POLITECNICO DI TORINO

Corso di Laurea Magistrale

in Ingegneria Biomedica

Tesi di Laurea Magistrale

Sviluppo di un CAD per la detection di tumori al seno non- mass con immagini di risonanza magnetica



Relatori

Prof. Rosati Samantha

Candidata

Anna Macula

Prof. Giannini Valentina

Anno accademico 2020/2021

Indice

1	Sor	nmai	rio 1	1
2	Intr	oduz	zione	2
	2.1	Obi	ettivo della tesi	2
	2.2	Stru	ittura della tesi	2
	2.3	Tun	nore al seno2	2
	2.4	MR	I	7
	2.4	.1	Tecnologia	7
	2.4	.2	MRI nel tumore al seno	3
3	Sta	to de	ll'arte11	1
4	Ma	teria	li e Metodi 21	1
	4.1	Dat	aset	1
	4.1	.1	Origine dei dati 21	1
	4.1	.2	Pre processing dei dati 22	2
	4.1	.3	Suddivisione Dataset	3
	4.2	Stru	uttura della rete	3
	4.3	Mo	dalità di allenamento 25	5
5	Ris	ultat	i	5
	5.1	Met	todi di valutazione dei risultati	5
	5.1	Nor	malizzazione dei dati di input27	7
	Z-s	core		7
	Min	n-Ma	ax Scaling	3
	5.2	Bin	arizzazione delle immagini predette 29)

Soglia round	29
Soglia di Otsu	30
Soglie della curva ROC	30
5.3 Calcolo dei parametri per ogni modello	31
Confronto tra loss functions	36
6 Valutazione delle prestazioni	39
Conclusione e prospettive future	52
Bibliografia	54
Ringraziamenti	56

1 Sommario

Nel seguente progetto di tesi è stato implementato un CAD per la detezione di tumore al seno non mass attraverso l'utilizzo delle librerie di deep learning di Python, Keras e TensorFlow. Dopo un excursus sulla problematica e le metodologie di screening, è stato approfondito l'utilizzo della Intelligenza Artificiale nella clinica. E' stata costruita una rete neurale convoluzionale di tipo U-Net sulla base di immagini di risonanza magnetica, tecnica che garantisce una sensibilità superiore all'esame mammografico, considerato ancora il "gold standard" per questo tipo di problematica. Sono stati implementati diversi modelli di rete al variare della normalizzazione dei dati di input o della loss function e successivamente sono state testate alcune binarizzazioni sui dati predetti, per osservare e valutare gli effetti del variare dei suddetti parametri sulle prestazioni del CAD. Si è mostrato come effettivamente i metodi scelti per l'allenamento della rete e la rappresentazione dei risultati siano decisamente influenti sulla performance della rete. Il lavoro di questa tesi può e deve sicuramente essere ampliato, perché i risultati della rete mostrano una ottima capacità di riconoscere la lesione, affiancata però dalla presenza di molti falsi positivi, dovuti soprattutto alla presenza degli organi rappresentati nell'MRI che costituiscono interferenza. I metodi di CAD, sia di detezione che di diagnosi, potrebbero rivelarsi davvero un ottimo supporto alla pratica clinica e la ricerca in questo campo è in pieno svolgimento.

2 Introduzione

2.1 Obiettivo della tesi

L'obiettivo del mio progetto di tesi è la costruzione di un "Computer aided detection" (CAD) per la segmentazione automatica di tumore al seno non-massivo. Ciò sarebbe di notevole utilità in clinica perché ad oggi il tumore viene segmentato manualmente dal radiologo: l'utilizzo di un CAD permetterebbe invece di automatizzare il processo, rendendolo estremamente più rapido ed anche meno affetto da variabilità inter e intra operatore. Il CAD è costruito e valutato su immagini di risonanza magnetica, attraverso il deep learning, e sono effettuati confronti tra differenti normalizzazioni delle immagini del dataset, diverse loss functions e binarizzazioni delle maschere predette dal CAD, in maniera tale da valutare e confrontare gli effetti che il variare di questi parametri ha sulle prestazioni del sistema.

2.2 Struttura della tesi

Dopo una breve introduzione sulla problematica del tumore al seno e sulla tecnica di imaging che è stata utilizzata per ricavare il Dataset, il capitolo sullo Stato dell'Arte approfondisce le tecniche di Intelligenza Artificiale e soprattutto di Deep Learning impiegate nella clinica e cita alcuni esempi di segmentazione automatica confrontabili con il lavoro svolto in questa tesi. Successivamente nel capitolo Materiali e Metodi sono descritti il dataset e la struttura della rete utilizzata, e tutte le prove effettuate, i cui risultati sono appunto riportati nel relativo capitolo. E' infine presente una sezione che riguarda le conclusioni e le prospettive future di ricerca nel campo.

2.3 Tumore al seno

2.26 milioni di diagnosi di tumore al seno e 685000 morti globali sono i dati relativi al 2020 forniti dall'organizzazione mondiale della sanità, che ha decretato il tumore al seno come il cancro ad incidenza più frequente nel mondo. [1]

La malattia può coinvolgere entrambi i generi, tuttavia la percentuale di uomini affetti è minima, circa dell'1%, e i protocolli di trattamento sono gli stessi adottati per le donne.

Anatomia

La mammella è una struttura ghiandolare esocrina composta da tre tipologie di tessuti: ghiandolare, che si occupa della produzione del latte, adiposo, che conferisce forma e grandezza al seno, e fibroso, responsabile del sostegno. Il seno è "denso" quando si ha prevalenza di tessuto ghiandolare e fibroso; questa condizione tendenzialmente varia con l'avanzare dell'età, con la quale aumenta la percentuale di tessuto adiposo.

Il tessuto ghiandolare è composto da lobuli, responsabili della produzione del latte, che si raggruppano formando 15 o 20 strutture più grandi chiamate lobi. Queste strutture sono collegate tramite i dotti mammari, che portano il latte prodotto verso i capezzoli.

Il funzionamento della ghiandola è coadiuvato dall'azione di ormoni femminili, cioè gli estrogeni, il progesterone e la prolattina.

Fattori di rischio

Il rischio di tumore al seno aumenta con l'età, in particolare nel periodo della menopausa: su un totale di 355.457 casi di tumore alla mammella stimati in Europa dall'ECIS (European Cancer Information System), 40.613 riguardano donne tra i 20 e i 44 anni (11%), mentre 191.955 sono i casi diagnosticati a donne tra i 45 e i 69 anni (54%) e 122.880 a donne dai 70 anni in su[2]. Risulta dunque che più del 80% dei casi di carcinoma al seno riguarda donne oltre i 45 anni. In particolare, coloro che sono coinvolte in una menopausa tardiva hanno maggior possibilità di contrarre la malattia, probabilmente a causa dell'esposizione prolungata ad estrogeni; i disequilibri ormonali, legati anche all'utilizzo di contraccettivi orali o alla terapia ormonale post menopausa, risultano dunque importanti indicatori di rischio. Al contrario, diversi studi hanno dimostrato come gravidanza in giovane età e l'allattamento al seno siano fattori riducenti dell'insorgenza di tumore al seno, perché responsabili del processo di differenziamento del tessuto mammario.[3]

Oltre ai fattori di rischio legati alla salute riproduttiva però, ci sono da considerare numerosi altri aspetti quali l'obesità, l'attività fisica, il fumo ed il consumo di e alcool, ma anche fattori anatomici, quali la densità del tessuto mammario, e genetici, ovvero la "storia" familiare e le mutazioni geniche, in particolare relative ai geni oncosoppressori BRCA1 e BRCA2. [3], [4], [5].

Prospettiva globale

Lo stile di vita "occidentale", caratterizzato da maggiore sedentarietà e da un cambiamento quasi radicale delle abitudini alimentari e riproduttive, sta contribuendo al diffondersi del cancro al seno. In accordo con un articolo del The Lancet Global Health, nonostante nei cosiddetti paesi con alto indice di sviluppo (HDI, Human Development Index), si abbia una maggiore incidenza di tumore al seno rispetto ai paesi in via di sviluppo [Fig.1], il tasso di mortalità precoci, relativo a donne pre menopausa, si attesta a 8.5 su 100000 nei paesi con più basso HDI contro i 2.8 su 100000 per i paesi con HDI più elevato (dati standardizzati per età)[6]. Questa apparente contraddizione è purtroppo imputabile alla forte disparità dovuta alla differenza di approccio alla malattia, che viene diagnosticata e trattata in maniera più specifica ed efficace nei paesi più ricchi.



Female breast cancer incidence and mortality rates, 2018

Fig.1 – Rates di incidenza e mortalità del cancro al seno, 2018 [7]

Classificazione

Il tumore al seno è dunque una malattia estremamente eterogenea, che può comparire con diverse manifestazioni, le quali si distinguono tra loro per le caratteristiche clinico patologiche, ma anche per la diagnosi e la modalità di trattamento. [2], [3].

E' necessario distinguere preliminarmente tra tumore non invasivo, che rimane circoscritto nella zona in cui si forma, più comunemente nei dotti lattiferi (Carcinoma Duttale in Situ, CDS) o nei lobuli (Carcinoma Lobulare in situ, LCIS), e tumore invasivo, le cui cellule si espandono attraverso il sistema immunitario o circolatorio e coinvolgono altri tessuti. Il tumore invasivo più comune è il carcinoma duttale, che costituisce il 70-80% dei casi di tumore al seno, mentre il secondo maggiormente diffuso è il carcinoma lobulare (10-15%). [8]

Il sistema TNM è il metodo di classificazione tumorale più comunemente utilizzato:

• T si riferisce alla grandezza e all'estensione del tumore, con i relativi biomarkers;

- N indica il numero e la grandezza dei linfonodi coinvolti, e dove si collocano;
- M da informazioni sulle metastasi, quindi se il tumore si è espanso in altri tessuti.

Alla lettera viene associato un numero, da 0 a 3, che indica il livello di gravità.

I medici raggiungono la prognosi, e dunque pianificano il piano di trattamento, combinando le classificazioni di T,N,M con le informazioni sul *grado* del tumore, ovvero una misura di quanto le cellule tumorali si discostano da cellule sane, e con i risultati dei test di positività a ER/PR e HER2, recettori ormonali target. [9], [10]

Screening e metodi di trattamento

La diagnosi precoce della malattia risulta essere l'obiettivo a cui puntare: individuare un tumore ai primi stadi permette di intervenire in maniera meno invasiva e debilitante per il paziente, e sicuramente più efficace. Secondo l'articolo di Cardoso et al., 25 paesi europei hanno messo in atto negli ultimi anni dei programmi di screening a livello nazionale o regionale, per permettere di rilevare il tumore al seno in uno stadio preclinico, dunque ben prima della comparsa dei sintomi. Le linee guida ESMO raccomandano una mammografia biennale alle donne tra i 50 e i 69 anni e una risonanza magnetica (e/o mammografia) annuale a donne con una forte storia familiare di cancro al seno. La mammografia è considerata il gold standard per lo screening del cancro al seno, tuttavia negli ultimi anni l'impiego dell'MRI si sta affermando sempre di più grazie alla elevatissima sensibilità. Sempre in accordo con le linee guida di ESMO, la risonanza magnetica è consigliata in casi quali storia familiare di cancro con mutazioni di BRCA, seno denso, cancro lobulare, sospetto di multifocalità/multicentricità, presenza di discrepanze tra imaging tradizionale e esaminazione clinica o risultati poco chiari. [11]

Nel trattamento del cancro al seno, l'intento è quello di prolungare l'aspettativa di vita garantendone però la qualità: si cerca di trovare una cura che elimini il tumore nella maniera meno invalidante per il paziente. La *chirurgia* è il metodo più diffuso per estirpare la massa tumorale, soprattutto nel caso questa sia rimasta circoscritta, ma anche per fasi più complesse della malattia. Ci sono diversi approcci, tra cui il più radicale è la mastectomia, l'eliminazione totale di un seno o di entrambi: questa scelta risulta la più conservativa e si effettua anche a scopo preventivo per donne ad alto rischio di tumore.

