di segmentazione manuale e sono state applicate operazioni geometriche per allinearsi alle rispettive immagini. In particolare, sono state utilizzate le operazioni di flip e rotazione sugli assi x e y tramite la libreria numpy.



(A) Pre-rotazione ed allineamento



(B) Post-rotazione ed allineamento

FIGURA 12 - ESEMPIO IMMAGINE GE: IMMAGINE E MASCHERA SOVRAPPOSTI PRIMA E DOPO L'APPLICAZIONE DELL'ALLINEAMENTO

In seguito, è stata applicata una trasformazione dei livelli di grigio tramite un'operazione lineare che fa uso dei valori di rescale slope e rescale intercept. Essi sono stati ricavati dai metadati DICOM delle immagini. Per le immagini acquisite con lo scanner GE, questi parametri non erano riportati nei metadati; sono stati, quindi, considerati i valori di default pari a 1 per rescale slope e 0 per rescale intercept.

Per ogni pixel è stata applicata la seguente equazione:

$$I = B * rescale_{slope} + rescale_{intercept}$$

dove:

- I: valore reale del pixel
- B: valore grezzo del pixel
- Rescale slope: coefficiente che corregge la scala valori
- Rescale intercept: valore additivo che va a compensare eventuali offset presenti

### 3.3.2 Ricampionamento e ridimensionamento

Per ottenere delle maschere di segmentazione automatica accurate e garantire un'alta qualità nell'addestramento della rete neurale, è necessario assicurare l'uniformità dei dati di input. Affinché il modello di deep learning possa focalizzarsi su pattern specifici e al contempo utilizzare meno risorse computazionali, sono state utilizzate per ciascun paziente solamente le slice positive, cioè le immagini in cui era presente la lesione tumorale.

Poiché le immagini del dataset presentavano poi dimensioni e valori differenti di pixel spacing (distanza fisica tra i centri di due pixel dell'immagine), è stato necessario applicare un'operazione preliminare di ricampionamento, garantendo uniformità dei parametri spaziali. Il ricampionamento è stato applicato con un valore mediano di pixel spacing pari a 0.96mm/pixel.

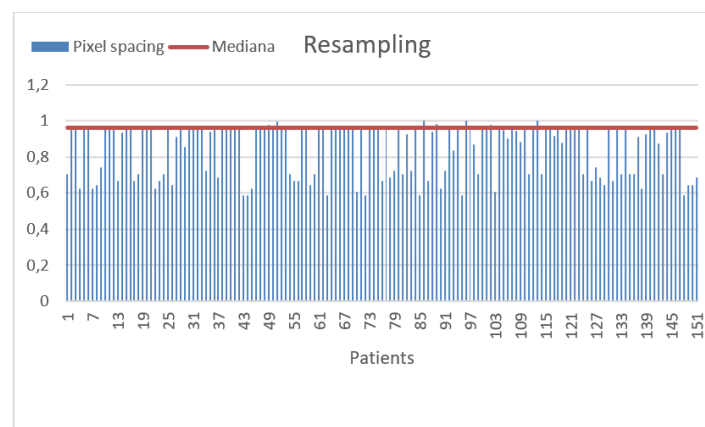


FIGURA 13 - GRAFICO DEL VALORE MEDIANO SCELTO PER IL RICAMPIONAMENTO

Ciò ha consentito di mantenere una corretta proporzione spaziale tra i pixel senza alterare i dettagli rilevanti, che risulta fondamentale in quanto le aree tumorali sono molto piccole rispetto all'intera immagine.

Successivamente, sono state ricalcolate le dimensioni di tutte le immagini ed è stato scelto un valore massimo di 431x431 pixel per dimensionarle ad un valore comune. Oltre alle immagini, il ricampionamento e il ridimensionamento sono stati applicati anche alle segmentazioni manuali, preservando la corrispondenza tra le lesioni presenti nelle immagini e nelle rispettive maschere. Per i dati con dimensioni inferiori a quella target, è stata applicata l'operazione di zero padding, la quale prevede l'aggiunta di pixel con valore zero lungo i bordi dell'immagine fino al raggiungimento della dimensione desiderata senza alterarne il contenuto.

### 3.3.3 Normalizzazione

Poiché la risonanza magnetica produce dati non calibrati, le differenze tra i diversi protocolli di acquisizione influenzano le caratteristiche delle immagini e i valori di intensità dei pixel. Affinché le immagini siano comparabili tra loro, dopo aver uniformato il dataset a livello dimensionale, è necessario anche uniformare il segnale.

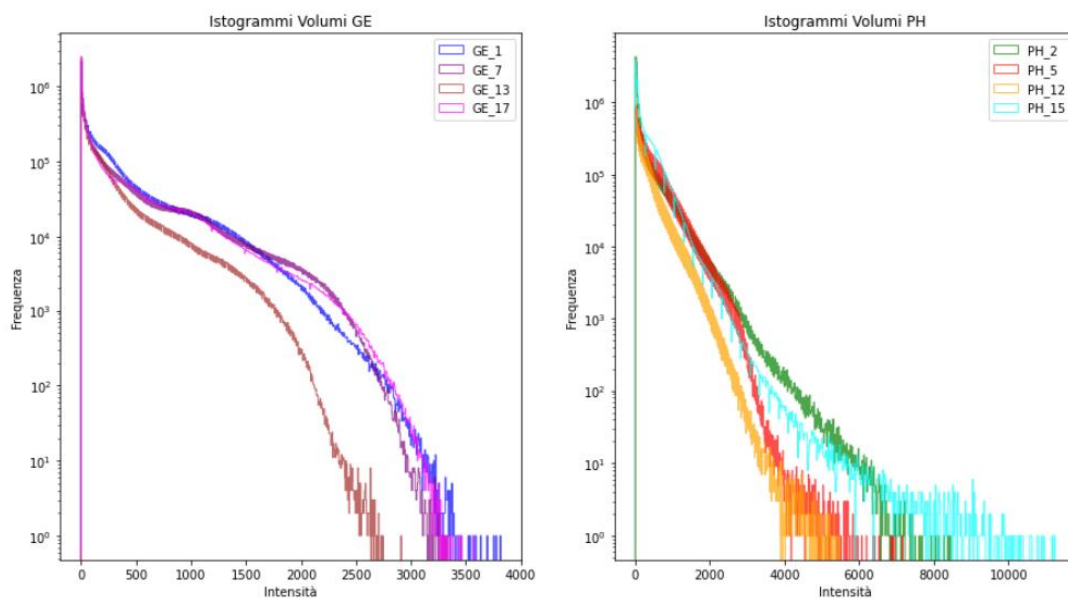


FIGURA 14 - ISTOGRAMMI DEI VOLUMI ORIGINALI DI ALCUNI CASI PHILIPS E GE

Gli scanner Ge e Philips producono immagini con range di intensità notevolmente diversi, come illustrato nella figura 15 che mostra gli istogrammi di un campione ridotto di pazienti. Questa discrepanza potrebbe introdurre bias nei processi di analisi, pertanto è stata ritenuta fondamentale una normalizzazione delle intensità per allineare le immagini su una scala comune. A tale scopo, sono state prese in considerazione e testate tre differenti tecniche di normalizzazione: Min-max Normalization, Z-score Scaling e Robust scaler.

### **Min-max Normalization**

La normalizzazione min-max è una tecnica molto comune per le immagini di risonanza magnetica. Essa riporta l'intensità di ciascun pixel in un intervallo[0,1), utilizzando i valori di intensità minima e massima di ciascuna immagine. La formula utilizzata è:

$$X_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

dove:

- $X_{norm}$ : intensità del pixel normalizzato
- $x$ : intensità originale del pixel
- $\min(x)$  e  $\max(x)$ : valori minimi e massimi di intensità presenti nell'immagine

### **Z-score scaling**

La normalizzazione Z-score utilizza i valori di media delle intensità dell'immagine e di deviazione standard per uniformare i pixel, tali che abbiamo tutti una media pari a zero e una deviazione standard pari ad uno. La formula utilizzata è:

$$X_{norm} = \frac{x - \mu}{\sigma}$$

dove:

- $X_{norm}$ : intensità del pixel normalizzato
- $x$ : intensità originale del pixel
- $\mu$ : media delle intensità nell'immagine
- $\sigma$ : deviazione standard delle intensità

## Robust scaler

La tecnica del robust scaler, a differenza delle precedenti, utilizza i valori di mediana dei pixel e della distanza interquartile (IQR) per normalizzare i dati e riportarli ad una scala di valori comune. La formula utilizzata è:

$$X_{scaled} = \frac{x - \text{mediana}(x)}{IQR(x)}$$

dove:

- $x$ : valore originale del pixel;
- $X_{scaled}$ : valore normalizzato del pixel;
- $\text{mediana}(x)$ : mediana del valore del pixel originale
- $IQR(x)$ : distanza interquartile, cioè la differenza  $Q3-Q1$ , dove  $Q3$  rappresenta il 75° percentile e  $Q1$  il 25° percentile delle intensità dell'immagine.

Dopo aver applicato il processo di normalizzazione, le immagini e le maschere manuali sono state salvate in formato PNG, scelto per la sua capacità di preservare la qualità dell'immagine.



### 3.4 Divisione dataset

Nella costruzione di un modello di deep learning, la corretta suddivisione del dataset in training, validation e test set costituisce un passaggio fondamentale. Durante l'addestramento, la rete neurale comprende le caratteristiche rilevanti delle immagini di input e durante la fase di validazione il set di validation viene utilizzato per ottimizzare gli iperparametri e monitorare le prestazioni del modello. Il gruppo di immagini di test, infine, permette di valutare le prestazioni del modello su dati mai visti, fornendo una misura della capacità di generalizzazione.

Per garantire una buona rappresentatività dei vari scenari clinici presenti nel dataset, è stata posta attenzione a parametri quali: i macchinari di risonanza magnetica utilizzati, distinguendo tra i modelli GE e Philips, l'area e la morfologia della lesione, categorizzandole in mass e non-mass.

In merito alle lesioni da segmentare, è importante tenere in considerazione le dimensioni variabili presenti nel dataset. Le lesioni più piccole e frammentate, in particolare, possono risultare difficili da individuare e segmentare, per cui risulta necessario assicurarsi che durante l'allenamento il modello acceda a dei campioni rappresentativi ciascuna tipologia di lesioni tumorali a disposizione.

