

POLITECNICO DI TORINO
Laurea Magistrale in Ingegneria Biomedica



Sviluppo di metodi basati sul Machine Learning per la caratterizzazione di lesioni in immagini di Tomosintesi e Mammografia sintetica

Relatori

Prof. Filippo MOLINARI

Ing. Massimo SALVI

Candidata

Federica TOSI

Luglio 2021

Sommario

Il carcinoma mammario è la neoplasia con maggiore tasso di incidenza nella donna, nonché la più diagnosticata nei soggetti prevalenti. L'introduzione della tomosintesi come tecnica di imaging ha permesso di migliorare la capacità di diagnosticare il tumore in fase iniziale, ma non di diminuire il numero di casi per cui si richiede la biopsia come ulteriore accertamento: sul totale delle biopsie eseguite, la percentuale di diagnosi di carcinoma è pari al 30%. In quest'ottica è chiara l'importanza di fornire un valido strumento di supporto per la caratterizzazione automatica di masse e opacità in immagini di tomosintesi.

Questo lavoro di tesi nasce da un progetto in collaborazione con Technologie Avanzate Srl e mira allo sviluppo di un sistema ad alta sensibilità col fine di negativizzare i falsi sospetti e, di conseguenza, il numero di biopsie effettuate su pazienti sane. In seguito ad una ricerca retrospettiva, sono state raccolte le immagini di tomosintesi e mammografia 2D sintetica di 82 esami acquisite con il Selenia Dimensions Mammography System (Hologic Inc.), per un totale di 87 lesioni che sono state segmentate manualmente da senologhe esperte del reparto di Senologia delle Molinette di Torino. Sono stati sviluppati in ambiente MATLAB diversi algoritmi basati sul Machine Learning in grado di discriminare le lesioni benigne dalle maligne, che comprendono Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Artificial Neural Network (ANN) e Convolutional Neural Network (CNN). Sono state elaborate entrambe le tipologie di immagini e le singole proiezioni separatamente oppure combinandole in uno studio multimodale. Dai risultati è emerso come la configurazione con performance migliori veda l'utilizzo delle mammografie 2D sintetiche e

le proiezioni considerate separatamente, con la quale l'SVM restituisce il più alto valore di AUC pari al 93,5% sul training set e un valore medio sul validation set pari all'87%. Con la stessa configurazione è stato raggiunto un livello di sensibilità medio pari all'89% sul training set e all'82% sul validation set. Risultati promettenti sono stati ottenuti anche unendo l'informazione di entrambe le proiezioni. Infine, l'analisi approfondita delle ROI estratte dalle due tipologie di immagini fornisce un quadro delle potenzialità del loro utilizzo in uno strumento di supporto alla decisione.

Indice

Elenco delle tabelle	VII
Elenco delle figure	VIII
Acronimi	XV
1 Introduzione	1
1.1 Nozioni di anatomia	1
1.1.1 Anatomia chirurgica	5
1.2 Il carcinoma mammario	7
1.2.1 Incidenza e fattori di rischio	7
1.2.2 Masse e opacità maligne e benigne	10
1.2.3 Esame obiettivo e ispezione mammaria	13
1.2.4 Diagnostica per immagini	14
1.2.5 Diagnosi basata sull'esame istologico	21
1.3 Immagini di Tomosintesi e Machine Learning : stato dell'arte	22
2 Materiali e Metodi	25
2.1 Dataset	25
2.2 Estrazione delle ROI	27
2.2.1 Pre processing	29
2.3 Estrazione delle features	32
2.3.1 Operatori di tessitura	34
2.3.2 Metodi Transform-based	40
2.3.3 Descrittori di forma	44

2.4	Feature selection	48
2.5	Metodi di classificazione	50
2.5.1	Metodi basati sul Machine Learning	51
2.5.2	Convolutional Neural Network	59
2.6	Metodi di validazione	64
2.6.1	Metriche di validazione	66
3	Risultati	70
3.1	Effetto dell'aumento della numerosità del Dataset . . .	70
3.2	Classificazione di lesioni su immagini 2D sintetiche . .	75
3.2.1	Risultati sul Training	75
3.2.2	Risultati sul Validation	79
3.3	Classificazione di lesioni su slices di Tomosintesi	83
3.3.1	Risultati sul Training	83
3.3.2	Risultati sul Validation	86
3.4	Performance CNN	90
3.5	Integrazione e confronto delle due modalità	91
4	Conclusioni e Sviluppi futuri	98
	Bibliografia	102

Elenco delle tabelle

1.1	Categorie BI-RADS e breve descrizione in base alle linee guida oncologiche.	20
2.1	Descrizione del Dataset	26
2.2	Statistica descrittiva del prim'ordine	35
2.3	Statistica descrittiva del second'ordine	37
3.1	Divisione in sotto-categorie delle features calcolate	73
3.2	Mammografia sintetica: percentuale di casi in cui le due proiezioni vengono assegnate a classi diverse dai diversi metodi di classificazione.	77
3.3	Tomosintesi: percentuale di casi in cui le due proiezioni vengono assegnate a classi diverse dai diversi metodi di classificazione.	86
3.4	Performance della CNN su Training e Validation	90