Un'alternativa alla mastectomia è una lumpectomia, intervento chirurgico di rimozione molto meno invasivo, coadiuvata dall'impiego della radioterapia, soprattutto nei casi di tumore ai primi stadi.

I tumori arrivati già in fase metastatica sono principalmente trattati con la chemioterapia, ovvero con l'utilizzo di farmaci antitumorali che tentano di uccidere le cellule tumorali. Negli ultimi anni l'abbinamento della termoterapia ha consentito di migliorare la prestazione della chemioterapia, rendendola più specifica ed efficace.

Ci sono altre metodologie di cura alternative, tra cui l'immunoterapia o le *terapie ormonali*, che risultano adatte alla cura di tumori positivi ai recettori di estrogeni, prolattina o HER2 [8].

2.4 MRI

La Risonanza Magnetica è una tecnica di imaging 3D non invasiva notevolmente diffusa in ambito clinico, adatta alla visualizzazione di tutti i tessuti ad esclusione di quello osseo. A differenza della tomografia computerizzata (CT), non utilizza alcun tipo di radiazione dannosa per l'organismo, tuttavia risulta più onerosa sia a livello di costi che di tempo.

In ambito oncologico è impiegata per la diagnosi e la valutazione e il monitoraggio del trattamento, e si sta rivelando una preziosa risorsa anche per lo screening tumorale, come suddetto.

2.4.1 Tecnologia

Il dispositivo per MRI si basa sulla generazione di un forte campo elettromagnetico statico, tipicamente dell'ordine di 0.5-3 Tesla, che fa allineare il momento magnetico dei nuclei degli atomi di idrogeno presenti nei tessuti con la sua direzione; gli atomi vengono eccitati a radiofrequenza, e modificano il loro orientamento, il quale avrà una direzione parallela e una trasversale alla linea di forza del campo statico; quando l'eccitazione viene interrotta, i nuclei si riallineano al campo: la variazione del vettore magnetizzazione crea

un campo elettromagnetico indotto su una bobina. Questo segnale è denominato FID (Free Induction Decay) ed ha espressione approssimata:

$$A = A_0 e^{\frac{-t}{T_1}} e^{\frac{-t}{T_2}}$$

L'ampiezza A_0 dipende dall'intensità del campo magnetico statico, che dunque controlla la risoluzione dell'immagine, mentre T_1 e T_2 sono i tempi di rilassamento, legati rispettivamente al recupero della componente longitudinale del vettore magnetizzazione (rilassamento spin-reticolo) e all'annullamento di quella trasversale (rilassamento spin spin). I tempi sono una caratteristica specifica del tessuto e variano notevolmente nel caso di tessuto patologico, quindi la loro rilevazione è fondamentale.

Il paziente è sottoposto anche a tre campi magnetici "gradiente" ovvero che variano linearmente in funzione dello spazio: ciò consente di mappare spazialmente l'informazione del FID e costruire l'immagine di risonanza magnetica, la cui scala di grigi rispecchia l'intensità del segnale FID, dunque la differente risposta dei tessuti dovuta alla diversa quantità protonica ed anche alla particolare sequenza di eccitazione utilizzata. Spesso per avere una migliore resa dell'immagine viene iniettato un mezzo di contrasto, tipicamente a base di gadolinio, un materiale in grado di modificare il comportamento magnetico dei protoni nei tessuti umani. (Contrast-Enhanced Perfusion MRI)

2.4.2 MRI nel tumore al seno

La sensibilità particolarmente elevata delle immagini MRI ha fatto si che negli ultimi anni questa affiancasse la mammografia in casi specifici quali donne con seno denso, o con mutazioni BRCA, in cui la TC ha difficoltà di rilevare chiaramente il tumore, o si hanno delle incertezze, soprattutto nei casi di tumore invasivo. Inoltre, l'elevata sensibilità nel rilevare anche i più precoci cambiamenti nel metabolismo intracellulare rende l'MRI un'ottima tecnica di monitoraggio del trattamento. [12]

L'MRI statica consente dunque di distinguere con nitidezza il tessuto fibro-ghiandolare dal tessuto adiposo tuttavia i valori di specificità risultano molto variabili, con elevato numero di falsi positivi, non permettendo di discernere con certezza le lesioni benigne dalle maligne; inoltre, studi comparativi tra mammografia e MRI riportati da Morrow et al. non hanno notato particolari differenze nella detezione dei tumori non invasivi DCIS o significativi vantaggi sulla riduzione del numero di operazioni chirurgiche o sul miglioramento del trattamento : la migliore sensibilità dell'MRI sembra non compensare le problematiche legate ai costi superiori e alle più complesse modalità di acquisizione.[12]

L'MRI dinamica (DCE MRI), affermatasi negli ultimi anni, ha prestazioni migliori, perché consente di rilevare anche caratteristiche funzionali del tumore: il paziente viene sottoposto ad una sequenza pesata T_1 sia prima che dopo l'iniezione di gadolinio: misurando la velocità di eliminazione dell'agente di contrasto è possibile individuare i tumori maligni, per i quali i rates di assorbimento ed eliminazione delle sostanze sono alterati a causa dell'ipervascolarizzazione tipica dei tessuti patologici. [13]

La DCE MRI è attualmente la tecnica di risonanza magnetica più utilizzata: essa garantisce una elevata sensibilità, anche migliore della mammografia, permettendo la detezione di tumori di piccola dimensione o non-massivi, tuttavia il numero di falsi positivi rimane ingente e si è notato che le fluttuazioni ormonali influenzano la capacità di assorbimento tissutale del gadolinio: ciò dimostra che ci sono dei limiti di utilizzo. Inoltre, l'impiego del gadolinio potrebbe costituire un problema a lungo termine perché è tossico per l'organismo, quindi si sta proseguendo la ricerca per individuare altri mezzi di contrasto altrettanto performanti.

La tipica procedura di DCE MRI consiste in una acquisizione pre-contrasto e cinque post contrasto, dunque successive all'iniezione del gadolinio. Solitamente si utilizzano sequenze di eccitazione che permettano di sopprimere il segnale derivante dal tessuto adiposo (fat suppression), come la "short T1 inversion recovery", per la quale quest'ultimo assume un colore scuro. Una volta acquisite le immagini a diversi time frame, si effettua il post- processing. Il primo step è la generazione di immagini sottratte, date dalla differenza di ognuna delle acquisizioni post contrasto con la pre-contrasto: in questo modo si valuta l'andamento del mezzo di contrasto nel tempo, facendo emergere la possibile lesione, che assumerà un colore chiaro a causa dell'accumulo di gadolinio. La "region of interest" (ROI) viene successivamente segmentata ed analizzata.

Secondo il lessico BI-RADS, a partire dal risultato della RM, bisogna classificare morfologicamente il tumore come lesione concentrata, massiva o non massiva. Le lesioni concentrate hanno dimensioni inferiori a 5mm quindi risultano difficili da caratterizzare; Una lesione massiva occupa uno spazio 3D ed è possibile descriverne forma e margini, al contrario una lesione non-mass non occupa uno spazio tridimensionale definito, dunque risulta complesso definirne i bordi e le caratteristiche interne. [14]

Ci sono altre tecniche emergenti basate sull'MRI, che cercano di superarne le problematiche. La diffusion-weighted MRI (DWI) misura il coefficiente di diffusione dell'acqua, che si muove di moto browniano all'interno dei tessuti: più il moto sarà limitato, più il tessuto avrà un livello di densità cellulare elevato. Nella pratica clinica si sta diffondendo la MRI multiparametrica, che include la DCE MRI con la DWI in modo da ottenere entrambe le informazioni. Un'altra tecnica promettente è la spettroscopia con tecnica di risonanza magnetica (MRS), considerata una "biopsia non invasiva", grazie alla capacità di caratterizzare biochimicamente il tessuto. Essa risulta dunque molto utile alla diagnosi differenziale delle lesioni tumorali [14]; un esempio è la spettroscopia protonica, che rileva i livelli di componenti cellulari contenenti colina: una quantità significativa è indicatrice di elevato turnover di membrana, marker di tumore maligno.

3 Stato dell'arte

Sistemi CAD

L'elaborazione e l'analisi di una risonanza magnetica richiede l'acquisizione di molte immagini e notevole tempo di processing da parte del radiologo, il quale ad oggi ha il compito di individuare e segmentare manualmente la lesione tumorale, dunque descriverne la morfologia, ma anche caratterizzarla dal punto di vista clinico-patologico.

L'interpretazione delle immagini nella attuale pratica clinica è quindi operatore dipendente: presupponendo che il clinico sia una persona esperta, è ugualmente evidente la necessità di avere un supporto informatico che possa fornire al radiologo un valido aiuto nella valutazione della lesione. Ciò accorcerebbe notevolmente i tempi di analisi, e renderebbe i risultati più oggettivi, meno soggetti all'interpretazione del singolo, dunque anche più riproducibili, e possibilmente meno affetti da errore.

Da oltre sessant'anni la ricerca sull'intelligenza artificiale (AI) produce notevoli risultati, ed è diventata un potente supporto in diversi ambiti, compresa la medicina. Nella clinica l'obiettivo è quello di estrarre e sintetizzare informazioni dall'imaging, in modo da fornire un valido contributo ai clinici in più ambiti, quali la diagnosi del cancro, il monitoraggio e la predizione della risposta al trattamento, permettendo di ottimizzare tempi impiegati e risultati ottenuti.

La ricerca si sta concentrando principalmente sulla creazione di Computer Detection Detection/diagnosis (CAD), ovvero sistemi che consentono di individuare automonamente la lesione (CADe) o sistemi che la caratterizzano, classificandola come maligna o benigna (CADx). Queste strutture si basano sul riconoscimento di alcune features peculiari delle lesioni, che consentono di distinguerle rispetto ai tessuti circostanti ma anche di caratterizzarle al meglio, in modo da poter effettuare una diagnosi.

Machine Learning vs Deep Learning

All'interno della AI sono distinguibili due approcci di pattern recognition: il più tradizionale Machine Learning e il Deep Learning. Le tecniche di machine learning, già

introdotte in ambito medico, non possono processare direttamente le immagini: è necessario infatti un passaggio precedente, di estrazione delle features, che sono predefinite, ovvero costruite manualmente; le caratteristiche, e.g. relative alla morfologia o alla texture del tumore, descrivono gli aspetti radiologici del tumore: l'estrazione di informazioni quantitative dalle lesione dall'analisi dell'imaging è ciò di cui si occupa la "Radiomica". Una volta estratte le features, si scelgono quelle che riescono a discriminare meglio il tumore rispetto al tessuto sano, seguendo il criterio di minimizzazione dell'errore di classificazione. Talvolta però features "ingegnerizzate", che risultano possibilmente più intuitive, non sono sufficientemente robuste per garantire generalizzazione nel processo di detezione tumorale. [15]

Il deep learning supera questo limite, perché le features da estrarre non sono stabilite preliminarmente, ma derivano direttamente dalla fase di learning, nella quale la rete ricava un vettore di features dalle immagini: questo è utilizzato come input da un classificatore, che procede al riconoscimento di pattern caratteristici dell'immagine. Il vettore di caratteristiche rappresenta dunque un optimum, perché appreso direttamente durante l'apprendimento della rete, e racchiude tutte le informazioni sia locali che globali che sono "nascoste" nelle immagini. A differenza delle features del machine learning, spesso le caratteristiche estratte non riescono ad essere espresse tramite formule matematiche che possano descriverne l'andamento, perciò risultano difficilmente interpretabili.

Il raggiungimento di prestazioni ottimali e di un valido vettore di features necessita dataset molto ampi ed elevata potenza di calcolo: ecco perché lo sviluppo di questa tecnologia è relativamente recente e in piena fase di sperimentazione. L'utilizzo di dataset ridotti infatti, aumenta il rischio di overfitting: la rete impara "troppo" dai dati in input, diventando eccessivamente specifica e dunque incapace di generalizzare; per cercare di ovviare, possono essere utilizzati metodi di data augmentation, applicando trasformazioni ai dati.