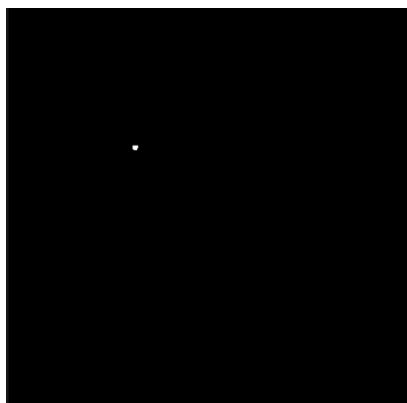
A tal proposito, si sono stabilite delle soglie per suddividere le aree delle lesioni in quattro categorie:

Tipologia	Dimensione (pixel)
Ignorate	<20
Piccole	21-430
Medie	431-1000
Grandi	1001-2000
Molto estese	>2000

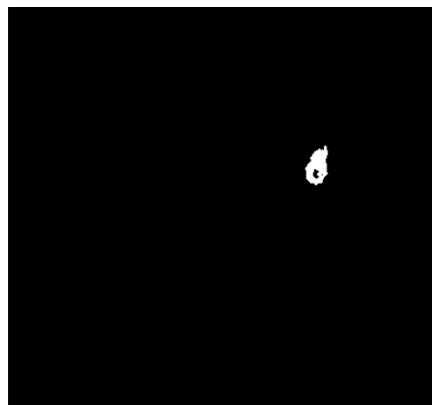
TABELLA 1 - SOGLIE DI CATEGORIZZAZIONE DELLE LESIONI TUMORALI PRESENTI NEL DATASET SECONDO LA DIMENSIONE

Le lesioni con un'area inferiore a 20 pixel non sono state incluse nell'allenamento del modello, poiché ritenute insufficienti per un processo di segmentazione.

### 3.4.1 Esempio di lesioni secondo la dimensione



(a) Lesione piccola



(b) Lesione media



(c) Lesione grande



(d) Lesione molto estesa

FIGURA 15 - ESEMPI DI MASCHERE DI SEGMENTAZIONE MANUALE PER OGNI CATEGORIA DI LESIONE SECONDO LA DIMENSIONE

Il dataset comprendeva complessivamente 4.325 immagini e per ciascuna categoria di lesione è stata stabilita una suddivisione tale che il 70% fosse incluso nel training set, il 20% nel validation set e il 10% nel test set. Questa ripartizione garantisce che ogni tipologia di lesione tumorale, in tutte le sue varianti, sia rappresentata in modo bilanciato in ciascun set.

Nella creazione dei set di allenamento, validazione e test, si è inoltre posta attenzione alla distribuzione dei pazienti, assicurandosi che ciascuno di essi fosse presente esclusivamente in uno dei tre set. Questo previene il fenomeno noto come data leakage, tipico delle applicazioni di machine learning e deep learning che si verifica quando le caratteristiche specifiche di un paziente sono presenti sia nei dati di allenamento che in quelli di validazione e/o test.

Si corre così il rischio che la rete possa apprendere dettagli del paziente, piuttosto che le caratteristiche delle aree patologiche. Una corretta suddivisione dei dati di input, quindi, assicura una stima più affidabile della capacità della rete di generalizzare sui nuovi casi mai visti.

<b>Dimensione</b>	<b>Training set</b>	<b>Validation set</b>	<b>Test set</b>
<b>Piccola</b>	1502	450	224
<b>Media</b>	771	273	226
<b>Grande</b>	421	108	88
<b>Molto estesa</b>	128	48	18
<b>Ignorate</b>	45	48	9

TABELLA 2 - SUDDIVISIONE DELLE IMMAGINI PER DIMENSIONE DELLE LESIONI PER CIASCUN SET

<b>Macchinario-Tipologia</b>	<b>Training set</b>	<b>Validation set</b>	<b>Test set</b>
<b>GE -Mass</b>	250	87	53
<b>GE-Non Mass</b>	287	81	50
<b>Philips-Mass</b>	1425	432	356
<b>Philips-Non Mass</b>	860	279	97

TABELLA 3 - SUDDIVISIONE DELLE IMMAGINI PER MACCHINARIO RM E TIPOLOGIA DELLA LESIONE PER CIASCUN SET

### 3.5 Configurazione modello U-Net

Il modello di rete utilizzato sfrutta un'architettura di tipo UNet, ampiamente utilizzata per le sue prestazioni di segmentazione di lesioni mammarie anche su seni con diverse densità e piccole lesioni. [10]

L'allenamento e la validazione della rete sono state eseguite utilizzando il framework Keras, una libreria open source di deep learning che permette di definire modelli di apprendimento automatico di alto livello. A differenza di altri framework, Keras è risultato più versatile e di semplice utilizzo per testare diversi modelli di rete con differenti combinazioni di parametri.

Per addestrare il modello, sono state caricate le immagini di training e validazione, insieme alle rispettive maschere di segmentazione manuale, e ad esse è stata applicata un'operazione di ridimensionamento.

L'architettura della rete UNet è costituita da operazioni di convoluzione e pooling che riducono in modo progressivo la dimensione spaziale dell'immagine per potenze di due, pertanto i dati di input devono avere una dimensione divisibile per 32, così da mantenere l'uniformità delle immagini tra i vari layer. Le immagini e le maschere manuali sono state quindi ridimensionate da 431x431 a 448x448 pixel, utilizzando la funzione `resize` della libreria `skimage`.

Dopo aver correttamente ridimensionato i dati di input, è stata costruita una matrice 4D per l'addestramento e la validazione della rete, sia per le immagini che per le reference manuali. La matrice ha struttura  $(n,h,w,c)$ , dove  $n$  rappresenta il numero di immagini,  $h$  e  $w$  le dimensioni dell'immagine e  $c$  il numero di canali, che in questo caso è pari a 1 dato che le immagini sono presenti in scala di grigi.

### 3.5.1 Backbone e funzioni di loss: prove effettuate

Le prove effettuate per trovare una configurazione ottimale del modello UNet utilizzato si sono basate sull'implementazione di diversi backbone e funzioni di loss (o funzioni di perdita/costo). Le reti backbone vengono utilizzate come encoder del modello di rete per apprendere ed estrarre le caratteristiche rilevanti dei dati di input tramite operazioni di convoluzioni e pooling. Nello specifico sono stati selezionati i modelli ResNet34 e ResNet50, entrambi con pesi pre-addestrati su ImageNet. Questo metodo, noto come transfer learning (apprendimento trasferito) viene implementato perché aiuta la rete a migliorare la propria capacità di generalizzazione su nuovi dati tramite la conoscenza pregressa delle immagini presenti sul dataset ImageNet, noto per la sua elevata numerosità. Tramite questa tecnica i pesi risultano già addestrati e i tempi di addestramento si riducono notevolmente, migliorando la stabilità alla rete sin dalle prime epoche.

Le architetture ResNet34 e ResNet50 appartengono entrambe al gruppo delle reti ResNet, le quali si distinguono per l'utilizzo di connessioni residue di tipo shortcut. Precisamente queste connessioni sommano l'input di un blocco all'output, migliorando la propagazione dell'errore durante il processo di backpropagation. In tal modo, si facilita il flusso del gradiente attraverso la rete e l'apprendimento di feature complesse a profondità maggiori rispetto ad altre architetture. La principale differenza tra le due reti risiede nella numerosità degli strati presenti, quindi nella profondità della rete. La ResNet34 è formata da 34 strati, quindi rappresenta una rete più veloce e semplice da allenare. D'altra parte, la ResNet50 è più profonda, con un totale di 50 strati e apparentemente più efficace nel apprendere feature complesse, ma richiede maggior costo computazionale e tempi di allenamento con un altrettanto elevato rischio di overfitting.

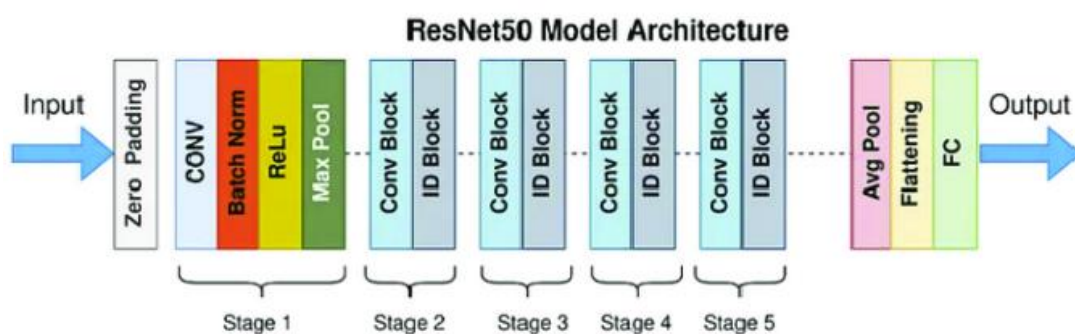


FIGURA 16 - ARCHITETTURA RETE RESNET50

Gli errori commessi dal modello sulle predizioni effettuate vengono calcolati tramite la funzione di loss. Essa viene richiamata dall'algoritmo di ottimizzazione per verificare la differenza tra le maschere predette dal modello e quelle reali, calcolando per ciascun pixel l'errore commesso. Tramite i gradienti durante la fase di backpropagation, i pesi vengono aggiornati e l'errore viene progressivamente minimizzato.

Poiché la funzione di perdita può condizionare l'accuratezza e la capacità di generalizzazione del modello, sono state testate tre diverse funzioni per confrontarne le prestazioni e trovare quella più efficiente per l'obiettivo di questo progetto. In particolare, sono state scelte la binary cross-entropy (BCE), la dice loss e la focal loss.

### Binary Cross-Entropy

La Binary Cross-Entropy (BCE) è una funzione di loss molto comune per task di segmentazione binaria. Viene utilizzata quando sono presenti classi sbilanciate e si vogliono penalizzare le previsioni errate che fa il modello. La formula utilizzata è:

$$BCE = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\check{y}_i) + (1 - y_i) \cdot \log(1 - \check{y}_i)]$$

dove:

- N: totalità dei pixel
- $y_i$ : vero valore target per il pixel  $i$  (0 per lo sfondo e 1 per la lesione)
- $\check{y}_i$ : rappresenta la probabilità predetta dal modello che il pixel  $i$  appartenga alla classe "1"

### Dice Loss

La Dice Loss viene comunemente sfruttata per segmentare immagini dove le lesioni sono molto piccole rispetto all'intera immagine. Sulla base del Dice Coefficient, questa funzione calcola e minimizza la sovrapposizione tra la maschera predetta e il ground truth. La formula utilizzata è:

$$Dice Loss = 1 - \frac{2 \sum_{i=1}^N y_i \check{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \check{y}_i}$$

## Focal loss

La Focal loss, invece, pone maggior attenzione ai pixel difficilmente classificabili, come quelli vicino ai bordi dell'area target. Viene infatti dato un maggior peso ai pixel delle lesioni più difficili da segmentare, riducendo l'influenza di quelli più facilmente predicibili. La formula utilizzata è

$$Focal Loss = -\frac{1}{N} \sum_{i=1}^N (1 - \check{y}_i)^\gamma \cdot y_i \cdot \log(\check{y}_i)$$

Dove il parametro  $\gamma$  regola l'attenzione sui pixel più difficili da classificare: maggiore è il suo valore, maggiore è il peso assegnato ai pixel classificati in modo errato.

## 3.5.2 Ottimizzazione dei parametri

Scegliere la giusta combinazione dei parametri del modello di rete è di notevole importanza nella fase di addestramento, poiché ne determina la capacità di apprendimento della rete. Di seguito vengono descritti i principali iperparametri utilizzati per migliorare le prestazioni del modello UNet adottato e le strategie di data augmentation utilizzate per aumentarne la robustezza.