Una delle strutture della AI con maggiore impiego è la ANN, Artificial Neural Network, che imita il processamento dell'informazione che compie il nostro cervello: essa si compone di layers caratterizzati da un certo numero di neuroni collegati tra loro attraverso dei pesi, che rappresentano le sinapsi cerebrali; il primo layer riceve le informazioni in ingresso, che vengono poi processate nei livelli successivi fino ad arrivare al livello di output in cui viene decretata la classe di appartenenza del dato in ingresso. Il layer di input ha un numero di neuroni pari al numero delle features che descrivono il dataset in input, mentre il layer in uscita ha quanti neuroni quante sono le categorie che devono essere apprese. La rete "apprende" a classificare nella fase di allenamento, nella quale i pesi vengono modificati ed adattati progressivamente ad ottimizzare l'output, secondo un processo di minimizzazione dell'errore tra output effettivo e output "desiderato". Il multilayer perceptron è uno dei modelli di rete neurale più popolare, ed basato su un numero fissato di layers e neuroni e su features predefinite. Si sono successivamente definiti i modelli di rete basati sul deep learning, con lo stesso principio di funzionamento ma con la capacità di estrarre autonomamente i pattern caratteristici. [16]

Convolutional Neural Networks (CNN)

La struttura di deep learning più comune è la Convolutional Neural Network, il cui principale componente è il filtro convoluzionale: la convoluzione è un'operazione lineare che consiste nella moltiplicazione di una matrice chiamata kernel di dimensioni NxN con una finestra del tensore, ovvero il vettore di input, di uguale dimensione. Il kernel dunque scorre lungo l'immagine con un passo (stride) prestabilito, ed i valori risultanti dal prodotto elemento per elemento vengono sommati, in modo da costruire progressivamente la feature map, dunque l'output di un livello di convoluzione. Questo processo è il primo passo per l'estrazione delle features dall'input: si è notato infatti che le immagini hanno dei gruppi di valori altamente correlati tra loro, che formano un pattern distintivo, il quale spesso si ripete in diverse parti dell'immagine.

Dopo la convoluzione, che mette in evidenza il pattern, si ha un livello di "pooling", che riduce la dimensione dell'immagine, comprimendo le features simili in una unica, in modo da rendere il risultato invariante rispetto a piccole distorsioni o traslazioni delle immagini. Spesso il criterio di pooling utilizzato è quello del massimo, ovvero considerando finestre di dimensione MxM della feature map, si seleziona solo il valore massimo, riducendo quindi le dimensioni della finestra ad un unico pixel (Fig.2).



Le feature maps ottenute dall'ultimo output dei blocchi di convoluzione sono trasformate in vettori 1D e connesse ad uno o più "strati densi", ovvero nei quali ogni input è connesso ad un output tramite dei pesi, che si modificano durante l'allenamento. L'ultimo strato completamente connesso rappresenta proprio il layer di output, che possiede un numero di nodi pari al numero di classi che la rete deve individuare. Durante il processo di training, la rete modifica i parametri "allenabili" dunque pesi e kernel secondo un algoritmo di ottimizzazione il cui scopo è quello di minimizzare la loss function [17]. La tipica struttura di una rete neurale è mostrata in Fig.3.

Più la rete è strutturata, quindi con un maggior numero di layers nascosti, più le caratteristiche estratte saranno specifiche, cioè aumenterà il livello di astrazione. Le reti deep learning infatti si basano su un principio gerarchico, per il quale le features di alto livello derivano dalla composizione di features più elementari: "combinazioni locali di

valori formano motivi, che formano parti, le quali assemblano parti e le parti creano oggetti" come affermano Yann LeCun et al. [18]



Il deep learning applicato ai dati di imaging trova impiego in molti settori della medicina, tra cui la segmentazione del seno e la rilevazione della lesione tumorale, come già accennato, ma risulta utile anche nel monitoraggio e nella previsione della risposta al trattamento, essendo capace di individuare i biomarkers del tumore. Un esempio di questa applicazione è lo studio condotto da K.Ravichandran et al., che hanno utilizzato una rete neurale convoluzionale per predire la risposta alla chemioterapia neoadiuvante (NAC), una terapia che viene utilizzata precedentemente all'operazione, che dovrebbe ridurre la dimensione della lesione e migliorare la risposta post chirurgica: purtroppo alcuni soggetti non traggono beneficio da questo tipo di cura quindi conoscere in anticipo quale sarà l'evoluzione del tumore dopo la terapia potrebbe aiutare ad individuare i soggetti più adatti a riceverla e far risparmiare costi e tempo utile a coloro per cui la NAC sarebbe inefficace [19].

La costruzione di un CAD completamente automatico che riesca a rilevare la massa tumorale a partire da imaging è un problema persistente, al quale la ricerca sta dedicando molte risorse. Un'acquisizione MRI del seno include altri organi quali i polmoni, il cuore, il fegato e i muscoli pettorali, dunque alcuni ritengono che al fine di fare una detezione specifica della massa tumorale, sia necessario isolare preliminarmente la zona delle mammelle. Metodi tradizionali di segmentazione possono basarsi su approcci pixel per pixel, nei quali ogni voxel è paragonato ad una soglia di intensità stabilita, o sul confronto con le referenze anatomiche ricavate dagli atlanti. Le maschere così create sono poi processate anche attraverso considerazioni geometriche che le rendono più aderenti alla realtà. [20]

La segmentazione automatica del seno rappresenta una sfida a causa della variabilità anatomica delle mammelle tra i diversi soggetti (densità, forma e grandezza dei seni e del muscolo pettorale), oltre che alla presenza di rumore e artefatti, quali effetti di alising o inomogeneità nell'intensità, nelle immagini acquisite [15]. Jiao et al. hanno implementato un particolare tipo di rete convoluzionale, la Unet ++, allenata con 50 epoche e con ottimizzatore Stochastic Gradient Descent (SGD); le prestazioni sono state valutate in termini di Dice coefficient, Jaccard coefficient, una misura di similarità, e sensibilità: i risultati sono stati rispettivamente 0.951, 0.908, 0.948, migliori rispetto a quelli di una Unet classica con le stesse caratteristiche. L'output della Unet ++ è stato utilizzato per la detezione della massa tumorale attraverso una Faster RCNN, allenata utilizzando le slices in cui era presente il tumore, marcando il punto centrale, la profondità e l'altezza della lesione. La Faster RCNN si compone di due moduli: una rete CNN che propone le regioni da segmentare e un detettore che le testa e seleziona le più plausibili. La sensibilità media di riconoscimento della lesione è stata di 0.874, un risultato molto promettente che secondo gli autori è stato raggiunto grazie alla rimozione degli organi di "interferenza" nel pre processing delle immagini. [21]

Unet

La tipologia di rete che rappresenta lo stato dell'arte del processo di segmentazione è un tipo di rete convoluzionale, chiamata U-net. Essa è caratterizzata da una struttura simmetrica, con una parte di contrazione dell'input e una seconda parte di espansione, come mostrato in Fig.4. Il primo modulo corrisponde quindi ad una classica rete convoluzionale, che riduce le dimensioni dell'immagine ed aumenta le caratteristiche

estratte: si alternano livelli convoluzionali, seguiti da normalizzazione e funzione di attivazione, e livelli di pooling. La parte simmetrica invece fa un'operazione di up-sampling, ovvero riporta la feature map ad alta risoluzione ottenuta dal primo modulo alle dimensioni delle immagini in input; in uscita dunque la U-net genera una maschera di segmentazione in cui ogni pixel è classificato nella sua classe di appartenenza e correttamente localizzato. [16], [21], [22].

L'architettura particolare della U-net consente prestazioni elevate nella segmentazione: essa fornisce risultati robusti anche utilizzando dataset limitati e consente di mantenere e rendere più precisa l'informazione spaziale combinando il processo di up sampling con le features ad alta risoluzione corrispondenti a quel livello.

La tecnica MRI dinamica, o anche la multiparametrica, producono delle immagini dense di informazioni, sia spaziali che temporali [21], che sono colte al fine di una migliore detezione e diagnosi della lesione: tuttavia analizzare una così ricca quantità di immagini, pre e post contrasto a diversi step causa elevati costi e lunghi protocolli di acquisizione. Al fine di facilitare ed abbreviare i protocolli MRI, Mehmet et. al hanno utilizzato una Unet 2D (Fig.5) sia per segmentare la zona del seno che per individuare le lesioni tumorali basandosi esclusivamente sull'immagine pre-contrasto e sulla prima post-contrasto (early-phase scan), in modo da testare le prestazioni di un CAD su immagini meno dense di informazioni . E' stato calcolato il Relative Enhancement Volume (RE), con la formula $RE = \frac{I_1 - I_0}{I_0}$, in cui II e IO sono rispettivamente le intensità delle MRI post e pre-contrasto. La rete utilizzata è stata allenata utilizzando sia soggetti sani che malati, in particolare utilizzando dei lotti di 40 slices, di cui metà contenenti la lesione e metà appartenenti ai pazienti sani. Nella fase successiva si è voluto ridurre il numero di falsi positivi: le lesioni individuate sono state date in input ad una rete convoluzionale 3D, formata da 3 strati densi caratterizzati da 5x5x5 filtri e seguiti da ReLu; la rete classifica basandosi sui volumi RE, dai quali si estraggono informazioni morfologiche nella regione circostante la lesione candidata ma anche sulla asimmetria di contrasto tra i due seni. Le prestazioni sono state valutate utilizzando la Computation Performance Metric (CPM) [23], che è risultata pari a 0.6429, superiore a quello ottenuto da un CADe basato su protocollo MRI completo, utilizzato per confronto; è da evidenziare inoltre che le

informazioni sulla asimmetria di contrasto tra i seni ha migliorato notevolmente le prestazioni.[24]

Un vantaggio delle U-net, oltre a quelli sovra citati è che è possibile utilizzare immagini intere di qualsiasi dimensione per il training, senza la necessità di dividerle in sub-regioni; di solito invece l'approccio "patch- based" è molto utilizzato nel deep learning. Questa possibilità fa sì che la rete abbia una percezione più realistica dell'input, senza che l'informazione fornita sia in qualche modo limitata dalla dimensione inferiore delle sezioni in cui l'immagine viene divisa. Tuttavia questa possibilità potrebbe portare ad altissimi costi computazionali. Dalmis et al. hanno confrontato le prestazioni di due U-net per la classificazione del tessuto adiposo, fibroghiandolare e del tessuto non appartenente al seno. La struttura delle U-net è uguale (Fig.5) ma la prima ha due classi di output, ovvero FGT/Adiposo, mentre la seconda distingue tutti e tre i tessuti.



Le prestazioni, misurate in termini di DSC coefficient, hanno mostrato come l'approccio U-net funzioni meglio degli approcci tradizionali basati su soglie o su atlanti anatomici, e che, tra le due U-net, quella caratterizzata da 2 classi in uscita abbia prestazioni leggermente migliori: il risultato è giustificato perché la rete aveva un solo obiettivo di classificazione, mentre l'altra due. [25]

Nonostante la natura 3D della Risonanza Magnetica, la tecnica attuale è quella di applicare la U-net ad ogni slice e poi ricostruire il volume. U-Net 3D iniziano ad essere implementate, come nell'articolo pubblicato nel 2021 da Cao et al., nel quale è descritto l'utilizzo di una U-net 3D per la detezione e segmentazione automatica di metastasi al cervello da immagini MRI. La rete ha una struttura asimmetrica, con due moduli di down sampling e un modulo di up sampling, con kernel rispettivamente di 3x3x3 e 1x1x3: le lesioni (<7mm) sono circondate da una grande quantità di tessuto sano, per questo il kernel con dimensione inferiore riesce a cogliere meglio le features delle metastasi, perché meno distratto dal tessuto circostante. Sono confrontate 6 diverse configurazioni della rete, di cui 4 con un unico kernel per entrambi i moduli "contrattivi", di dimensioni diverse, e una che viene definita Vanilla U-net, con struttura simmetrica ma stessi parametri di allenamento. Le prestazioni migliori sono state quelle della rete implementata nello studio, quindi asimmetrica con due differenti kernel, raggiungendo il 100% di sensibilità per lesioni maggiori di 6 mm, con 1.1 di rapporto fasi positivi per paziente [26]. Ciò dimostra risultati molto promettenti nell'utilizzo di questa tipologia di rete, ma non sono stati trovati esempi in letteratura che riguardassero il tumore al seno, argomento di questo progetto di tesi.

La U-net originaria implementata da Ronneberger et al. è stata modificata e potenziata nella versione U-net ++ (Fig. 6): oltre alla classica struttura ad U infatti sono stati inseriti dei livelli "ponte", mostrati in verde, delle connessioni dense, rappresentati in blu, e un modulo di supervisione, in rosso. I nuovi livelli hanno lo scopo di ridurre il gap semantico tra le feature maps corrispondenti dei moduli di down e up sampling. Collegarle fa si che siano ricevute feature maps simili, facilitando il processo di ottimizzazione. I link in blu permettono un collegamento diretto tra le due parti della rete, in modo da migliorare l'accuratezza di segmentazione, perché si assicurano che tutte le feature maps create arrivino al nodo corrente, garantendone la massima risoluzione. Il modulo di supervisione è infine inserito per controllare il livello di complessità, dunque per bilanciare la performance con la velocità di processo. [27]



Fig. 6 Struttura U-net++ [27]

4 Materiali e Metodi

4.1 Dataset

4.1.1 Origine dei dati

I pazienti considerati in questo progetto di tesi sono 72 donne affette da tumore al seno che si sono sottoposte a MRI di controllo. I dati provengono da due differenti strutture ospedaliere: 32 pazienti dall'ospedale Molinette di Torino e 40 pazienti provenienti dall'IRCCS di Candiolo.

Su un totale di 72 pazienti, 20 hanno età superiore ai 60 anni, e il range di età varia tra i 35 e i 77 anni, con prevalenza di donne sui 50 anni, categoria più rischiosa anche a causa della prevalenza di tessuto fibroghiandolare che rende il seno estremamente denso.

Le dimensione delle immagini mediche sono differenti tra i due gruppi: quelle acquisite all'IRCCS di Candiolo hanno dimensione fissa 512x512, mentre le MRI provenienti dalle Molinette sono caratterizzate da dimensioni variabili, anche se nella maggior parte dei casi sono 352x352. Si tratta principalmente di MRI assiali, anche se all'interno dei dataset del Candiolo erano presenti anche immagini con proiezione sagittale. Le immagini sono state acquisite tra i 7 ed i 14 giorni del ciclo mestruale, nel caso di donne premenopausa, utilizzando, per i pazienti provenienti dalle Molinette, uno scanner da 1.5 T a 7 canali con una sequenza assiale STIR e pesata T1 non fat sat combinata con una DWI; successivamente è stata effettuata una DCE-MRI 3D nella quale sono state acquisite 1 immagine pre contrasto e 5 post-contrasto con una sequenza T1 echo con fat suppression.

Nel caso di Candiolo il magnete utilizzato è un 1.5 T con 8 canali: precedentemente sono state acquisite immagini sagittali pesate in T2 e DWI assiali e lo studio è stato poi completato con DCE-MRI 3D assiale, con sequenza pesata in T1 e con fat suppression, che ha generato 1 immagine pre- contrasto e 5 post somministrazione dell'agente di contrasto.

Si è deciso di creare un dataset unico, in modo da allenare la rete nella maniera più generica possibile: i tumori sono principalmente di tipo non mass, anche se con varietà nella morfologia, e le MRI nelle due strutture sono state ottenute tramite protocolli di acquisizione differenti ed anche con diverse proiezioni.

Per ogni paziente si avevano in possesso diverse acquisizioni, di cui una pre-contrasto e le altre post-contrasto a differenti time frame, e l'immagine sottratta, ottenuta dalla differenza della terza immagine post-contrasto, in cui l'assorbimento del mezzo di contrasto era massimo, con la baseline. Inoltre, per ogni paziente era presente la maschera della lesione, segmentata manualmente da una radiologa esperta.

4.1.2 Pre processing dei dati

L'allenamento della rete è stato effettuato fornendo in input l'immagine sottratta e la maschera manuale come class label. Preliminarmente sono stati effettuati dei controlli sulle immagini: sono stati corretti alcuni errori di orientamento e si è verificato che tutte le maschere manuali avessero solo valori binarizzati, quindi pari a 1 per i pixel appartenenti alla lesione e 0 per lo sfondo. Inoltre, si è deciso di riportare a 0 tutti i pixel dell'immagine tissutale che avevano valori inferiori a 0: questi infatti, analizzando l'immagine, costituivano rumore o più genericamente non rappresentavano la lesione.

È stato scelto un approccio patch-based, ovvero sia immagine MRI che maschera sono state suddivise in ROI di dimensione 16x16 non sovrapposte; per consentire ciò, si è effettuato zero padding sulle immagini, aggiungendo pixel di valore pari a 0 in modo che la dimensione fosse correttamente divisibile per 16.

Considerando le maschere manuali, sono state selezionate per il training della rete soltanto le ROI che avessero almeno 5 pixel tumorali, ovvero con valore pari a 1, a cui ovviamente sono state abbinate le corrispondenti ROI del tessuto.

Seguendo i criteri sopra riportati, sono state ottenute 6771 ROI, memorizzate mantenendo l'informazione sul paziente e la slice di appartenenza, e le coordinate spaziali della ROI all'interno dell'immagine.

4.1.3 Suddivisione Dataset

La composizione di Construction Set e Validation Set è avvenuta considerando una suddivisione per paziente: il 70% dei soggetti ha costituito il construction set e il rimanente 30% il validation. Dunque 50 pazienti sono stati utilizzati per costruire la rete, mentre i rimanenti 22 per testarne la performance.

All'interno del costruction set, si è ripetuta la suddivisione 70-30% per costituire il training set ed il test set. Il training set rappresenta il dataset utilizzato per l'addestramento della rete, ed è costituito da 35 pazienti, invece il test set è composto dai rimanenti 15 pazienti ed è necessario per l'aggiornamento dei pesi della rete neurale durante l'allenamento.

La suddivisione dei pazienti tra i tre diversi set è avvenuta in maniera del tutto casuale, tramite codice Python, ponendo attenzione tuttavia a mantenere una corretta proporzione tra i due dataset in modo da creare Train, Test e Validation equilibrati con pazienti provenienti da entrambe le strutture ospedaliere.

4.2 Struttura della rete

La rete neurale che è stata implementata è una U-net, considerata la tipologia di rete neurale più adatta a scopi di segmentazione. Il codice è stato implementato utilizzando l'API Keras in Python 3.8 configurata con la libreria Tensor Flow, dedicata al Deep Learning.

Schema della rete

L'architettura della rete è riportata in Fig. 5: essa è costituita da una prima parte di "down sampling" che riporta la struttura classica di una rete CNN: un livello convoluzionale, seguito da un livello di maxpooling, che riduce la dimensione dell'input.

Il livello convoluzionale è costituito da una serie di operazioni: l'input è convoluto con il filtro, successivamente viene normalizzato tramite normalizzazione Botch, che consolida i gradienti instabili presenti nei layers nascosti, e infine viene utilizzata la funzione di attivazione ReLu (Rectified Linear Activation Unit), che lascia invariati gli output positivi e pone a 0 i rimanenti. Nella rete implementata sono presenti due blocchi

convoluzionali, e sono stati impostati 16 filtri. Corrispondentemente ai due blocchi sono presenti dunque due livelli di MaxPooling, che riducono l'input, ROI 16x16, ad una dimensione di 4x4. Il secondo modulo della rete è esattamente speculare al primo, con blocchi convoluzionali e livelli corrispondenti delle due parti, in modo di creare strati densi, per un totale di 130833 parametri. L'ultimo layer è convoluto in modo da dare in output una matrice di dimensione 16x16x1 ed ha funzione di attivazione Sigmoide. L'ottimizzatore scelto è Adam e le loss function testate si sono basate sulla Binary Cross Entropy e il coefficiente DSC.



4.3 Modalità di allenamento

Mantenendo la struttura della rete fissa, sono state implementate quattro modelli di rete che si differenziano a causa delle diverse normalizzazioni delle ROI di input. Normalizzare i dati è estremamente importante perché variabili misurate a scale diverse non contribuirebbero equamente all'allenamento della rete e potrebbero creare dei bias, facendo predominare alcune features su altre. Dunque è importante riportare tutti i valori compresi tra 0 e 1. Per farlo, si è accorsi a due differenti normalizzazioni:

- Z-score standardization, che si comporta meglio con gli outliers anche se non produce dati normalizzati nella stessa scala esatta; questo metodo utilizza media e deviazione standard dei valori secondo la formula <u>x-mean(x)</u> <u>std(x)</u>. Per calcolare la media e la deviazione standard sono stati utilizzati esclusivamente i pixel con intensità diversa da 0 in modo da non creare bias.
- Entrambe le normalizzazioni sono state effettuate considerando massimo, minimo, dev.standard e media del paziente o della singola slices, per un totale di 4 reti differenti:



Qual è la normalizzazione con la quale si ottengono migliori risultati, a parità di struttura della rete e con gli stessi dati in input? Come le diverse normalizzazioni influiscono sulle prestazioni della rete? Queste sono le domande a cui si è cercato di rispondere.

Successivamente, è stato allenato un quinto modello di rete, variando la loss function: è stata costruita una loss function basata sul Dice coefficient, in modo da valutare anche in che modo questa condizionasse i risultati della rete.

5 Risultati

5.1 Metodi di valutazione dei risultati

Al fine di valutare e confrontare i modelli implementati, sono stati calcolati i parametri di Precision, Recall e Dice coefficient.

Confrontando i pixel predetti con i reali, si compone la confusion matrix, contando il numero di veri positivi, falsi positivi, veri negativi e falsi negativi. Per chiarezza, viene considerato positivo un pixel che ha valore 1, che è quindi classificato come tumorale, mentre i pixel negativi hanno valore 0 e sono classificati come appartenenti a tessuto sano. Il conteggio di TP,FP,TN,FN è effettuato sia considerando le singole fette, che per ogni paziente, in modo da capire se la rete riesce a distinguere tra lesioni appartenenti a pazienti diversi.

La Precision è una misura di accuratezza, espressa dal rapporto tra il numero delle previsioni positive che corrette che ha avuto il modello e il numero dei positivi totali predetti dalla rete.

$$Precision = \frac{TP}{TP + FP}$$

Dunque un modello è preciso quando ci sono pochi falsi positivi: nel caso dell'obiettivo di questa tesi, avere una precision alta significa che un pixel individuato come positivo sia davvero positivo, ovvero aver individuato selettivamente la lesione, senza avere altre zone che la rete classifica come tumorali.

La Recall rappresenta invece la percentuale di positivi riconosciuti correttamente.

$$Recall = \frac{TP}{TP + FN}$$

Questo parametro misura dunque la sensibilità del modello, la sua capacità di individuare la lesione tumorale. Una recall bassa significa che ci sono molti falsi negativi, ovvero che la rete non è riuscita a classificare correttamente i pixel tumorali, scambiandoli per appartenenti a tessuto sano.

Il DSC, Dice Coefficient riassume le due informazioni, essendo infatti una misura di similarità, ovvero quantifica il livello di sovrapposizione tra la maschera predetta dalla rete e la maschera originale.

$$DSC = \frac{2 * TP}{2 * TP + FP + FN}$$

5.1 Normalizzazione dei dati di input

Preliminarmente, le immagini costituenti il dataset dovevano essere normalizzate. Si è scelto di utilizzare due normalizzazioni differenti per verificare che tipo di effetto potesse avere la normalizzazione dei dati sulla rete. Entrambe le normalizzazioni sono state effettuate considerando i valori di ogni fetta oppure sui valori dei singoli pazienti ai quali le slices appartenevano.