### Dimensione del batch

La dimensione del batch è un iperparametro che definisce il numero di campioni in cui vengono suddivise le immagini di input e trasmesse al modello di rete prima di calcolare i gradienti e aggiornare i pesi. La scelta della dimensione del batch si basa su efficienza e accuratezza: un batch più piccolo porta ad un aggiornamento frequente dei pesi, un utilizzo di memoria ridotto ma tempi di addestramento maggiori; d'altra parte un batch più grande implica una convergenza più stabile e un'accelerazione dell'addestramento, ma richiede più memoria. Per l'architettura utilizzata si è scelto un valor medio di batch pari a 32.

### Ottimizzatore

Nel corso dell'addestramento, la rete utilizza una funzione di costo per calcolare l'errore tra le previsioni effettuate ed i risultati attesi. In seguito, tramite l'algoritmo di ottimizzazione, il modello modifica i propri pesi per portare questo errore al minimo e migliorare le prestazioni. Nel modello implementato è stato utilizzato l'algoritmo di ottimizzazione Adam (Adaptive Moment Estimation), che regolarizza il tasso di apprendimento per una convergenza più rapida.

In particolare, tale algoritmo calcola la media mobile esponenziale del gradiente  $m_t$  con un fattore di decadimento  $\beta_1$  (generalmente 0.9) per memorizzare il gradiente precedente con la formula:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

dove  $g_t$  rappresenta il gradiente al passo  $t$ .

Successivamente, viene calcolata la media esponenziale dei quadrati dei gradienti  $v_t$  con un fattore di decadimento  $\beta_2$  (generalmente 0.999) con la formula:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

Le due medie  $m_t$  e  $v_t$  vengono poi combinate per calcolare l'incremento di aggiornamento  $\Delta\omega_t$  e aggiornare il peso  $\omega_t$ . [20].

$$\Delta\omega_t = -\eta \frac{v_t}{\sqrt{(s_t + e)}} g_t$$

$$\omega_{t+1} = \omega_t + \Delta\omega_t$$

## Callback

La funzione di callback detta EarlyStopping è stata inserita per interrompere l'allenamento qualora non si presentassero miglioramenti significativi nella metrica di valutazione prescelta. Nel modello adottato, si è deciso di monitorare `val_binary_accuracy`, con un valore di `min_delta` pari a 0.01 e un parametro di `patience` pari a 8. Questo vuol dire che l'allenamento viene interrotto quando l'accuratezza di validazione non mostra un aumento di almeno 0.01 rispetto al valore precedente, registrato per un periodo di 8 epoche consecutive. Combinando l'early stopping con un numero limitato di epoche, si incrementa l'efficacia dell'addestramento del modello, inducendo l'arresto nel momento ottimale. Un numero troppo grande di epoche potrebbe compromettere la capacità di generalizzazione del modello sui dati di validazione e test, per questo stato scelto un limite di 20 epoche.



## Data augmentation

La tecnica di data augmentation viene utilizzata per migliorare la capacità di generalizzazione del modello automatico, attraverso l'applicazione di trasformazioni casuali alle immagini di training. Il dataset viene così ampliato generando virtualmente nuove immagini, quindi nuovi esempi da cui imparare senza inserire manualmente nuovi dati. Comunemente vengono applicate trasformazioni come rotazioni, traslazioni, zoom e variazioni di luminosità, per permettere alla rete di riconoscere le caratteristiche rilevanti delle immagini.

Per implementare il data augmentation, si è scelto di utilizzare un generatore di dati tramite la classe *ImageDataGenerator* di Keras. Questa classe carica e trasforma le immagini in tempo reale durante l'allenamento della rete, portando ad una riduzione del carico sulla memoria e migliorando l'efficienza dell'allenamento. In particolare, il generator combina le immagini e le maschere di segmentazione per poi restituirle in batch tramite la funzione `flow`.

In questo progetto è stata testata una sola operazione di data augmentation: la variazione di luminosità nell'intervallo da 0.8 a 1.2. Data la presenza di macchinari diversi di risonanza magnetica con diversi range di intensità dei pixel, questa tecnica è stata scelta per simulare variazioni di contrasto e intensità. Non sono state aggiunte ulteriori trasformazioni per porre maggior attenzione sugli elementi rilevanti per la segmentazione, senza complicare ulteriormente l'allenamento della rete.

Inoltre, poiché la trasformazione di data augmentation è randomica, è stato impostato un training ripetibile tramite un valore `seed` pari a 1, garantendo che la stessa trasformazione venga applicata coerentemente ad ogni epoca di addestramento.

### 3.6 Post- processing

Per migliorare la qualità delle maschere automatiche ottenute e confrontarle accuratamente con le maschere manuali, è importante valutare i risultati ottenuti sul piano tridimensionale. Nel caso di lesioni tumorali su immagini mediche, risulta infatti necessario comprendere chiaramente la distribuzione spaziale delle lesioni nell'intero volume.

In particolare, per ciascun paziente del test set, sono stati salvati sia i volumi originali delle immagini che le rispettive maschere manuali in formato .nrrd (Nearly Raw Raster Data) tramite la libreria SimpleITK.

Dopo averle ordinate secondo il numero di slice e ridimensionate per mantenere le dimensioni originali, anche le maschere automatiche sono state salvate in formato .nrrd, così da poterle confrontare con le maschere manuali assicurandosi la medesima risoluzione.

Successivamente, per migliorare la bontà delle segmentazioni automatiche riducendo il numero di falsi positivi presenti, è stato analizzato ogni volume segmentato automaticamente dalla rete tramite due funzioni della libreria scikit-image: labels e regionprops. La funzione label identifica le singole regioni connesse all'interno di ogni volume, attribuendo a ciascuna un'etichetta. Una volta etichettate le lesioni, la funzione regionprops calcola per ognuna le caratteristiche geometriche come area, volume, perimetro, bounding box e assi principali. L'estensione della lesione è stata considerata la caratteristica più rilevante, in quanto fornisce informazioni utili per comprendere la distribuzione spaziale della lesione.

Tramite l'analisi di ciascun volume è stato possibile identificare e rimuovere le aree erroneamente segmentate tramite due criteri di selezione:

- **Distribuzione volumetrica:** generalmente una lesione tumorale si sviluppa su più piani di volume, pertanto sono state rimosse le componenti presenti su un unico piano.
- **Area della lesione:** tenendo conto della morfologia delle lesioni tumorali segmentate manualmente, è stata definita una soglia di 80 pixel per l'area della lesione. Tutte le componenti con un'area inferiore a tale soglia sono state rimosse.

### 3.6.1 Operatori morfologici

E' stata analizzata la morfologia e la continuità delle regioni segmentate e si è scelto di applicare tecniche di post-processing per rendere le lesioni più fedeli alle maschere manuali. In particolare, sono stati utilizzati due operatori: erosione e dilatazione. L'erosione si applica ad un'immagine binaria per affinare le dimensioni delle regioni segmentate rimuovendo i pixel attorno ai bordi ad esse. Ciascun pixel viene sostituito con il valore minimo dell'area locale definita da un elemento strutturale che, in questo progetto, è stato rappresentato da un disco. Questa operazione aiuta ad eliminare pixel isolati più piccoli della regione di interesse e non del tutto connesse, segmentate per errore a causa probabilmente di rumore o artefatti presenti nell'immagine.

La dilatazione, contrariamente, espande le regioni segmentate, aggiungendo pixel lungo i suoi bordi della lesione individuata. Questo operatore sostituisce ciascun pixel con il valor massimo presente nell'area locale definita dall'elemento strutturale. Ciò è utile per unire aree non connesse per errore e colmare spazi vuoti tra regioni connesse.

In aggiunta, la dilatazione consente di ripristinare aree rimosse erroneamente durante la fase di erosione, ottenendo regioni continue coerenti con le lesioni tumorali di interesse. Poiché l'ordine di applicazione degli operatori morfologici può influenzare i risultati finali, sono state effettuate delle prove per determinare quale fosse la sequenza ottimale. Applicare un'operazione di apertura, cioè un'erosione seguita da una dilatazione, aiuta a rimuovere efficacemente piccoli oggetti e perfezionare i bordi delle lesioni più grandi, rischiando però di perdere alcuni dettagli. L'operazione opposta di chiusura va invece a riempire gli spazi vuoti delle regioni segmentate.

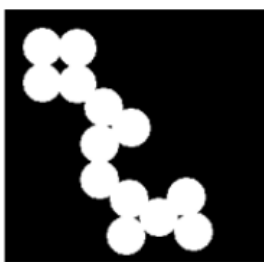
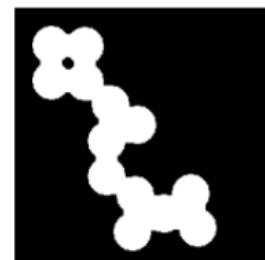


IMMAGINE ORIGINALE



(A) APERTURA



(B) CHIUSURA

FIGURA 17 - ESEMPIO DI APPLICAZIONE DELLE OPERAZIONI DI APERTURA (A) E CHIUSURA (B)

### 3.7 Metriche di valutazione

Per monitorare e valutare le prestazioni del modello nella previsione della maschera di segmentazione automatica durante la fase di allenamento, è stata utilizzata la binary accuracy. Questa metrica misura la qualità delle predizioni e fornisce una misura di stima circa l'accuratezza della segmentazione prodotta. La binary accuracy viene calcolata come:

$$\text{Binary Accuracy} = \frac{\text{Numero di pixel classificati correttamente}}{\text{Numero totale di pixel}}$$

Per ottenere un'analisi completa dell'efficienza e della robustezza della rete in fase di validazione e testing, sono state utilizzate le seguenti metriche:

- Dice similarity coefficient (DSC): misura la sovrapposizione tra la maschera predetta (A) e quella reale (B). Il valore varia tra 0 (nessuna sovrapposizione) a 1 (sovrapposizione perfetta).

$$DSC = \frac{2 * |A \cap B|}{|A| + |B|}$$

- Precision: misura la proporzione tra i pixel correttamente classificati come positivi (TP) sulla totalità delle predizioni positive del modello (TP+FP).

$$\text{Precision: } \frac{TP}{TP + FP} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

- Recall (Sensibilità): misura la proporzione tra i pixel correttamente classificati come positivi (TP) sulla totalità dei pixel realmente positivi (TP+FN).

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- F1-score: è la media armonica di Precision e Recall nei casi di squilibrio tra le classi.

$$F1 - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Intersection over Union (IoU): misura l'area di sovrapposizione tra la maschera predetta e il ground truth, cioè la maschera manuale che rappresenta la verità.

$$IOU = \frac{|A \cap B|}{|A \cup B|}$$

- Specificity: misura la proporzione tra i pixel correttamente classificati come negativi (TN) sulla totalità dei pixel realmente negativi (TP+FN)

$$Specificity = \frac{TN}{TN + FP} = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

- Balanced Accuracy: misura la media tra Specificity e recall.

$$Balanced\ Accuracy = \frac{Recall + Specificity}{2} = \frac{\left(\frac{TN}{TP + FN} + \frac{TN}{TN + FP}\right)}{2}$$

Per valutare la bontà delle maschere segmentate in termini volumetrici, sono state utilizzate altre due metriche specifiche:

- Volume Overlap Error (VOE): misura l'errore di sovrapposizione tra la maschera automatica e la reference manuale.

$$VOE = 1 - \frac{|A \cap B|}{|A \cup B|}$$

- True Negative Rate (TNR): misura il numero di voxel classificati correttamente come negativi, rispetto al totale dei voxel realmente negativi.

$$TNR = \frac{TN}{TN + FP}$$

# Capitolo 4

## Risultati

In questo capitolo vengono presentati ed analizzati i risultati ottenuti in seguito all'applicazione dei metodi descritti precedentemente e delle prove sperimentali effettuate. L'analisi mira a valutare la robustezza del modello di rete proposto e la qualità delle segmentazioni automatiche generate, confrontandole con le maschere manuali di riferimento.

In particolare, vengono mostrati gli effetti delle tecniche di normalizzazione testate in relazione alle caratteristiche del dataset, giustificando su tale base la scelta finale della tecnica considerata più adeguata. Successivamente, vengono valutate le metriche ottenute dalle diverse configurazioni di rete sperimentate, correlate da immagini di esempio che supportano le decisioni effettuate relativamente all'architettura e ai parametri del modello scelto.

Infine, vengono attenzionate le metriche di valutazione ottenute sui volumi tridimensionali ricostruiti ed elaborati, i quali permettono di avere una visione approfondita e completa delle prestazioni del modello implementato. In particolare, si mettono a confronto le tecniche di post-processing testate, identificando l'approccio più adatto per il progetto. Ad integrazione dei risultati quantitativi, vengono presentati e descritti anche alcuni casi rappresentativi, la cui analisi critica mette in luce i punti di forza e limiti di questo progetto.

## 4.1 Valutazione delle tecniche di normalizzazione

Per valutare quale fosse la tecnica di normalizzazione più adatta per il dataset utilizzato e l'obiettivo di segmentazione, sono stati generati gli istogrammi relativi a un gruppo ridotto di casi GE e Philips.

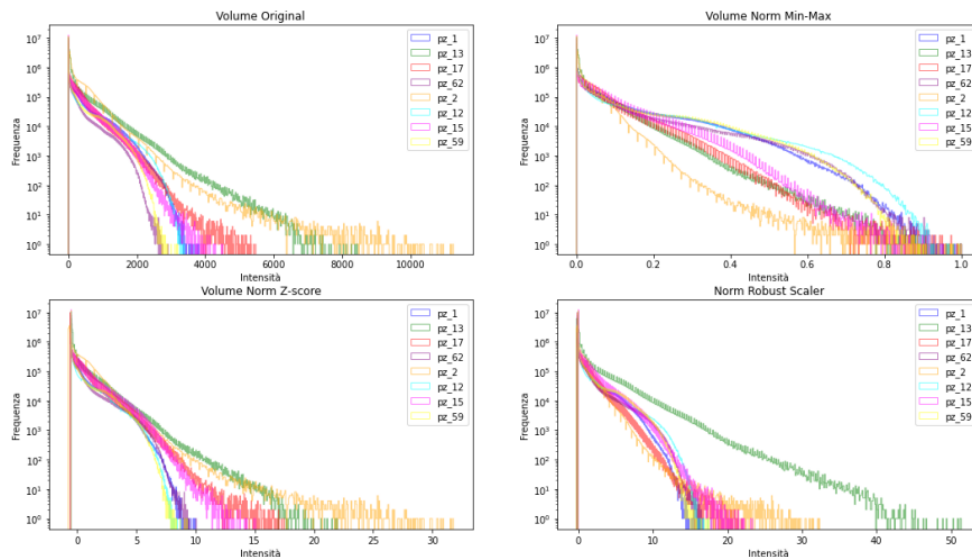


FIGURA 18 - APPLICAZIONE DELLE TECNICHE DI NORMALIZZAZIONE SU UN GRUPPO DI CASI PHILIPS E GE

Data la presenza di differenti range di intensità tra i due scanner, la tecnica del Min-Max scaling è stata scartata. Con questo metodo la distribuzione dei dati si presenta poco allineata: i valori acquisiti dal sistema GE, ad esempio, potrebbero non avere una diretta corrispondenza con quelli ottenuti dal sistema Philips, sebbene facciano riferimento alla stessa area di interesse. Questo rappresenta un limite significativo della tecnica di min-max scaling poiché una distribuzione non allineata delle intensità dei pixel porta ad un'errata interpretazione dei dati, compromettendo le successive analisi. Per quanto riguarda le normalizzazioni Z-Score e Robust Scaler, entrambe mostrano una distribuzione più conforme. Tuttavia, si è scelto di applicare la normalizzazione robust scaler, in quanto ritenuta più adatta all'obiettivo di segmentazione. Le immagini di risonanza magnetica, difatti, sono spesso influenzate da valori anomali o outliers) dovuti a rumore o artefatti generati durante l'acquisizione, che possono variare tra protocolli di acquisizione differenti. La tecnica z-score non è stata considerata adeguata, poiché è molto sensibile a valori estremi che potrebbero causare distorsione dei dati. Differentemente, la robust scaler si è rilevato un metodo meno affetto da tali distorsioni, tramite l'utilizzo della mediana che garantisce una misura più stabile rispetto alla media. L'intervallo interquartile IQR, inoltre, permette di gestire più facilmente le distribuzioni non simmetriche preservando la coerenza tra i dati.

## 4.1 Architettura della rete: backbone

Al fine di trovare la giusta combinazione di architettura e parametri del modello di rete neurale, sono stati inizialmente testati due backbone della tipologia ResNet, in particolare ResNet34 e ResNet50, descritti nel capitolo precedente, utilizzando la funzione di loss Binary Cross-Entropy. Le metriche di valutazione sono state calcolate nei tre set di training, validation e test, per valutare la capacità di generalizzazione dei modelli sul piano 2D.

Metriche	ResNet34			ResNet50		
	Training	Validation	Test	Training	Validation	Test
Dice coefficient	0.8229	0.5111	0.6408	0.7630	0.4930	0.5957
Precision	0.9085	0.7317	0.7836	0.8851	0.6968	0.7640
Recall	0.7675	0.4438	0.5986	0.6770	0.4420	0.5421
F1-score	0.8229	0.5111	0.6408	0.7630	0.4930	0.5957
IOU	0.7225	0.3963	0.5237	0.6597	0.3889	0.4805
Specificity	0.9998	0.9997	0.9996	0.9998	0.9997	0.9997
Balanced Accuracy	0.8837	0.7218	0.7992	0.8484	0.7208	0.7709

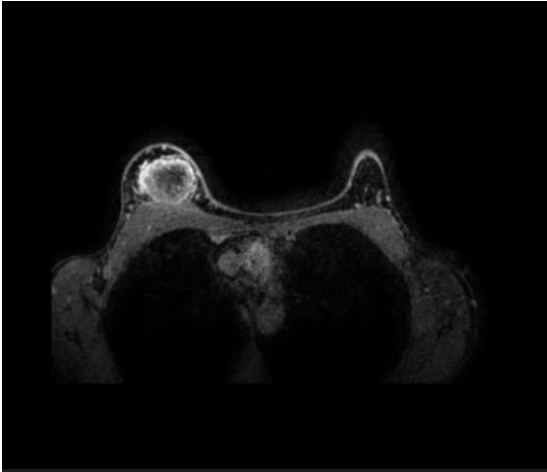
TABELLA 4 - CONFRONTO METRICHE DI VALUTAZIONE DEI BACKBONE RESNET34 E RESNET50

La ResNet34 si è dimostrata essere più prestante nella maggior parte delle metriche, con ottimi risultati durante l'allenamento e una discreta capacità di generalizzazione. Nel set di training vengono infatti ottenuto un Dice coefficient di 0.8229 con una precisione e una specificità pari a 0.9085 e 0.9998. Tali valori mostrano l'alta efficienza del modello nel saper distinguere tessuto sano da tessuto patologico. Nonostante le prestazioni siano diminuite nei casi di validazione, il modello riesce a individuare e segmentare le lesioni di interesse mantenendo un'alta specificità circa costante nei tre set. La Balanced accuracy finale sul test di test pari a 0.7992 dimostra un buon bilanciamento di sensibilità e specificità, quindi un ridotto numero di falsi negativi e falsi positivi. La ResNet50, invece, mostra delle performance inferiori, con un Dice Coefficient e una precisione rispettivamente di 0.7630 e 0.8851 durante la fase di allenamento, che si riducono a 0.5957 e 0.7640 nel test set. Nonostante la specificità e l'accuratezza siano simili a quelle raggiunte dalla ResNet34, questa tipologia di backbone più profonda non ha portato migliorie nelle prestazioni del modello.

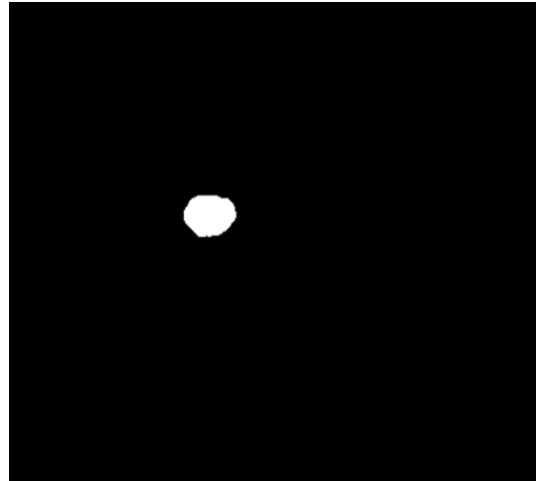


Di seguito vengono riportati esempi di maschere automatiche generate dalle due architetture testate, con le rispettive maschere manuali a confronto.