Z-score



Fig. 7 Andamento rete normalizzata per fetta tramite z-score standardization



Fig. 8 Andamento rete normalizzata per paziente tramite z-score standardization





Fig. 9 Andamento rete normalizzata per fetta tramite min max scaling



Fig. 10 Andamento rete normalizzata per paziente tramite min max scaling

Sono stati salvati i best models, raggiunti dopo poche epoche: la rete ha mostrato dunque un discreto overfitting, probabilmente dovuto al basso numero di filtri utilizzati. Si nota però come la z-score normalization continui ad imparare fino alla 12/13 epoca mentre il min- max scaling arriva al best model già alla quinta epoca per la rete normalizzata per paziente.

5.2 Binarizzazione delle immagini predette

Una volta allenata la rete, ne sono state valutate le prestazioni su training e test set ed infine sul validation set. Le maschere calcolate hanno valori decimali compresi tra 0 e 1 quindi devono essere binarizzate; anche per questa operazione, si è scelto di testare differenti binarizzazioni in modo da confrontarne gli effetti.

Soglia round

La prima soglia testata è stata 0.5, ottenuta tramite la funzione di approssimazione round, tenuta fissa per tutti i set di dati.

Soglia di Otsu

Il metodo di Otsu è una modalità di binarizzazione automatica, che sceglie la soglia più adatta per ogni immagine, nel caso di questo progetto di tesi per ogni roi, secondo un criterio di minimizzazione della varianza intra classe tra "tessuto sano" e "tessuto tumorale": non si ha dunque una soglia generica ma una per ogni ROI predetta.

Soglie della curva ROC

Le ultime soglie applicate vengono ricavate sulla base della curva ROC, ovvero una tecnica statistica che misura l'accuratezza di un test diagnostico: sulle ascisse si ha la specificità, mentre sulle ordinate la sensibilità della rete.

A partire da questa curva, si decide di scegliere una soglia che ottimizzi un determinato parametro, utilizzando i dati di training set ed applicando poi la soglia stabilita ai rimanenti set.

In questo studio è stato scelto di ottimizzare la specificità, ovvero:

$$SPEC = \frac{VN}{VN + FP}$$

essa rappresenta la proporzione di veri negativi predetti dalla rete rispetto al numero totale di negativi, ovvero di soggetti sani;

un altro parametro che si è scelto di ottimizzare è il Positive Predictive Value, il quale è equivalente alla precision, dunque numero reale di pixel positivi, ovvero tumorali, rispetto a tutti i positivi che sono stati predetti [28];

infine si è utilizzato lo Youden Index, che è il punto di best cut-off della curva ROC, e dunque ottimizza il rapporto tra sensibilità e specificità.

Youden Index = Sensibilità + Specificità
$$-1$$

Anche le diverse binarizzazioni sono state calcolate per singola fetta o per paziente.

5.3 Calcolo dei parametri per ogni modello

Min max Scaling per fetta



Fig.11: 11a:Andamento di specificità, sensibilità 11b: Andamento del PPV 11c: Andamento dello Youden Index

Tab.1 Andamento delle diverse binarizzazioni. 1a Training set 1b Test set 1c Validation set

FETTA																					
		S	OGLIA ROUN	D			SOGLIA OTSI	J		SC	GLIA ppv=0,	46		SOC	GLIA yuden=0	0,02		SO	GLIA spec=0	,78	
			train				train				train				train				train		
	recall		precision	DSC		recall	precision	DSC		recall	precision	DSC		recall	precision	DSC		recall	precision	DSC	
MEDIA		0,348	0,830		0,475	0,779	0,536		0,630	0,382	0,845		0,508	0,533	0,751		0,609	0,321	0,862		0,450
DEV, STD		0,139	0,164		0,154	0,123	0,098		0,098	0,159	0,139		0,165	0,174	0,125		0,151	0,151	0,161		0,169
0 esclusi	si													0,539	0,760		0,616	0,326	0,876		0,458
														0,166	0,097		0,138	0,146	0,116		0,160
	SOGLIA ROUND						SOGLIA OTSI	J		SC	GLIA ppv=0,	46		SOC	GLIA yuden=0	0,02		SO	GLIA spec=0	,78	
			train				train				train				train				train		
	recall		precision	DSC		recall	precision	DSC		recall	precision	DSC		recall	precision	DSC		recall	precision	DSC	
MEDIA		0,331	0,785		0,451	0,784	0,511		0,611	0,371	0,821		0,489	0,521	0,705		0,582	0,309	0,821		0,432
DEV, STD		0,140	0,222		0,163	0,129	0,101		0,100	0,166	0,170		0,185	0,197	0,153		0,180	0,152	0,225		0,185
0 esclusi										0,381	0,845		0,504	0,536	0,726		0,599	0,328	0,870	÷	0,459
										0,155	0,097		0,167	0,178	0,095		0,152	0,136	0,101		0,156

Tab 1a

FETTA															
	SO	GLIA ROU	ND	S	DGLIA OTS	U	SO	GLIA ppv=0	0,46	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,78
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,314	0,580	0,373	0,680	0,556	0,577	0,112	0,153	0,112	0,166	0,187	0,143	0,094	0,155	0,099
DEV, STD	0,286	0,381	0,310	0,233	0,165	0,165	0,141	0,164	0,119	0,174	0,190	0,135	0,129	0,174	0,110
0 esclusi							0,140	0,191	0,140	0,198	0,223	0,171	0,119	0,196	0,125
							0,145	0,162	0,118	0,172	0,187	0,131	0,134	0,174	0,110
	SOGLIA ROUND			S	DGLIA OTS	U	SO	GLIA ppv=0	0,46	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,78
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,362	0,703	0,464	0,726	0,533	0,602	0,119	0,158	0,128	0,186	0,158	0,162	0,098	0,157	0,113
DEV, STD	0,216	0,311	0,246	0,192	0,109	0,133	0,076	0,095	0,078	0,115	0,095	0,096	0,066	0,098	0,073
0 esclusi							0,127	0,169	0,137	0,199	0,170	0,174	0,105	0,168	0,122
							0,072	0,088	0,073	0,107	0,088	0,088	0,063	0,091	0,069

Tab 1b

FETTA															
	SO	GLIA ROU	ND	S	OGLIA OTS	SU	SO	GLIA ppv=(),46	SOG	LIA yuden:	=0,02	SOC	GLIA spec=	0,78
		validation			validation			validation			validation	l.		validation	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,345	0,725	0,454	0,775	0,490	0,593	0,055	0,164	0,073	0,173	0,254	0,194	0,046	0,163	0,064
DEV, STD	0,120	0,145	0,123	0,108	0,103	0,098	0,070	0,168	0,088	0,090	0,144	0,092	0,060	0,176	0,080
0 esclusi							0,065	0,194	0,086	0,175	0,257	0,197	0,056	0,197	0,077
							0,072	0,166	0,090	0,088	0,142	0,090	0,062	0,175	0,082
	SOGLIA ROUND			S	OGLIA OTS	SU	SO	GLIA ppv=(),46	SOG	LIA yuden:	=0,02	SOC	GLIA spec=	0,78
		validation			validation			validation			validation	l.		validation	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,292	0,684	0,397	0,743	0,464	0,564	0,041	0,137	0,052	0,158	0,223	0,177	0,036	0,137	0,048
DEV, STD	0,110	0,219	0,129	0,103	0,128	0,121	0,049	0,164	0,058	0,078	0,112	0,083	0,043	0,169	0,054
0 esclusi							0,050	0,167	0,063	0,158	0,223	0,177	0,044	0,168	0,058
							0,049	0,167	0,058	0,078	0,112	0,083	0,043	0,172	0,054



Min max Scaling per paziente



12c: Andamento dello Youden Index

Fab.2 Andamento delle diver	se binarizzazioni. 2a	Training set 2b	Test set 2c Validation set
-----------------------------	-----------------------	-----------------	----------------------------

FETTA															
	SO	GLIA ROUN	ND	S	oglia ots	U	SO	GLIA ppv=	0,6	SOG	LIA yuden=	=0,02	SOC	GLIA spec=	0,42
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,364	0,825	0,492	0,424	0,696	0,521	0,328	0,834	0,456	0,595	0,684	0,624	0,363	0,823	0,489
DEV, STD	0,141	0,149	0,156	0,118	0,155	0,128	0,137	0,177	0,156	0,164	0,116	0,135	0,142	0,163	0,154
0 esclusi							0,334	0,849	0,465	0,599	0,690	0,629	0,369	0,835	0,497
							0,131	0,136	0,145	0,157	0,100	0,124	0,135	0,129	0,143
PAZIENTE															
	SO	GLIA ROUN	ND	S	oglia ots	U	SO	GLIA ppv=	0,6	SOG	LIA yuden=	=0,02	SOC	GLIA spec=	0,42
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,344	0,785	0,465	0,414	0,657	0,499	0,312	0,797	0,433	0,588	0,643	0,601	0,346	0,780	0,464
DEV, STD	0,142	0,181	0,166	0,113	0,180	0,139	0,135	0,225	0,162	0,189	0,148	0,164	0,144	0,221	0,165
0 esclusi							0,330	0,845	0,460	0,605	0,662	0,618	0,367	0,827	0,493
							0,115	0,115	0,125	0,162	0,100	0,129	0,120	0,113	0,122



FETTA															
	SO	GLIA ROU	ND	S	oglia ots	U	SO	GLIA ppv=	0,6	SOG	LIA yuden=	=0,02	SOC	GLIA spec=	0,42
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,305	0,613	0,363	0,380	0,604	0,437	0,112	0,153	0,112	0,216	0,193	0,167	0,123493	0,152	0,119
DEV, STD	0,275	0,355	0,290	0,233	0,257	0,230	0,149	0,165	0,122	0,206	0,181	0,142	0,158	0,159	0,127
0 esclusi				0,385	0,612	0,442	0,141	0,193	0,141	0,246	0,220	0,190	0,172	0,212	0,166
				0,231	0,249	0,226	0,155	0,164	0,121	0,203	0,178	0,137	0,162	0,150	0,121
PAZIENTE															
	E SOGLIA ROUND			S	oglia ots	U	SO	GLIA ppv=	0,6	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,42
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,345	0,685	0,446	0,405	0,665	0,492	0,096	0,145	0,108	0,227	0,214	0,174	0,111	0,147	0,118
DEV, STD	0,203	0,299	0,233	0,154	0,250	0,188	0,075	0,104	0,078	0,146	0,207	0,096	0,081	0,102	0,081
0 esclusi							0,103	0,155	0,116	0,243	0,229	0,186	0,119	0,157	0,127
							0,073	0,099	0,075	0,137	0,206	0,086	0,078	0,097	0,077

Tab 2b

FELIA															
	SC	GLIA ROU	ND	S	OGLIA OTS	U	SO	GLIA ppv=	0,6	SOG	LIA yuden:	=0,02	SOG	GLIA spec=	0,42
		validation			validation			validation			validation			validation	1
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,345	0,725	0,454	0,402	0,632	0,481	0,045	0,167	0,064	0,207	0,259	0,216	0,102	0,248	0,135
DEV, STD	0,120	0,145	0,123	0,105	0,158	0,107	0,054	0,169	0,073	0,094	0,129	0,087	0,066	0,143	0,076
0 esclusi				0,402	0,632	0,481	0,052	0,194	0,074	0,209	0,261	0,218	0,105	0,254	0,138
				0,105	0,158	0,107	0,055	0,168	0,074	0,092	0,128	0,086	0,065	0,140	0,074
PAZIENTE															
	SOGLIA ROUND			S	oglia ots	U	SO	GLIA ppv=	0,6	SOG	LIA yuden:	=0,02	SOC	GLIA spec=	0,42
		validation			validation			validation			validation			validation	1
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,299	0,677	0,406	0,362	0,599	0,439	0,030	0,141	0,044	0,185	0,223	0,195	0,097	0,234	0,130
DEV, STD	0,094	0,210	0,117	0,059	0,182	0,094	0,033	0,153	0,046	0,083	0,099	0,081	0,066	0,120	0,077
0 esclusi				0,362	0,599	0,439	0,035	0,163	0,051	0,193	0,233	0,204	0,102	0,245	0,136
				0,059	0,182	0,094	0,033	0,153	0,046	0,074	0,088	0,071	0,064	0,111	0,073



Z-score per fetta





FETTA															
	SO	GLIA ROU	ND	S	OGLIA OTS	U	SO	GLIA ppv=0),72	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,72
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,485	0,808	0,596	0,502	0,715	0,584	0,440	0,839	0,566	0,570	0,760	0,642	0,447	0,835	0,571
DEV, STD	0,152	0,152 0,134 0,146			0,149	0,136	0,148	0,143	0,151	0,156	0,113	0,137	0,149	0,141	0,151
0 esclusi							0,446	0,850	0,574	0,575	0,768	0,648	0,453	0,846	0,579
							0,141	0,107	0,138	0,147	0,086	0,123	0,141	0,105	0,137
	SO	GLIA ROU	ND	S	OGLIA OTS	U	SO	GLIA ppv=0),72	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,72
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,458	0,769	0,561	0,482	0,683	0,556	0,415	0,801	0,533	0,541	0,719	0,608	0,422	0,798	0,538
DEV, STD	0,168	0,170	0,175	0,145	0,177	0,160	0,163	0,185	0,178	0,180	0,161	0,171	0,165	0,185	0,178
0 esclusi							0,428	0,825	0,549	0,557	0,740	0,625	0,435	0,821	0,554
							0,149	0,126	0,155	0,156	0,106	0,137	0,150	0,126	0,155

Tab.3 Andamento delle diverse binarizzazioni. 3a Training set 3b Test set 3c Validation set

Tab.3a

SO	GLIA ROU	ND	S	OGLIA OTS	U	SOC	GLIA ppv=0),72	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,72
	test			test			test			test			test	
recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
0,356	0,680	0,408	0,382	0,631	0,448	0,118	0,173	0,118	0,163	0,206	0,146	0,120	0,172	0,119
0,287	0,285	0,285	0,242	0,224	0,238	0,148	0,163	0,121	0,172	0,181	0,126	0,149	0,160	0,121
						0,146	0,213	0,146	0,189	0,239	0,169	0,147	0,210	0,146
						0,151	0,155	0,118	0,171	0,174	0,120	0,152	0,152	0,118
SO	GLIA ROU	ND	S	OGLIA OTS	U	SOC	GLIA ppv=0),72	SOG	LIA yuden=	=0,02	SOG	GLIA spec=	0,72
	test			test			test			test			test	
recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
0,395	0,708	0,475	0,427	0,703	0,514	0,117	0,169	0,127	0,173	0,253	0,158	0,120	0,168	0,129
0,233	0,245	0,235	0,183	0,145	0,181	0,080	0,109	0,079	0,110	0,219	0,078	0,081	0,110	0,080
						0,126	0,181	0,136				0,128	0,180	0,138
i						0,077	0,103	0,074				0,077	0,104	0,074
	SC recall 0,356 0,287 SC recall 0,395 0,233	SOGLIA ROU test recall precision 0,356 0,680 0,287 0,285 SOGLIA ROU test recall precision 0,395 0,708 0,233 0,245	SOGLIA ROUNT recall precision DSC 0,356 0,680 0,408 0,287 0,285 0,285 SOGLIA ROUNT test recall precision 0,395 0,708 0,475 0,233 0,245 0,235	SUBLIA ROUND S test second 0,355 0,680 0,408 0,382 0,287 0,285 0,285 0,242 SUBLIA ROUND second second second SUBLIA ROUND SEcond SEcond second	SOGLIA ROUNT SOGLIA ROUNT recall precision DSC recall precision 0,355 0,680 0,408 0,382 0,631 0,287 0,285 0,242 0,224 SOGLIA ROUNT SOGLIA ROUNT recall precision DSC recall precision 0,395 0,708 0,427 0,703 0,333 0,245 0,235 0,183 0,145	SOGLIA ROUND SOGLIA OSU recall precision DSC recall precision DSC 0,356 0,680 0,408 0,382 0,631 0,448 0,287 0,285 0,285 0,242 0,224 0,238 SOGLIA ROUND SOGLIA ROUND Test Test Test <	SOGLA ROUND SOGLA ROUND	No No	Nome Nom Nome Nome <th< td=""><td>$\begin{array}{ c c c c c c } \hline \begin{tabular}{ c c c c c } \hline \begin{tabular}{ c c c c c c c } \hline \begin{tabular}{ c c c c c c c c c c c c c c c c c c c$</td><td>$\begin{array}{ c c c c c c c } \hline \begin{tabular}{ c c c c c c c } \hline \begin{tabular}{ c c c c c c c c c c c c c c c c c c c$</td><td>Image: Normal System Image: N</td><td>$\begin{array}{ c c c c c c c c c c c c c c c c c c c$</td><td>Image: constraint of the state of</td></th<>	$ \begin{array}{ c c c c c c } \hline \begin{tabular}{ c c c c c } \hline \begin{tabular}{ c c c c c c c } \hline \begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c } \hline \begin{tabular}{ c c c c c c c } \hline \begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	Image: Normal System Image: N	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	Image: constraint of the state of

Tab.3b

FETTA															
	SC	GLIA ROU	ND	S	oglia ots	U	SO	GLIA ppv=0),72	SOG	LIA yuden=	=0,02	SOC	GLIA spec=	0,72
		validation	l.		validation			validation			validation	I		validation	l.
	recall	precision	DSC												
MEDIA	0,435	0,685	0,513	0,468	0,617	0,519	0,064	0,160	0,081	0,178	0,250	0,195	0,065	0,160	0,082
DEV, STD	0,149	0,132	0,132	0,134	0,143	0,117	0,076	0,154	0,090	0,090	0,136	0,089	0,077	0,153	0,091
0 esclusi							0,074	0,185	0,093	0,179	0,252	0,196	0,074	0,182	0,093
							0,077	0,151	0,090	0,089	0,135	0,088	0,078	0,151	0,091
	SC	GLIA ROU	ND	S	OGLIA OTS	U	SO	GLIA ppv=0),72	SOG	LIA yuden=	=0,02	SOC	GLIA spec=	0,72
		validation	l.		validation			validation			validation	I		validation	l.
	recall	precision	DSC												
MEDIA	0,394	0,657	0,478	0,431	0,562	0,474	0,045	0,128	0,055	0,153	0,220	0,173	0,045	0,127	0,055
DEV, STD	0,123	0,167	0,123	0,113	0,174	0,121	0,054	0,126	0,061	0,076	0,120	0,085	0,054	0,127	0,061
0 esclusi							0,055	0,156	0,067	0,161	0,231	0,182	0,055	0,156	0,067
							0,055	0,123	0,062	0,070	0,112	0,078	0,055	0,124	0,062

Tab.3c





Tab.4 Andamento delle diverse binarizzazioni. 4a Training set 4b Test set 4c Validation set

FETTA															
	SO	GLIA ROU	ND	S	DGLIA OTS	U	SOC	GLIA ppv=0),74	SOG	LIA yuden=	=0,06	SOC	GLIA spec=	D,66
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,460	0,832	0,581	0,509	0,678	0,574	0,431	0,847	0,559	0,536	0,787	0,626	0,454	0,836	0,576
DEV, STD	0,151	0,130	0,146	0,124	0,161	0,131	0,147	0,132	0,148	0,158	0,123	0,141	0,150	0,131	0,147
0 esclusi							0,436	0,857	0,566	0,542	0,796	0,633	0,459	0,845	0,583
							0,141	0,097	0,136	0,148	0,091	0,126	0,143	0,097	0,134
PAZIENTE															
	SO	GLIA ROU	ND	S	oglia ots	U	SOC	GLIA ppv=0),74	SOG	LIA yuden:	=0,06	SOC	GLIA spec=	0,66
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,429	0,795	0,546	0,490	0,641	0,549	0,399	0,809	0,524	0,506	0,747	0,592	0,423	0,800	0,542
DEV, STD	0,155	0,177	0,170	0,110	0,183	0,137	0,149	0,181	0,168	0,173	0,168	0,173	0,154	0,178	0,169
0 esclusi							0,411	0,833	0,539	0,521	0,769	0,610	0,435	0,824	0,558
							0,135	0,118	0,145	0,152	0,111	0,143	0,138	0,115	0,143

Tab 4a

FETTA															
	SC	GLIA ROU	ND	S	OGLIA OTS	U	SO	GLIA ppv=0),74	SOG	LIA yuden=	=0,06	SOC	GLIA spec=	0,66
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,348	0,727	0,408	0,412	0,603	0,462	0,126	0,188	0,123	0,159	0,189	0,144	0,132	0,187	0,127
DEV, STD	0,282	0,268	0,288	0,240	0,250	0,236	0,158	0,182	0,126	0,173	0,164	0,128	0,161	0,178	0,126
0 esclusi							0,153	0,228	0,150	0,188	0,223	0,169	0,159	0,225	0,153
							0,162	0,176	0,123	0,173	0,155	0,122	0,164	0,172	0,123
PAZIENTE															
	SC	GLIA ROU	ND	S	OGLIA OTS	U	SO	GLIA ppv=0),74	SOG	LIA yuden=	=0,06	SOC	GLIA spec=	0,66
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,396	0,696	0,487	0,475	0,609	0,512	0,139	0,174	0,143	0,175	0,165	0,158	0,145	0,172	0,146
DEV, STD	0,229	0,297	0,244	0,177	0,230	0,190	0,111	0,111	0,093	0,135	0,103	0,097	0,113	0,109	0,093
0 esclusi							0,149	0,186	0,153	0,188	0,176	0,169	0,156	0,184	0,156
							0,108	0,104	0,088	0,131	0,096	0,091	0,109	0,102	0,088

Tab 4b

FETTA															
	SO	GLIA ROU	ND	S	OGLIA OTS	U	SOC	GLIA ppv=0),74	SOG	LIA yuden=	=0,06	SOG	GLIA spec=	D,66
	validation				validation			validation			validation			validation	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,448	0,701	0,528	0,488	0,612	0,530	0,067	0,172	0,085	0,166	0,253	0,188	0,071	0,171	0,088
DEV, STD	0,152	0,127	0,129	0,131	0,145	0,117	0,078	0,154	0,091	0,083	0,134	0,086	0,081	0,151	0,092
0 esclusi							0,078	0,197	0,097	0,168	0,257	0,191	0,080	0,193	0,099
							0,078	0,149	0,091	0,081	0,131	0,083	0,081	0,147	0,092
PAZIENTE															
	SO	GLIA ROU	ND	S	OGLIA OTS	U	SOC	GLIA ppv=0),74	SOG	LIA yuden=	=0,06	SOG	GLIA spec=	0,66
	validation				validation			validation			validation			validation	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,411	0,657	0,496	0,454	0,558	0,490	0,049	0,139	0,061	0,148	0,217	0,169	0,052	0,138	0,064
DEV, STD	0,124	0,191	0,138	0,091	0,175	0,117	0,050	0,127	0,057	0,075	0,116	0,084	0,052	0,123	0,059
0 esclusi							0,060	0,170	0,075	0,155	0,227	0,177	0,063	0,169	0,078
							0,049	0,120	0,055	0,070	0,109	0,078	0,051	0,115	0,056

Tab	4c
-----	----

Dati tutti i risultati raccolti, si nota come la normalizzazione sul paziente risulti preferibile rispetto a quella per fetta, anche da un punto di vista concettuale: è più lineare che i parametri di minimo, massimo, media e deviazione standard siano raccolti sul tumore intero rispetto che sulla singola fetta; come tipo di normalizzazione la z-score ha prodotto precision, recall e dice più alti.

E' stato dunque deciso di portare avanti il modello con dati normalizzati con z-score standardization per paziente.

5.4 Confronto tra loss functions

Si è scelto poi di effettuare un'ulteriore prova sul modello, variando la loss function per l'allenamento della rete.

In generale una funzione di costo o appunto loss function, si utilizza per l'aggiornamento dei pesi, che vengono modificati in modo che questa sia minimizzata, diminuendo il più possibile il divario tra la probabilità di classe ipotizzata e la classe effettiva.