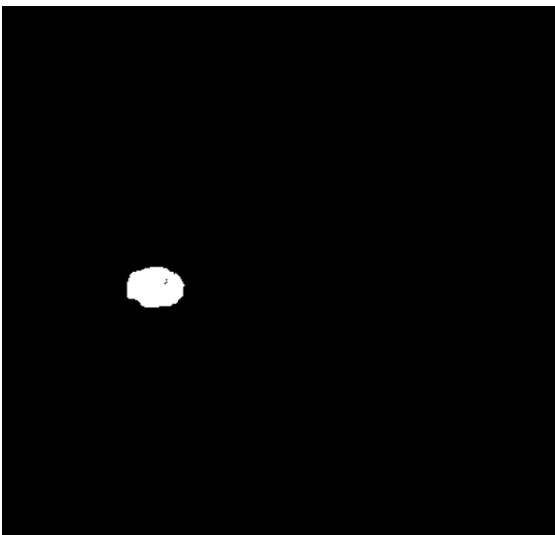
Immagine originale



Maschera manuale



Maschera automatica ResNet34



Maschera automatica ResNet50

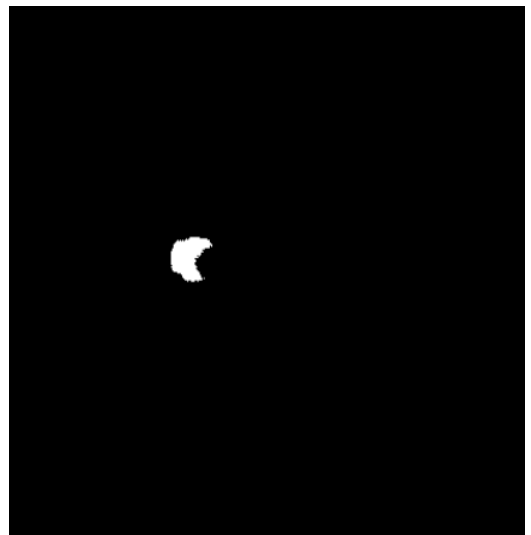
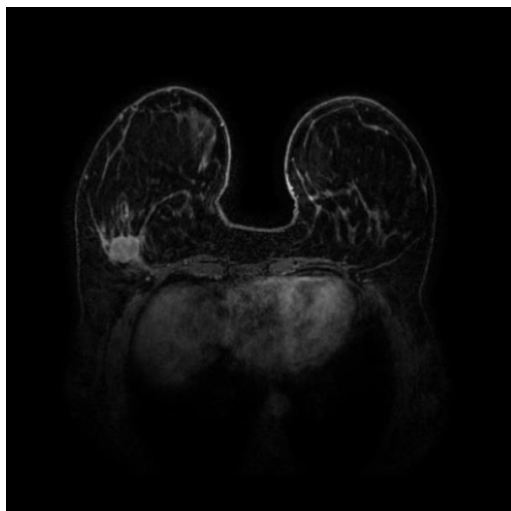
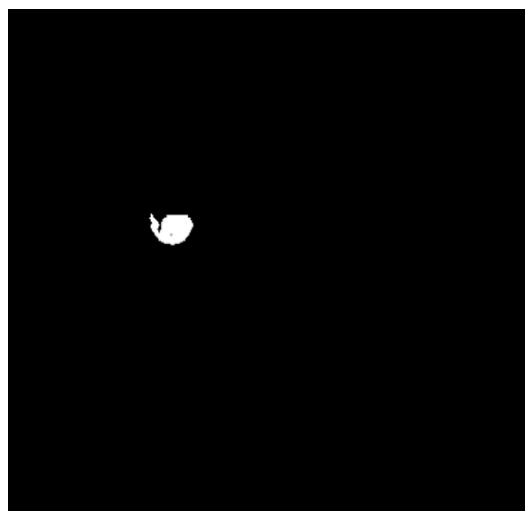


FIGURA 19 - CONFRONTO MASCHERE AUTOMATICHE RESNET34 E RESNET50

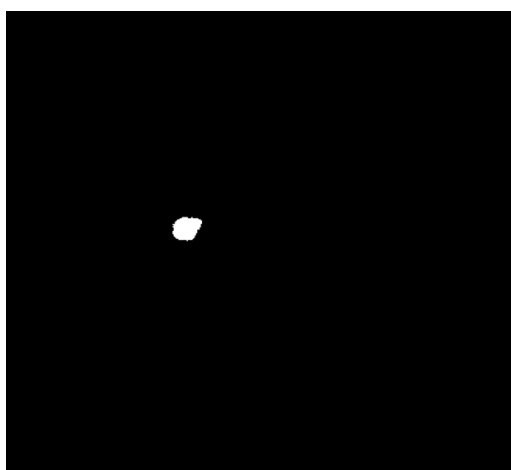
Immagine originale



Maschera manuale



Maschera automatica ResNet34



Maschera automatica ResNet50

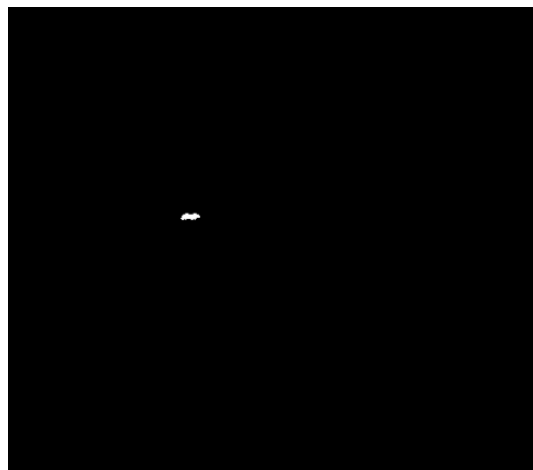


FIGURA 20 - CONFRONTO MASCHERE AUTOMATICHE RESNET34 E RESNET 50

Dalle immagini a confronto, si osserva che la ResNet34 genera delle maschere automatiche più fedeli a quelle manuali, rispetto alla ResNet50. Le dimensioni e i bordi risultano più chiari e definiti, come ci si aspettava dalle metriche ottenute. Nei casi presentati, le segmentazioni automatiche in output dall'architettura ResNet50 risultano frammentate e di ridotte dimensioni, evidenziando la difficoltà che ha avuto la rete nell'imparare i dettagli fini delle lesioni. Nonostante la ridotta profondità di layer di cui è costituita, la ResNet34 è stata selezionata come architettura migliore per l'obiettivo di segmentazione, con risultati nettamente migliori quantitativamente e qualitativamente.

## 4.2 Parametri di rete: funzioni di loss

Definita l'architettura del backbone, sono state testate tre diverse funzioni di loss: Binary Cross-Entropy (BCE), Focal Loss e Dice Loss. In seguito, vengono riportate le metriche di valutazione per i tre set di dati ed esempi di immagini di maschere automatiche in output dai rispettivi modelli.

Metriche	BCE Loss			Focal Loss			Dice Loss		
	Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
Dice coefficient	0.8229	0.5111	0.6408	0.8455	0.5283	0.2580	0.7630	0.4930	0.5957
Precision	0.9085	0.7317	0.7836	0.9067	0.7203	0.7621	0.8851	0.6968	0.7640
Recall	0.7675	0.4438	0.5986	0.8071	0.4773	0.1617	0.6770	0.4420	0.5421
F1-score	0.8229	0.5111	0.6408	0.8445	0.5283	0.2580	0.7630	0.4930	0.5957
IOU	0.7225	0.3963	0.5237	0.7530	0.4106	0.1574	0.6597	0.3889	0.4805
Specificity	0.9998	0.9997	0.9996	0.9988	0.9996	0.9999	0.9998	0.9997	0.9997
Balanced Accuracy	0.8837	0.7218	0.7992	0.9034	0.7385	0.5808	0.8484	0.7208	0.7709

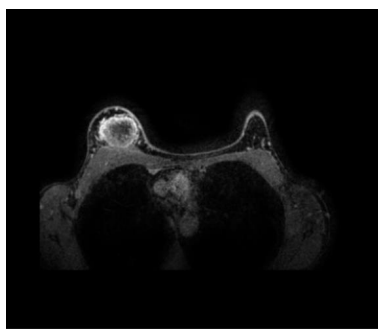
TABELLA 5 - CONFRONTO PERFORMANCE DEL MODELLO DI RETE CON TRE DIVERSE FUNZIONI DI LOSS

Visti i risultati raggiunti in termini di Dice Coefficient, Precision e Balanced Accuracy, il modello di rete che ha adottato la binary cross-entropy (BCE) come funzione di loss ha mostrato prestazioni superiori su tutti i set di dati. Per l'obiettivo di segmentazione automatica di questa tesi, è richiesto un modello di rete che non solo sappia individuare correttamente le aree di interesse e generare una maschera di segmentazione automatica accurata, ma che sia anche capace generalizzare in modo efficace su nuovi dati mai visti. In questo scenario, la BCE ha raggiunto dei valori di recall e precisione più o meno equilibrati, dimostrando di essere la scelta più valida.

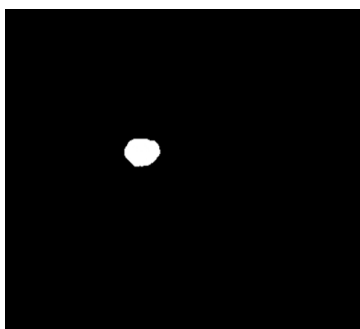
Nel caso della focal loss, implementata per dar maggior peso agli esempi più difficili da segmentare, i risultati si sono presentati ottimi in fase di allenamento ma drasticamente inferiori sui dati di test. Questo dimostra che il modello perde capacità di generalizzazione quando vengono presentati in input casi nuovi mai visti prima, probabilmente dovuti al fatto che si è posta molta attenzione ai dati di allenamento tanto da non riuscire a adattarsi a quelli di test.

La Dice Loss, infine, ampiamente utilizzata per task di segmentazione, ha ottenuto risultati migliori della Focal Loss, ma inferiori rispetto alla BCE. Questa funzione si basa sul livello di sovrapposizione tra ground truth e maschera automatica, che in questo caso non sono stati ottimali e sufficienti. Di seguito si riportano gli esempi di maschere prodotte dalle tre funzioni a sostegno di quanto detto.

Immagine originale



Maschera manuale



Maschera automatica- BCE

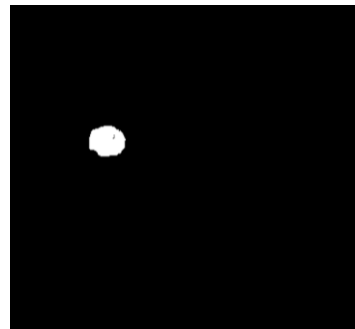
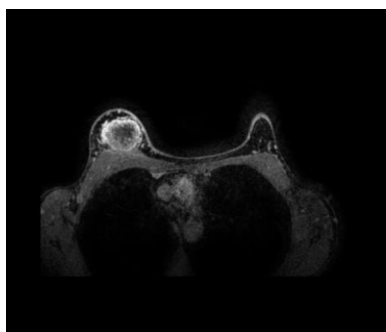
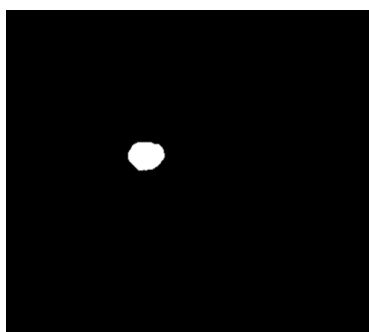


Immagine originale



Maschera manuale



Maschera automatica- Focal

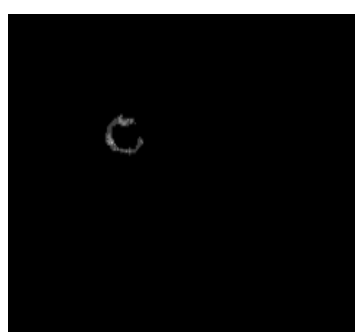
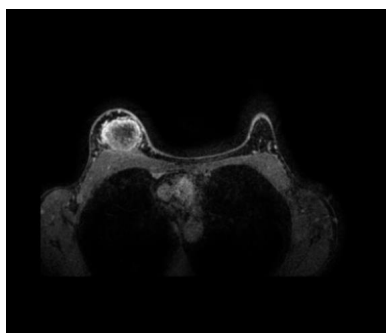
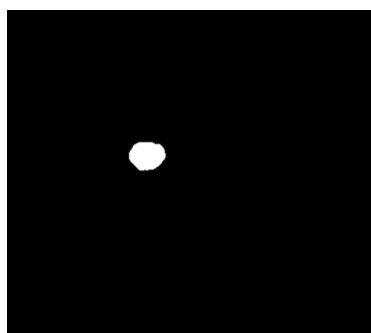


Immagine originale



Maschera manuale



Maschera automatica- Dice

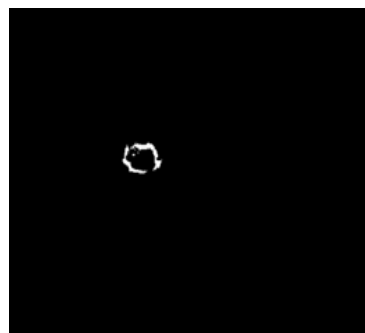


FIGURA 21- ESEMPIO CONFRONTO MASCHERE AUTOMATICHE E MANUALI PER LE TRE LOSS SPERIMENTATE

Dalle immagini presenti nella figura 21 si nota come il modello di rete che ha utilizzato la BCE loss, sia stato in grado di individuare e segmentare la regione di interesse nella sua interessa, con bordi e forme fedeli alla rispettiva maschera manuale. Ciò non può essere altrettanto detto per la Dice Loss, la quale ha prodotto maschere incomplete e non del tutto definite.