Tutte le prove preliminari erano state effettuate con la Binary Cross Entropy, che rappresenta la funzione di costo di default per problemi di classificazione. Ad ogni epoca

di training, si calcola il punteggio/perdita dell'output della rete, di natura logaritmica, che sarà più alto quanto sarà grande la distanza dal valore atteso effettivo.

La Dice Loss, implementata in questa tesi, poteva risultare più adatta rispetto alla Binary Cross Entropy a causa della tipologia di dataset: infatti a partire dalla definizione di Dice Coefficient:

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

il numeratore rappresenta le attivazioni comuni tra A e B, ovvero maschera predetta e originale, mentre il denominatore coinvolge il numero di attivazioni delle due maschere separatamente; ciò consente di normalizzare i risultati della loss sulla maschera target, in modo che questi non siano influenzati dal fatto che una classe sia più rappresentata dell'altra, come nel caso delle segmentazioni tumorali, nelle quali alla classe "tumore" appartengono pochi elementi rispetto alla classe "non tumore"[29]



Fig.15 15a:Andamento di specificità, sensibilità. 15b: Andamento del PPV. 15c: Andamento dello Youden Index

FETTA															
	SO	GLIA ROU	ND	S	oglia ots	U	SOC	GLIA ppv=0),86	SOG	LIA yuden:	=0,06	SOC	GLIA spec=	0,66
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,364	0,825	0,492	0,779	0,536	0,630	0,754	0,560	0,637	0,789	0,527	0,627	0,744	0,567	0,638
DEV, STD	0,141	0,149	0,156	0,123	0,098	0,098	0,140	0,100	0,107	0,129	0,095	0,100	0,141	0,101	0,108
0 esclusi															
PAZIENTE															
	SO	GLIA ROU	ND	S	OGLIA OTS	U	SOC	GLIA ppv=0),86	SOG	LIA yuden:	=0,06	SOC	GLIA spec=	0,66
		train			train			train			train			train	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC
MEDIA	0,344	0,785	0,465	0,784	0,511	0,611	0,759	0,538	0,623	0,795	0,506	0,612	0,745	0,543	0,621
DEV, STD	0,142	0,181	0,166	0,129	0,101	0,100	0,149	0,100	0,110	0,134	0,092	0,097	0,152	0,102	0,112
0 esclusi															

Tab 5a

FETTA															
	SO	GLIA ROU	ND	S	oglia ots	U	SO	GLIA ppv=0),86	SOG	LIA yuden=	=0,06	SOC	GLIA spec=	0,66
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,305	0,613	0,363	0,680	0,556	0,577	0,292	0,205	0,217	0,324	0,203	0,227	0,284	0,204	0,213
DEV, STD	0,275	0,355	0,290	0,233	0,165	0,165	0,186	0,131	0,120	0,193	0,126	0,122	0,185	0,131	0,120
0 esclusi										0,336	0,211	0,236	0,295	0,212	0,222
										0,185	0,122	0,115	0,179	0,127	0,114
PAZIENTE															
	SO	GLIA ROU	ND	S	OGLIA OTS	U	SOC	GLIA ppv=0),86	SOG	LIA yuden=	=0,06	SOC	GLIA spec=	0,66
		test			test			test			test			test	
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,345	0,685	0,446	0,726	0,533	0,602	0,315	0,191	0,212	0,346	0,184	0,217	0,301	0,192	0,207
DEV, STD	0,203	0,299	0,233	0,192	0,109	0,133	0,147	0,099	0,077	0,156	0,089	0,077	0,142	0,106	0,078
0 esclusi															

FETTA															
	SC	GLIA ROU	ND	S	oglia ots	U	SO	GLIA ppv=(),86	SOG	LIA yuden=	=0,06	SOG	GLIA spec=	0,66
	validation				validation			validation			validation			validation	L. C. C.
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,345	0,725	0,454	0,775	0,490	0,593	0,160	0,174	0,150	0,337	0,251	0,279	0,156	0,174	0,148
DEV, STD	0,120	0,145	0,123	0,108	0,103	0,098	0,139	0,116	0,115	0,105	0,105	0,093	0,137	0,116	0,114
0 esclusi							0,176	0,191	0,165				0,171	0,191	0,162
							0,136	0,108	0,110				0,134	0,108	0,109
PAZIENTE															
	SC	GLIA ROU	ND	S	oglia ots	U	SO	GLIA ppv=(),86	SOG	LIA yuden=	=0,06	SOG	GLIA spec=	0,66
	validation				validation			validation			validation			validation	1
	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	DSC	recall	precision	dsc
MEDIA	0,299	0,677	0,406	0,743	0,464	0,564	0,121	0,146	0,115	0,307	0,225	0,252	0,119	0,146	0,114
DEV, STD	0,094	0,210	0,117	0,103	0,128	0,121	0,099	0,111	0,081	0,092	0,089	0,084	0,097	0,111	0,080
0 esclusi							0,141	0,169	0,133				0,137	0,169	0,132
							0,093	0,102	0,073				0,092	0,101	0,071

Tab 5b

Tab 5c

Come si può facilmente notare confrontando i dati riportati nelle tabelle 4 e 5, nel caso della Dice Loss Function la Recall risulta essere in tutti i set più alta della Precision, mentre per la rete allenata con Binary Cross Entropy Loss Function questo andamento è invertito, ovvero la Precision assume valori maggiori rispetto alla Recall.

Ad una prima approssimazione quindi, sembra che la rete allenata con Dice Loss sia più adatta agli scopi di questo studio perché riporta meno falsi negativi e risulta essere più conservativa verso la lesione rispetto alla Binary Cross Entropy Loss.

6 Valutazione delle prestazioni

A partire dai risultati ottenuti nel testing dei modelli, si è notato come, tra le binarizzazioni, quella effettuata tramite la soglia di Otsu ottiene i risultati migliori, ed anche più equilibrati tra construction e validation set; si è deciso di portare avanti anche la binarizzazione caratterizzata dalla soglia per ottimizzare lo Youden Index che, tra i parametri derivati dalla curva ROC, è risultato il più adatto allo studio.

Inoltre, le prestazioni della rete sono state valutate sia per la rete allenata con Binary Cross Entropy loss che per quella con la Dice loss function.

Sono stati valutati, dunque, quattro modelli in totale:

- Binary Cross Entropy Loss
 - o Soglia Otsu
 - Soglia Youden Index
- Dice Loss
 - Soglia Otsu
 - Soglia Youden Index

Le quattro differenti reti sono state applicate in questo caso alle immagini intere di training, test e validation set, attraverso un codice nel quale una finestra 16x16 scorre per tutta l'immagine e sottopone la ROI selezionata alla rete. Il risultato è la maschera predetta non binarizzata. Questa viene successivamente binarizzata utilizzando le soglie

di Otsu e Youden Index, creando le maschere corrispondenti, che possono essere confrontate con la maschera originale. Le soglie di Otsu sono state calcolate progressivamente per ogni fetta considerata, mentre le soglie dello Youden Index sono le originali stabilite secondo l'andamento della curva ROC sul training.

Si è poi fatto un ulteriore passaggio calcolando l'oggetto connesso tra le maschere predette e binarizzate e la maschera originale: il codice creato estrae i pixel comuni tra le maschere, riuscendo quindi ad individuare la lesione.

Sono stati calcolati i risultati di Dice, Precision e Recall medii per ogni paziente e poi specifici per ogni fetta del paziente degli oggetti connessi calcolati per i quattro modelli considerati.



Grafici andamenti sui pazienti

Fig.16 Andamenti del Dice Coefficient sul Train per i 4 modelli allenati



Fig.17 Andamenti della Precision sul Train per i 4 modelli allenati



Fig.18 Andamenti della Recall sul Train per i 4 modelli allenati

	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
dice	0,458	0,438	0,502	0,539
precision	0,372	0,344	0,754	0,696
recall	0,722	0,739	0,423	0,497

Tab.6 Valori medii sui pazienti per il train

I grafici mostrano un andamento abbastanza altalenante sui pazienti: bisogna sottolineare che alcuni soggetti, soprattutto appartenenti al dataset delle Molinette, come il paziente 9, avevano tumori davvero molto piccoli, le cui maschere sono composte da un bassissimo numero di pixel, dunque le reti create hanno avuto molta difficoltà a riconoscerli. Un altro esempio particolare è il paziente 69, ovvero il paziente 37 di Candiolo: il tumore ha dimensioni molto contenute ed inoltre il radiologo ha individuato in alcune slices delle parti di tumore infinitesime, che hanno contribuito a penalizzare il risultato di classificazione medio del paziente.



Fig.19 Andamenti del Dice Coefficient sul Test per i 4 modelli allenati



Fig.20 Andamenti della Precision sul Test per i 4 modelli allenati



Fig.21 Andamenti della Recall sul Test per i 4 modelli allenati

	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
dice	0,479	0,465	0,484	0,510
precision	0,413	0,392	0,766	0,715
recall	0,733	0,745	0,427	0,492

Tab.7 Valori medii sui pazienti per il test



Fig.22 Andamenti del Dice Coefficient sul Validation per i 4 modelli allenati



Fig.23 Andamenti della Precision sul Validation per i 4 modelli allenati



Fig.24 Andamenti della sul Validation per i 4 modelli allenati

	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
dice	0,399	0,184	0,449	0,469
precision	0,301	0,124	0,592	0,544
recall	0,743	0,800	0,423	0,495

Tab.8 Valori medii sui pazienti per il test

I grafici mostrano l'andamento di Dice, Precision e Recall sui diversi pazienti. I risultati non sono ottimali, tuttavia ci sono pazienti che raggiungono Dice oltre gli 0.8 per il Train ed il Test set, e pazienti con Dice oltre 0.7 per il Validation.

E' evidente che le differenze maggiori si abbiano sui modelli allenati con diverse loss function: come osservato precedentemente, la Recall dei modelli allenati con Dice Loss è molto più alta rispetto ai modelli allenati con Binary Cross Entropy Loss, e ciò comporta una lesione "sovrastimata", dunque con più falsi positivi ma comunque ben individuata, anche con picchi di Recall pari a 1 ovvero senza falsi negativi; al contrario, i modelli allenati con BCE Loss sono più precisi, dunque selettivi nella segmentazione.

Non c'è un modello che prevale, per ogni paziente un'alternativa risulta migliore rispetto all'altra: ciò può dipendere da diversi fattori quali la grandezza della lesione, il tipo, ovvero mass o non mass, l'orientamento o le modalità di acquisizione. Gli ultimi due criteri in realtà non sembrano essere determinanti, perché pazienti di Candiolo o delle Molinette hanno risultati di classificazione diversi indipendentemente dalla provenienza dei dati, e quindi anche dalle modalità di acquisizione differenti, ed anche l'orientamento sagittale o assiale sembra non influire sulle differenza tra le prestazioni delle due loss functions, anche se si può osservare che per le proiezioni saggittali, sebbene in minoranza nel dataset, le lesioni sono classificate in maniera più efficace, come nel caso dei pazienti 40, 41,46.

Un esempio di classificazione ben riuscita è il paziente 9 di Candiolo, parte del train:



46

	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
DICE	0,840	0,826	0,842	0,861
PRECISION	0,795	0,763	0,961	0,930
RECALL	0,890	0,901	0,749	0,801

Tab.6 Dati relativi a paziente 9 di Candiolo, slice 26

Il paziente è risultato un ottimo soggetto di allenamento, con un tumore abbastanza esteso e di tipo massivo. La forma particolare della maschera della slice 26 è stata predetta quasi perfettamente dalle reti allenate, raggiungendo Dice medii davvero alti.

Alcune maschere predette, in particolare quelle del modello allenato con Dice Loss, presentano un pattern nelle ROI di sfondo, che viene ovviamente eliminato nel calcolo dell'oggetto connesso; questo fenomeno è attribuibile al fatto che le reti sono state allenate con ROI aventi almeno 5 pixel tumorali, dunque nessuna ROI del construction set era tutta nera, come tipicamente è lo sfondo d'aria delle MRI. Inoltre in tutte le maschere sono identificati come tumore gli organi interni, ad esempio il cuore: questo è facilmente ovviabile con un pre o post processing delle immagini, anche se in ogni caso sicuramente questo tipo di falsi positivi è escluso a priori dai radiologi.



	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
DICE	0,530	0,537	0,085	0,129
RECALL	0,398	0,779	0,044	0,070
PRECISION	0,795	0,410	0,951	0,939

Tab 7. Dati relativi al paziente 16 di Molinette.

Un esempio di paziente appartenente al test set: le maschere predette assumono le stesse caratteristiche già descritte nel paziente del train che è stato riportato sopra. Pattern presenti nelle maschere dei modelli allenati con Dice Loss, ma oggetto connesso in questo caso più fedele alla maschera originale. Molti meno falsi positivi nelle maschere relative al modello allenato con BCE Loss ma lesione individuata in maniera del tutto marginale, come si nota dai risultati in Tab.7.



	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
DICE	0,285	0,255	0,707	0,561
RECALL	0,991	0,991	0,826	0,890
PRECISION	0,166	0,146	0,618	0,410

Tab.8 Dati relativi al paziente 6 di Candiolo

In questo paziente, anch'esso appartenente al test set, si può notare come le reti allenate con BCE abbiano segmentato in maniera più accurata rispetto che alle reti allenate con DICE: ciò è chiaro a partire dal Dice coefficient, misura di similarità, che risulta essere nettamente più alto nel primo caso; bisogna evidenziare che la recall nel caso dei modelli allenati con Dice Loss risulti essere quasi 1: ciò significa che quasi ogni pixel del tumore è stato correttamente classificato, ma che la rete ha delineato una lesione di dimensioni maggiori rispetto a quella reale, cioè ha sovra segmentato.



	OTSU DICE	YOUDEN DICE	OTSU BIN	YOUDEN BIN
dice	0,64	0,66	0,27	0,34
recall	0,61	0,69	0,16	0,21
precision	0,67	0,63	0,93	0,91

Tab.9 Dati relativi al paziente 22 di Molinette

Il paziente 22 del dataset Molinette è appartenente al validation test, dunque del tutto esterno dall'allenamento della rete: in questo caso si nota che il modello con Dice loss (28c-d) riesce meglio a delineare la maschera, mentre quello con BinaryCrossEntropy loss sottostima notevolmente la lesione, nonostante la precision sia prossima ad 1.

Conclusione e prospettive future

In questa tesi, ci si è concentrati sul creare un modello di rete U-net che potesse segmentare autonomamente il tumore alla mammella a partire da immagini di risonanza magnetica. Sono stati costruiti differenti modelli che si differenziavano per la normalizzazione e la successiva manipolazione dei dati e per la loss function della rete.

In generale le performance di classificazione non sono ottime ma questo studio può costituire un ottimo punto di partenza per approfondimenti successivi, in cui la struttura della rete può essere ampliata e resa più articolata, o possono essere manipolati i dati applicando un data augmentation in modo tale da eliminare il problema di overfitting ed aumentare le prestazioni di classificazione. Non sono stati trovati in letteratura approcci di deep learning che partissero dall'immagine intera senza aver prima segmentato la zona del seno: sicuramente una pre-segmentazione che eliminasse tutti gli organi interni del torace avrebbe migliorato i risultati, ma questo ha costituito un elemento di novità di questo lavoro di tesi. Ciò che si sarebbe potuto fare è effettuare un post-processing sulle immagini che aiutasse a limitare la quantità di falsi positivi dovuti appunto alla presenza di organi, per "pulire" l'immagine ed ottenere valori di Dice Precision e Recall più alti.

Bisogna sottolineare che molti pazienti utilizzati in questo studio erano caratterizzati da lesioni non-mass, che sono estremamente più complesse da individuare: in letteratura i tradizionali metodi di machine learning non riescono a creare features che consentano di caratterizzare questo tipo di tumori, non avendo una morfologia caratteristica o dei bordi delineati. Il metodo di deep learning utilizzato in questa tesi invece supera questo ostacolo, andando ad estrarre autonomamente pattern caratteristici: tuttavia, date le caratteristiche morfologiche delle lesioni non-mass, i componenti tumorali non sono concentrati ma si estendono tra il tessuto adiposo e fibro-ghiandolare, rendendo estremamente complessa la segmentazione corretta. Sicuramente però una segmentazione automatica è ottimale rispetto alla manuale, perché spesso questa tipologia di tumore ha delle caratteristiche che possono essere nascoste all'occhio umano e alla logica delle features costruite manualmente.

E' stato scelto di utilizzare per la costruzione della rete ROI 16x16 che avessero almeno 5 pixel tumorali: di conseguenza la rete ha imparato a vedere tumore anche se magari non presente, ecco perché nelle maschere generate si hanno spesso dei pattern di pixel classificati come tumorali in ROI che invece dovrebbero essere tutte categorizzate come "non tumore".

Si è notato come la dice loss risulti più adatta ad avere una valutazione conservativa, perché nella maggior parte dei casi la recall è risultata particolarmente elevata, mentre la binary cross entropy loss è risultata molto più precisa ma spesso con problemi di sotto stima della lesione.

Dagli esempi riportati si è dunque compreso che una buona segmentazione è garantita quando c'è un equilibrio tra i valori dei parametri calcolati: Recall eccessivamente elevate indicano una sovrastima nella segmentazione mentre Precision molto alte garantiscono che i punti individuati appartengano al tumore, ma spesso la lesione viene eccessivamente sotto segmentata. Al fine quindi di valutare le prestazioni della rete vanno sempre tenuti in considerazione entrambi i criteri ed il Dice è un'ottima sintesi tra i due precedenti.

Per concludere, un dataset più ricco ed una rete di maggiore complessità aiuterebbero sicuramente i risultati di classificazione, ma lo studio effettuato in questa tesi è solo la fase iniziale di questo progetto che sembra comunque avere grandi prospettive di sviluppo.

Bibliografia

- [1] 'Breast cancer'. https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed Jun. 15, 2021).
- [2] European Cancer Information System, 'European Commission. Breast cancer burden in EU-27.' Jan. 2021. [Online]. Available: https://ecis.jrc.ec.europa.eu/pdf/Breast_cancer_factsheet-Dec_2020.pdf
- [3] Z. Tao, A. Shi, C. Lu, T. Song, Z. Zhang, and J. Zhao, 'Breast cancer: epidemiology and etiology', *Cell Biochem. Biophys.*, vol. 72, no. 2, pp. 333–338, 2015.
- [4] H. D. Nelson *et al.*, 'Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis', *Ann. Intern. Med.*, vol. 156, no. 9, pp. 635–648, May 2012, doi: 10.7326/0003-4819-156-9-201205010-00006.
- [5] CDCBreastCancer, 'What Are the Risk Factors for Breast Cancer?', Centers for Disease Control and Prevention, Sep. 14, 2020. https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm (accessed Jun. 15, 2021).
- [6] M. Oluwasanu and O. I. Olopade, 'Global disparities in breast cancer outcomes: new perspectives, widening inequities, unanswered questions', *Lancet Glob. Health*, vol. 8, no. 8, pp. e978–e979, Aug. 2020, doi: 10.1016/S2214-109X(20)30307-7.
- [7] 'Breast Cancer', *The Cancer Atlas*. http://canceratlas.cancer.org/QxP (accessed Jun. 16, 2021).
- [8] M. Akram, M. Iqbal, M. Daniyal, and A. U. Khan, 'Awareness and current knowledge of breast cancer', *Biol. Res.*, vol. 50, no. 1, p. 33, Oct. 2017, doi: 10.1186/s40659-017-0140-9.
- [9] 'Breast Cancer Stages', *Cancer.Net*, Jun. 25, 2012. https://www.cancer.net/cancer-types/breast-cancer/stages (accessed Jun. 18, 2021).
- [10] 'Cancer Staging National Cancer Institute', Sep. 03, 2015. https://www.cancer.gov/about-cancer/diagnosis-staging/staging (accessed Jun. 18, 2021).
- [11] F. Cardoso *et al.*, 'Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Ann. Oncol.*, vol. 30, no. 8, pp. 1194–1220, 2019.
- [12] M. Morrow, J. Waters, and E. Morris, 'MRI for breast cancer screening, diagnosis, and treatment', *The Lancet*, vol. 378, no. 9805, pp. 1804–1811, 2011.
- [13] J. V. Rispoli, 'Current and Emerging Magnetic Resonance-Based Techniques for Breast Cancer', *Front. Med.*, vol. 7, p. 17, 2020.
- [14] 'Le tecniche di indagine MRS'. http://www.neuroscienzedipendenze.it/mrs.html (accessed Jun. 20, 2021).
- [15] N. I. R. Yassin, S. Omran, E. M. F. El Houby, and H. Allam, 'Machine learning techniques for breast cancer computer aided diagnosis using different image

modalities: A systematic review', *Comput. Methods Programs Biomed.*, vol. 156, pp. 25–45, Mar. 2018, doi: 10.1016/j.cmpb.2017.12.012.

- [16] A. Meyer-Bäse, L. Morra, U. Meyer-Bäse, and K. Pinker, 'Current Status and Future Perspectives of Artificial Intelligence in Magnetic Resonance Breast Imaging', *Contrast Media Mol. Imaging*, vol. 2020, pp. 1–18, Aug. 2020, doi: 10.1155/2020/6805710.
- [17] 'Convolutional neural networks: an overview and application in radiology | Insights into Imaging | Full Text'. https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9 (accessed Jun. 28, 2021).
- [18] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [19] K. Ravichandran, N. Braman, A. Janowczyk, and A. Madabhushi, 'A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI', p. 12.
- [20] G. Piantadosi, M. Sansone, R. Fusco, and C. Sansone, 'Multi-planar 3D breast segmentation in MRI via deep convolutional neural networks', *Artif. Intell. Med.*, vol. 103, p. 101781, Mar. 2020, doi: 10.1016/j.artmed.2019.101781.
- [21] H. Jiao, X. Jiang, Z. Pang, X. Lin, Y. Huang, and L. Li, 'Deep Convolutional Neural Networks-Based Automatic Breast Segmentation and Mass Detection in DCE-MRI', *Comput. Math. Methods Med.*, p. 12.
- [22] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', *ArXiv150504597 Cs*, May 2015, Accessed: Jun. 28, 2021. [Online]. Available: http://arxiv.org/abs/1505.04597
- [23] M. Niemeijer, M. Loog, M. D. Abramoff, M. A. Viergever, M. Prokop, and B. van Ginneken, 'On combining computer-aided detection systems', *IEEE Trans. Med. Imaging*, vol. 30, no. 2, pp. 215–223, Feb. 2011, doi: 10.1109/TMI.2010.2072789.
- [24] M. U. Dalmış, S. Vreemann, T. Kooi, R. M. Mann, N. Karssemeijer, and A. Gubern-Mérida, 'Fully automated detection of breast cancer in screening MRI using convolutional neural networks', *J. Med. Imaging*, vol. 5, no. 1, p. 014502, 2018.
- [25] M. U. Dalm and A. Gubern-Me, 'Using deep learning to segment breast and fibroglandular tissue in MRI volumes', *Med. Phys.*, p. 15, 2017.
- [26] Y. Cao *et al.*, 'Automatic detection and segmentation of multiple brain metastases on magnetic resonance image using asymmetric UNet architecture', *Phys. Med. Biol.*, vol. 66, no. 1, p. 015003, Jan. 2021, doi: 10.1088/1361-6560/abca53.
- [27] J. (Hong Jing), 'Biomedical Image Segmentation: UNet++', *Medium*, Nov. 27, 2020. https://towardsdatascience.com/biomedical-image-segmentation-unet-991d075a3a4b (accessed Jun. 28, 2021).
- [28] G. D'Arrigo, F. Provenzano, C. Torino, C. Zoccali, and G. Tripepi, 'I TEST DIAGNOSTICI E L'ANALISI DELLA CURVA ROC', p. 6, 2011.
- [29] 'An overview of semantic image segmentation.' https://www.jeremyjordan.me/semantic-segmentation/ (accessed Jul. 07, 2021).

Ringraziamenti