Diversamente, le maschere automatiche generate in output dall'utilizzo della focal loss, si presentano frammentate e con pochi pixel positivi individuati, come previsto dai risultati numerici sopra elencati.

Nel complesso la BCE Loss è stata scelta come funzione di loss più adeguata e performante per il modello di rete implementato, con buone capacità di generalizzazione e maggior accuratezza nel rilevamento delle regioni tumorali di interesse

### 4.3 Volumi 3D: scelta del post-processing migliore

Prima di applicare le tecniche di post-processing stabilite, sono state calcolate le metriche di valutazione sui volumi tridimensionali costruiti, al fine di avere una misura dell'accuratezza della segmentazione automatica eseguita dalla rete sullo spazio 3D.

<b>Metriche</b>	<b>Senza post-processing</b>
<b>Dice coefficient</b>	0.6924
<b>Precision</b>	0.7607
<b>Recall</b>	0.7067
<b>IOU</b>	0.5517
<b>VOE</b>	0.6538
<b>TNR</b>	0.9999

TABELLA 6- METRICHE DI VALUTAZIONE DEI VOLUMI 3D PRIMA DELL'APPLICAZIONE DEL POST-PROCESSING

Come mostra la tabella 6, i risultati volumetrici ottenuti prima dell'applicazione del post-processing appaiono coerenti a quelli ottenuti sul piano bidimensionale, ma non del tutto soddisfacenti. Il dice coefficient pari a 0.6924, ad esempio, evidenzia una discreta sovrapposizione tra la maschera manuale e quella automatica generata dal modello di rete, ma non ancora ottimale.

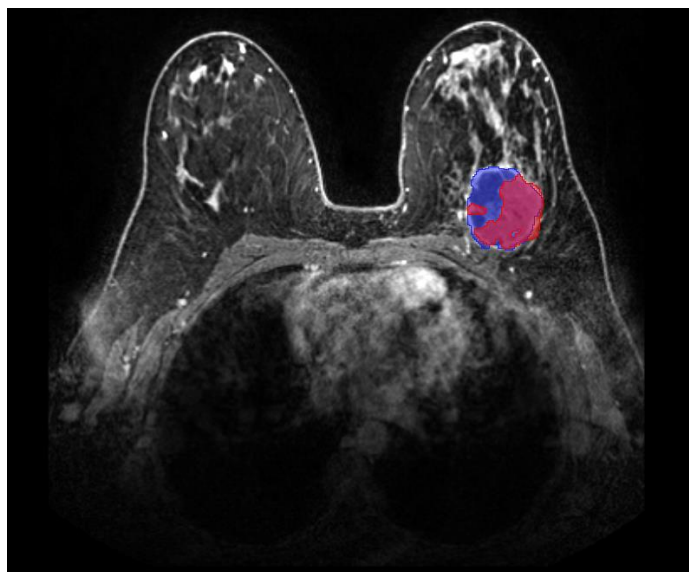


FIGURA 22 - ESEMPIO PAZIENTE DI TEST CON MASCHERA MANUALE (IN BLU) E MASCHERA AUTOMATICA (IN ROSSO) SOVRAPPOSTE

Nella figura 22 si mostra un caso in cui la segmentazione automatica (in rosso) ha dimensioni notevolmente diverse rispetto alla maschera manuale (in blu), con discrepanze anche nei bordi e aree non completamente piene. Ciò genera dei falsi positivi che riducono le metriche calcolate, come si evince dalla precision (0.7607) e dalla recall (0.7076). Inoltre, i valori di IOU (Intersection over Union) di 0.5517 e VOE (Volumetric Overlap Error) di 0.6538 indicano che il modello di deep learning implementato non riesce a catturare correttamente l'estensione delle aree patologiche. Per migliorare questi risultati ed affinare le segmentazioni automatiche prodotte, sono state applicate due tecniche di post-processing con operatori morfologici: apertura e chiusura. Per entrambe è stato definito un elemento strutturale cubico di dimensione 3x3x3.

<b>Metriche</b>	<b>Apertura</b>	<b>Chiusura</b>
<b>Dice coefficient</b>	0.7223	0.7455
<b>Precision</b>	0.7985	0.7411
<b>Recall</b>	0.7031	0.7925
<b>IOU</b>	0.5805	0.6061
<b>VOE</b>	0.6389	0.6273
<b>TNR</b>	0.9999	0.9998

TABELLA 7 - CONFRONTO METRICHE DI VALUTAZIONE TRA TECNICHE DI POST-PROCESSING

## ESEMPIO IMMAGINI – POST PROCESSING

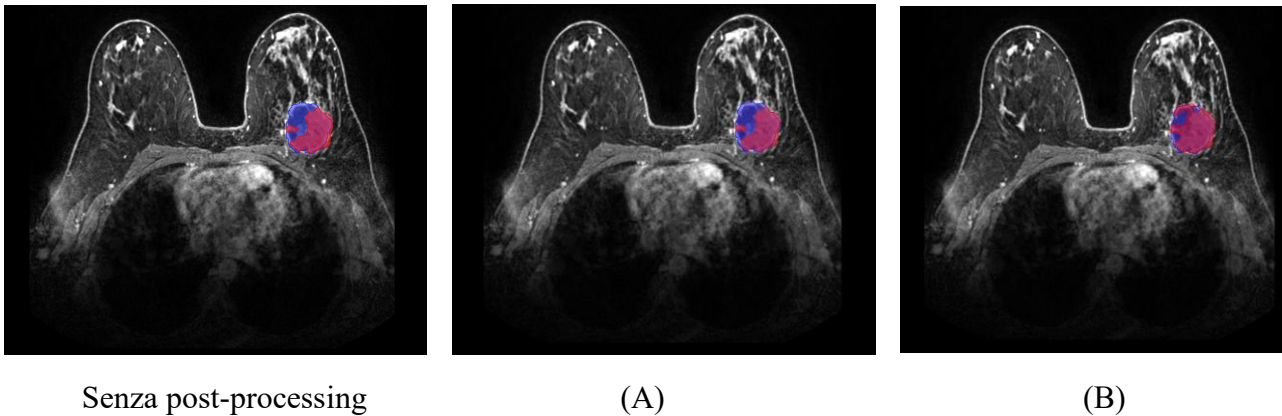


FIGURA 23 - ESEMPIO POST PROCESSING CON APERTURA (A) E CHIUSURA (B). VENGONO SOVRAPPOSTE LA MASCHERA MANUALE (BLU), QUELLA AUTOMATICA PRIMA DEL POST PROCESSING (ROSSO) E DOPO IL POST PROCESSING (ROSSO)

Nella figura 23 vengono mostrate le immagine relativa alle tecniche di post-processing applicate, cioè apertura (A) e chiusura (B). Le metriche riportate nella tabella 7 mostrano come i risultati dopo l'applicazione degli operatori morfologici siano nettamente migliorati. In particolare, l'operazione di apertura ha ridotto il numero di falsi positivi aumentando di conseguenza il valore di precision da 0.7607 a 0.7985.

Dall'immagine 29A si nota, difatti, che alcuni buchi sono stati riempiti e i bordi sono diventati più definiti e meglio allineati con quelli della maschera manuale. Questo ha portato ad un aumento dell'accuratezza della sovrapposizione tra maschera automatica e ground truth, come confermato dall'incremento del dice coefficient che è passato da 0.6924 a 0.7358. Nonostante ciò, la chiusura è stata scelta come l'operazione di post-processing migliore. Essa ha mostrato risultati superiori in termini di dice con una segmentazione più precisa e completa, ma ha permesso anche di riempire alcune lacune individuando meglio l'estensione della patologia.

Questo risultato è evidente non solo visivamente, come mostrato nella figura 23B ma anche nella Recall (0.7925), il cui aumento dimostra un ridotto numero di falsi negativi, cioè di pixel erroneamente classificati come negativi appartenenti allo sfondo ma che nella realtà fanno parte della lesione.

Riassumendo, se da un lato l'apertura è ottimale per eliminare spot isolati, è altrettanto vero che presenta limiti perdendo dettagli su forme e bordi. Con la chiusura, invece, applicando prima la dilatazione che espande le regioni segmentate e successivamente l'erosione, molte aree mancanti vengono riempite, ottenendo una segmentazione più affidabile e completa.

#### 4.4 Casi critici: Discussione

Nonostante i risultati ottenuti dal modello di rete e dalle tecniche di post-elaborazione implementate siano risultati molto positivi, l'analisi critica ha evidenziato anche alcuni casi in cui la segmentazione automatica appare molto imprecisa e non fedele alla corrispettiva segmentazione manuale. Questi casi critici non solo hanno messo in luce alcuni errori fatti del modello di rete implementato, ma anche dei limiti derivanti dalle caratteristiche delle immagini di risonanza magnetica dinamica.

##### Caso 1: Segmentazione di tessuti non tumorali

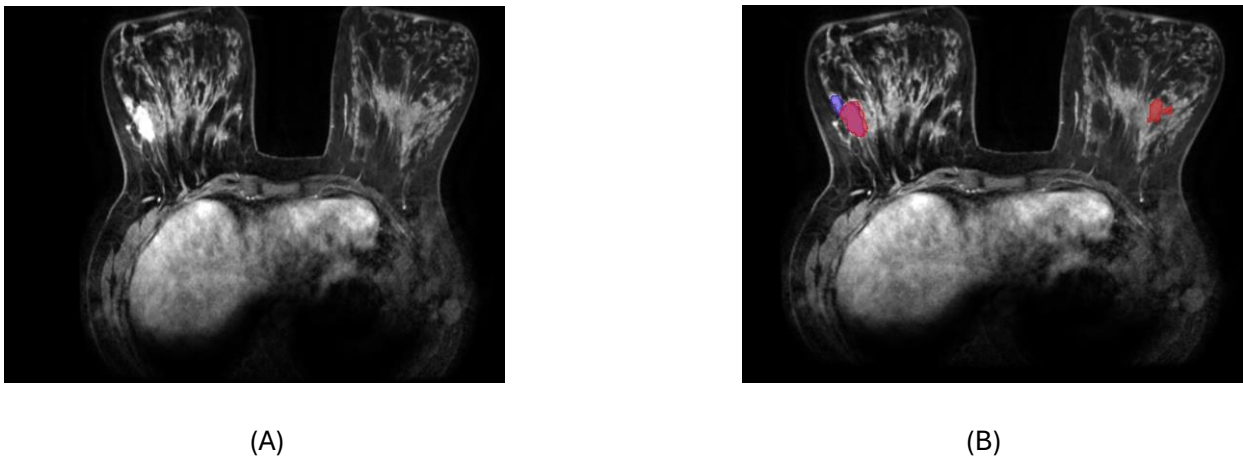


FIGURA 24 - CASO CRITICO 1: CONFRONTO IMMAGINE ORIGINALE (A) E SEGMENTAZIONI (B). IN BLU È PRESENTE LA SEGMENTAZIONE MANUALE, IN ROSSO LA SEGMENTAZIONE AUTOMATICA

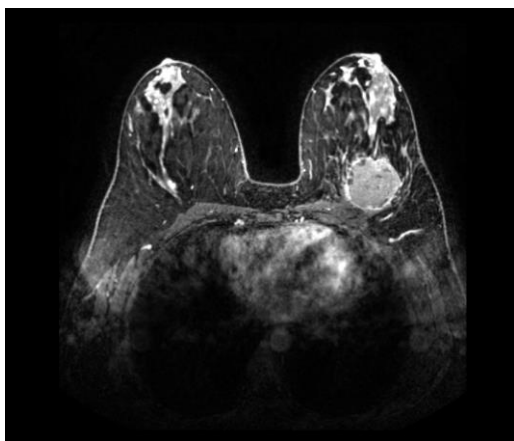
Il primo caso analizzato mostra gli errori che la rete ha compiuto nell'individuare e segmentare zone non tumorali, identificate erroneamente come patologiche. La figura 24A mostra l'immagine originale dove il tumore è ben evidenziato da un'alta intensità di segnale, che la rete riesce facilmente ad individuare. Nonostante ciò, nella figura 30B sono presenti la maschera manuale (in blu) e la maschera automatica (in rosso), la quale include un'area nel seno destro non corrispondente ad una zona patologica.



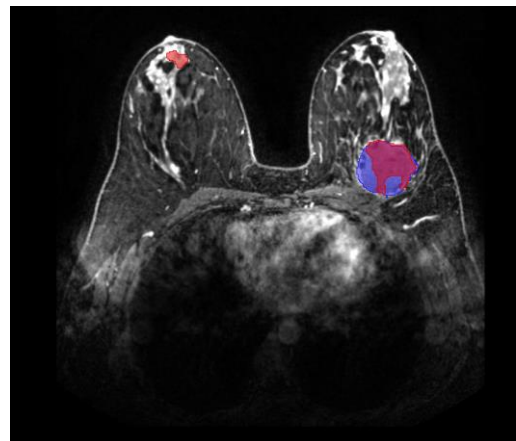
Questo errore può essere associato ad alcuni limiti della rete neurale applicata, la quale catturando pattern locali tramite filtri convoluzionali, riesce efficacemente ad individuare dettagli fini dell'immagine, ma presenta anche una forte dipendenza dall'intensità di segnale a cui viene associato il tumore. Di conseguenza, tessuti molto densi con caratteristiche simili, come bordi o intensità ravvicinate, possono essere erroneamente classificati come patologici.

Dai risultati descritti precedentemente, è risultato evidente anche il fenomeno dell'overfitting, per cui la rete apprende bene le caratteristiche delle lesioni presenti nelle maschere manuali durante l'allenamento, ma le prestazioni si riducono nel momento in cui viene richiesto di generalizzare su casi nuovi mai visti. La seconda area segmentata, nonostante abbia un'intensità di segnale molto diversa da quella del tumore, potrebbe essere stata erroneamente identificata come patologica a causa di un eccessivo adattamento ai pattern appresi nei dati di training.

## Caso 2: Segmentazione di strutture non patologiche con intensità simili



(A)



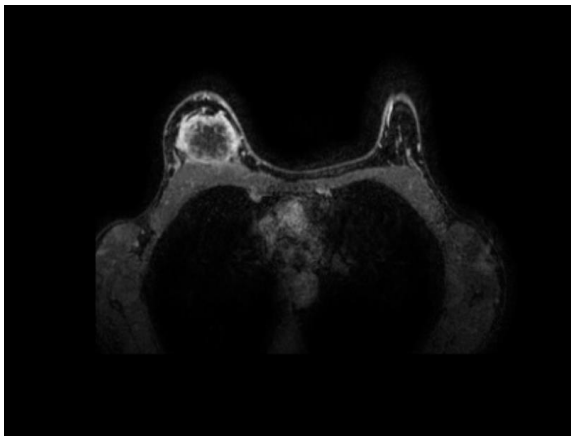
(B)

FIGURA 25 - CASO CRITICO 2: CONFRONTO IMMAGINE ORIGINALE (A) E SEGMENTAZIONE (B). IN ROSSO È PRESENTE LA SEGMENTAZIONE MANUALE, IN BLU LA SEGMENTAZIONE AUTOMATICA

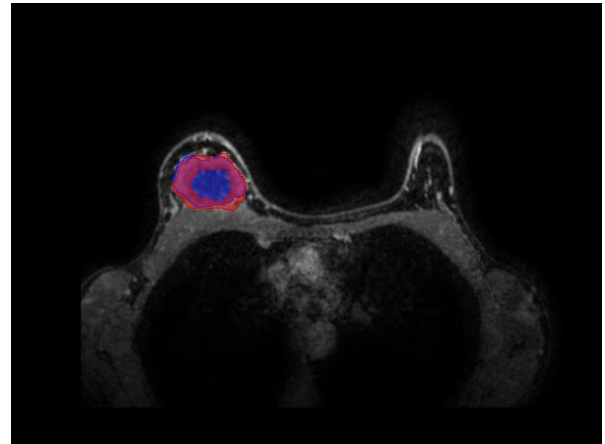
Il secondo caso si differenzia dal precedente in quanto le ulteriori zone segmentate presentano un'alta intensità di segnale, che potrebbe essere associata a un'area patologica. Nella figura 25B, infatti, si nota in rosso una piccola area segmentata in prossimità del capezzolo sinistro della paziente che, nonostante non sia stata classificata come tumorale, presenta un'intensità di segnale simile alla lesione di interesse. Questo errore è attribuibile all'imaging di risonanza magnetica, che si basa sulla diffusione del contrasto tra i tessuti, influenzato dalla rete di vascolarizzazione e dalla densità del tessuto mammario.

Le lesioni tumorali, si presentano con un'alta intensità di segnale durante il picco di diffusione del contrasto, ma anche altre strutture, come vasi sanguigni e linfonodi, possono avere caratteristiche simili. Poiché la rete neurale ha imparato a distinguere aree patologiche e aree sane principalmente sulla base dell'intensità del segnale, essa identifica come tumorale aree con intensità elevate, indipendentemente dalla struttura anatomica sottostante.

### Caso 3: Segmentazioni incomplete



(A)



(B)

*FIGURA 26 - CASO CRITICO 3: CONFRONTO IMMAGINE ORIGINALE (A) E SEGMENTAZIONE (B). IN VERDE È PRESENTE LA SEGMENTAZIONE MANUALE, IN BLU LA SEGMENTAZIONE AUTOMATICA*

Il terzo caso analizzato pone l'attenzione su segmentazioni incomplete che ha effettuato la rete neurale rispetto alla segmentazione manuale realizzata dal senologo esperto. Nella figura 26B la maschera manuale in blu include l'intera area delle massa tumorale, mentre la segmentazione automatica (in rosso) ne delimita solamente i contorni, escludendo l'area interna. Questa differenza è dovuta alla diversa vascolarizzazione tra i bordi e la zona centrale caratterizzata dalla presenza di necrosi. Il mezzo di contrasto utilizzato nella risonanza magnetica, difatti, si diffonde facilmente nelle aree vascolarizzate producendo un segnale evidente e definito. La presenza di necrosi all'interno dell'area tumorale, indica una scarsa perfusione e di conseguenza una ridotta intensità di segnale.

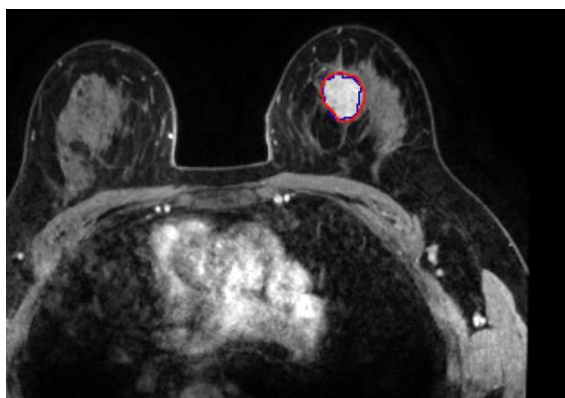
Poiché la rete neurale distingue aree patologiche attraverso l'alta vascolarizzazione e l'intensità del segnale prodotto, essa ha correttamente individuato la lesione e i suoi confini ma, posta a confronto con la reference manuale, la segmentazione nel completo viene classificata come incompleta.

Questo dimostra come la rete soddisfi l'obiettivo di segmentazione del tessuto mammario identificando correttamente la zona e i confini della lesione. Tuttavia, la differenza nella sovrapposizione delle due maschere ha penalizzato le metriche di valutazione calcolate.

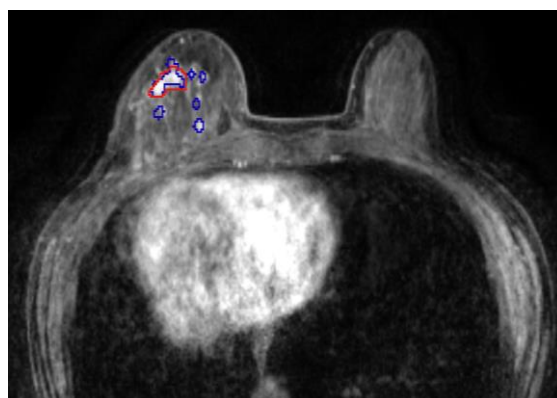
### **Lesioni mass e non mass**

Come descritto nei precedenti capitoli, il dataset utilizzato era composto da lesioni di tipo mass e non-mass, che si differenziano per le loro caratteristiche morfologiche.

Nell'analisi critica delle segmentazioni automatiche prodotte in output dal modello implementato, è stato anche valutato se esso fosse in grado di produrre una segmentazione accurata per entrambe le tipologie di lesioni.



(A) Lesione di tipo mass



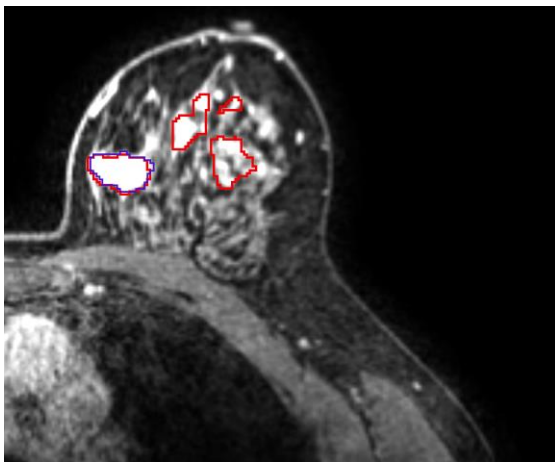
(B) Lesione di tipo non mass

*FIGURA 27 - DIFFERENZE TRA LE LESIONI DI TIPO MASS (A) E NON MASS (B). IN ROSSO È RIPORTATA LA SEGMENTAZIONE AUTOMATICA E IN BLU LA SEGMENTAZIONE MANUALE.*

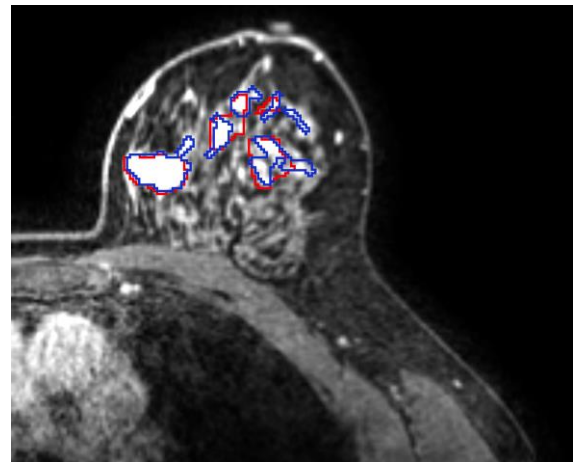
La figura 27 sopra riportata mostra come per le lesioni di tipo mass la segmentazione automatica in rosso presenta dei contorni coerenti con quelli della maschera manuale. Per il caso non mass, invece, i contorni si presentano più irregolari e poco accurati, evidenziando come il modello riscontrato difficoltà nell'individuare il volume della lesioni. Alcuni piccoli spot di lesione tumorale, infatti, non sono stati rilevati, suggerendo che la rete non riesca a identificare bene tessuti patologici frammentati e diffusi.

## Analisi dell'Impatto delle Annotazioni Complete sulla Segmentazione

Le maschere manuali utilizzate per il lavoro di tesi provengono da un precedente studio di radiomica in cui si richiedeva di individuare solamente la lesione principale al seno di ciascun paziente. Tuttavia, per migliorare la completezza delle annotazioni e verificare se questo potesse essere impattante per i risultati finali ottenuti, sono stati riesaminati i volumi dei pazienti del test set insieme al supporto di una senologa esperta. Questa ulteriore ispezione ha permesso di integrare maggiori annotazioni, includendo tutte le lesioni tumorali presenti nelle immagini ed è stato determinato per analizzare l'accuratezza delle segmentazioni automatiche in un quadro più completo.



(A) Annotazioni primarie



(B) Annotazioni complete

*FIGURA 28 - DIFFERENZE TRA LE MASCHERE MANUALI UTILIZZATE E LE MASCHERE MANUALI CON ANNOTAZIONI COMPLETE. IN ROSSO È RIPORTATA LA SEGMENTAZIONE AUTOMATICA E IN BLU LA SEGMENTAZIONE MANUALE.*

Nella figura 28A si nota come rete ha individuato perfettamente la lesione presente, con contorni perfettamente sovrapponibili alla maschera manuale. Nonostante ciò, vengono identificate dal modello anche ulteriori aree del seno come patologiche, risultando di conseguenza come falsi positivi che influenzano in modo negativo i risultati delle metriche di valutazione.

Al contrario, nella figura 29B, le aree identificate si sono rivelate effettivamente delle lesioni tumorali, sebbene ci sia una precisione inferiore rispetto alla lesione principale, grazie all'inserimento delle annotazioni complete.

Questo risultato dimostra come alcuni errori si verificano per le segmentazioni manuali incomplete presenti nel dataset che, una volta aggiunte, permetterebbero di migliorare le prestazioni e ridurre il numero di falsi positivi.

# Capitolo 5

## Conclusioni e sviluppi futuri

In questo lavoro di tesi si è sviluppato un modello di rete neurale basato su deep learning per effettuare una segmentazione automatica delle lesioni tumorali al seno. Lo scopo è stato quello di creare uno strumento di supporto per il radiologo per ridurre i tempi di lavoro nell'identificazione delle aree patologiche e l'influenza della variabilità inter e intra-operatore. A differenza dei metodi tradizionali, l'uso del deep learning permette di ottimizzare il flusso di lavoro, migliorando l'accuratezza della diagnosi con risultati più attendibili.

I risultati migliori sono stati ottenuti dal modello di rete UNet con backbone ResNet34 e funzione di loss binary cross-entropy, con un valori finali di dice e recall rispettivamente di 0.6924 e 0.7067 sui volumi ricostruiti. Durante l'allenamento sono stati raggiunti alti livello di accuratezza già dalle prime epoche, evidenziando la presenza del fenomeno di overfitting. Esso ha compromesso le prestazioni sui dati di test, indicando una ridotta capacità di generalizzazione del modello. Le tecniche di post-processing adottate hanno permesso di rimuovere imperfezioni presenti nelle segmentazioni automatiche, come spot spuri e buchi, aumentando i valori di dice a 0.7455 e recall a 0.7925. Tuttavia, nonostante la presenza di falsi positivi presenta ancora un limite importante, il modello di rete sviluppato ha mostrato di essere in grado di delineare in modo accurato i margini delle lesioni tumorali presenti nelle immagini di risonanza, prerogativa essenziale per l'obiettivo di segmentazione scelto.

Le maschere manuali utilizzate, inoltre, presentavano solo la lesione principale individuata dalle senologhe esperte. Si è dimostrato, a tal proposito, che alcuni errori compiuti della rete che hanno portato ad un aumento dei falsi positivi, erano in realtà legati alla mancata presenza di annotazioni complete sulle segmentazioni manuali. Di conseguenza, risulta rilevante utilizzare maschere completamente annotate per avere un aumento delle prestazioni del modello di rete.

Per migliorare l'accuratezza del modello di rete e ridurre il fenomeno dell'overfitting presente si auspica l'introduzione di ulteriori operazioni di data augmentation e l'utilizzo di tecniche avanzate di regolarizzazione. L'utilizzo di dropout durante l'allenamento, ad esempio, potrebbe aiutare la rete a non adattarsi rapidamente ai dati di allenamento, migliorando così la capacità di generalizzazione sui nuovi casi.

Inoltre, si potrebbe considerare un'architettura di rete più complessa e profonda o l'impiego di un metodo come l'ensemble learning (apprendimento d'insieme) che combina diversi modelli e aumenta la robustezza del sistema di deep learning.

Anche l'uso della multimodalità potrebbe portare dei vantaggi: la sequenza DWI (Diffusion Weighted Imaging), metterebbe in evidenza il diverso comportamento di diffusione tra lesioni tumorali e strutture vascolari, favorendo la riduzione di aree erroneamente segmentate.

In conclusione, questa tesi ha mostrato come l'utilizzo di tecniche di deep learning potrebbe pienamente soddisfare l'esigenza di avere una segmentazione automatica delle lesioni. Nonostante siano ancora presenti alcuni limiti, gli sviluppi futuri suggeriti offrono approcci innovativi che mirano a migliorare ulteriormente le prestazioni ottenute, rendendo la segmentazione automatica uno strumento di supporto sempre più affidabile per la diagnosi precoce e l'ottimizzazione del flusso di lavoro.



# Bibliografia

- [1] World Health Organization. (2021). Breast cancer. Available at: <https://www.who.int/news-room/factsheets/detail/breast-cancer>.
- [2] Mammografia: Cos'è, come si effettua, FAQ | AIRC. (n.d.). <https://www.airc.it/cancro/affronta-la-malattia/guida-agli-esami/mammografia>
- [3] Mann, R. M., Kuhl, C. K., Kinkel, K., & Boetes, C. (2008). Breast MRI: guidelines from the European Society of Breast Imaging. *European radiology*, 18(7), 1307–1318. <https://doi.org/10.1007/s00330-008-0863-7>
- [4] Insider Radiology. (n.d.). *Gadolinium contrast medium*. Insider Radiology. Available at: <https://www.insideradiology.com.au/gadolinium-contrast-medium/>
- [5] A.D. Elster. *DCE analysis*. (n.d) Questions and Answers in MRI. Available at: <https://mriquestions.com/how-is-dce-analyzed.html#/>
- [6] Mafraji, M. A. (2023, November 4). Imaging a risonanza magnetica. Available at: <https://www.msmanuals.com/it/professionale/argomenti-speciali/principi-di-imaging-radiologico/imaging-a-risonanza-magnetica>
- [7] Thomassin-Naggara, I., Trop, I., Lalonde, L., David, J., Péloquin, L., & Chopier, J. (2012). Tips and techniques in breast MRI. *Diagnostic and interventional imaging*, 93(11), 828-839. <https://doi.org/10.1016/j.diii.2012.06.004>
- [8] Gordon, Y., Partovi, S., Müller-Eschner, M., Amarteifio, E., Bäuerle, T., Weber, M. A., Kauczor, H. U., & Rengier, F. (2014). Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. *Cardiovascular diagnosis and therapy*, 4(2), 147–164. <https://doi.org/10.3978/j.issn.2223-3652.2014.03.01>
- [9] Türkbey, B., Thomasson, D., Pang, Y., Bernardo, M., & Choyke, P. L. (2010). The role of dynamic contrast-enhanced MRI in cancer diagnosis and treatment. *Diagnostic and interventional radiology (Ankara, Turkey)*, 16(3), 186–192. <https://doi.org/10.4261/1305-3825.DIR.2537-08.1>
- [10] Bougias, H., & Stogiannos, N. (2022). Breast MRI: Where are we currently standing?. *Journal of Medical Imaging and Radiation Sciences*, 53(2), 203-211. <https://doi.org/10.1016/j.jmir.2022.03.072>



- [11] Alshuhri, M. S., Al-Musawi, S. G., Al-Alwany, A. A., Uinarni, H., Rasulova, I., Rodrigues, P., Alkhafaji, A. T., Alshanberi, A. M., Alawadi, A. H., & Abbas, A. H. (2023). Artificial intelligence in cancer diagnosis: Opportunities and challenges. *Physiological Reports*, 11(5), e154996. <https://doi.org/10.1016/j.prp.2023.154996>
- [12] Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3), 257-273.
- [13] Bhowmik, A., & Eskreis-Winkler, S. (2022). Deep learning in breast imaging. *BJR|Open*, 4(1), 20210060. <https://doi.org/10.1259/bjro.20210060>
- [14] Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in neural information processing systems*, vol. 25, 2012
- [15] Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., ... & Merhof, D. (2024). Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Doi: [10.1109/TPAMI.2024.3435571](https://doi.org/10.1109/TPAMI.2024.3435571)
- [16] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234-241). Springer. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [17] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
- [18] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- [19] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [20] Mahendra, Sanksshep. 2023. "What Is the Adam Optimizer and How Is It Used in Machine Learning." *Artificial Intelligence +*. Retrieved (<https://www.aiplusinfo.com/blog/what-is-the-adam-optimizer-and-how-is-it-used-in-machine-learning/>).