Elenco delle figure

1.1	Anatomia della mammella e rappresentazione della struttura interna [1].	2
1.2	Rappresentazione del cambiamento di conformazione della mammella a riposo e durante l'allattamento [1].	2
1.3	Rappresentazione e indicazione della nomenclatura dei principali vasi e nodi del sistema linfatico nella regione toracica divisi nei tre livelli di appartenenza [3].	5
1.4	Mammografia destra in proiezione cranio caudale (a) e mammografia sinistra in proiezione medio laterale obliqua (b) con indicazione dei quadranti visibili.	6
1.5	Analisi a livello nazionale dell'attività di screening mammografico 2012-2018 [6].	8
1.6	Esempi di opacità di natura maligna osservate su immagini di mammografia	11
1.7	Esempi di opacità di natura benigna: (A) Steatonecrosi con clips chirurgiche visibili; (B) Cisti settata; (C) FAD calcifico; (D) FAD.	12
1.8	Schema del principio di funzionamento della tomosintesi.	16
1.9	Visualizzazione delle singole slices dell'immagine BTO su un visualizzatore DICOM : confronto tra le singole slices e cambiamento nella visualizzazione della lesione.	17
1.10	Confronto tra immagine mammografica FFDM (Full Field Digital Mammography), a sinistra, e C-View, a destra [22].	18

1.11	Visualizzazione di un tumore duttale infiltrante in DM e SM: evidenza della migliore rappresentazione delle masse spiculate nell'immagine sintetica [22].	19
1.12	Indice del numero di articoli in letteratura per macroargomento dal 2001 a oggi in base ai dati di NCBI. . .	23
2.1	Esempio di segmentazione manuale su monitor diagnostico <i>SecureView</i> in entrambe le proiezioni.	28
2.2	Sovrapposizione immagine 2D sintetica e Key-slice BTO	29
2.3	Esportazione del contorno della lesione sulla key-slice BTO	29
2.4	Step di pre-processing delle immagini di mammografia 2D sintetica.	31
2.5	Effetto dello sharpening sull'immagine della lesione sulla slice BTO	31
2.6	Schema di estrazione delle ROI dalle immagini C-View	32
2.7	(a) Visualizzazione del contorno manuale, in rosso, e del contorno della maschera dilatata in seguito al padding proporzionale, in arancione. (b) Maschera del tessuto interno identificabile come lesione. (c) Maschera della porzione di tessuto esterno	33
2.8	A sinistra l'immagine in grayscale, a destra l'istogramma delle luminosità che descrive la distribuzione dei valori di livelli di grigio.	34
2.9	Schema di calcolo della matrice GLCM nelle quattro direzioni utilizzate e a distanza generica D	36
2.10	Rappresentazione delle matrici GLCM calcolate su una ROI di esempio	36
2.11	Risultato del filtraggio con filtro di Sobel lungo X e lungo Y. L'ultima immagine è la composizione delle due precedenti, ed è quella che è stata utilizzata per il calcolo dei parametri	38
2.12	Schema del calcolo dei Local binary patterns per valori generici dei pixel [37].	38

2.13	Esempi di immagini ottenute in seguito al filtraggio con i filtri di Laws, dalle quali sono stati estratti i parametri descritti. Per ogni immagine è riportato il nome del filtro utilizzato come combinazione dei vettori originari. . . .	40
2.14	Rappresentazione del filtro di Gabor implementato con orientazione apri a 45° e lunghezza d'onda pari a 10. Sono visibili sia la componente gaussiana che la sinusoidale	41
2.15	Rappresentazione dello stesso filtro ma in 2D: l'onda ha forma simile alle wavelet, con ampiezza attenuata dalla funzione di Gauss.	41
2.16	Risultato del filtraggio dell'immagine lungo le quattro direzioni scelte e lunghezza d'onda pari a 10.	42
2.17	Risultato del filtraggio nelle stesse direzioni ma con lunghezza d'onda pari a 5	42
2.18	Rappresentazione dell'ellisse equivalente, a destra, e dell'inviluppo complesso, a sinistra, per una lesione del Dataset. Per confronto è stato riportato sempre il contorno manuale in ciano.	45
2.19	I grafici riportano due profili di lesioni a confronto, tracciati in base alla distanza dal centroide della regione e alla posizione angolare dei punti del contorno.	47
2.20	Features in ordine decrescente in base al valore di F. Sono riportati inoltre il valore critico come soglia orizzontale e il numero di features selezionate come soglia verticale	50
2.21	Esempio di divisione degli elementi di due classi con l'analisi di discriminazione lineare [50]	53
2.22	Esempio di modello SVM per la classificazione binaria [47].	55
2.23	Esempio di struttura di una rete ANN [47]	57
2.24	Tabella riassuntiva del setting dei classificatori con le migliori combinazioni di parametri	59
2.25	Esempio della struttura di una rete convoluzionale [31].	60
2.26	Distribuzione delle dimensioni delle ROI del Dataset con indicazione del valore medio e del valore scelto per il rescaling	62

2.27	Esempi di ROI in seguito al data augmentation, con indicazione della classe di appartenenza: verde = Benigne, rosso = Maligne.	63
2.28	Schema di rappresentazione di una Confusion Matrix	66
2.29	Esempi di curve ROC.	69
2.30	Esempi di curve Precision - Recall [60].	69
3.1	Grafici di confronto per i valori medi di AUC al variare del numero di dati, da 96 a 174 immagini, e della tipologia di immagini, indicando con 2D le immagini sintetiche e con 3D le tomosintesi	71
3.2	Grafici di confronto per i valori medi di Accuracy e Recall	72
3.3	Composizione del gruppo di features utilizzate per la classificazione delle lesioni su 96 e 174 immagini 2D sintetiche (A e B) e su 96 e 174 immagini di tomosintesi (C e D).	74
3.4	Confusion Matrix dei quattro metodi su immagini 2D sintetiche e tabella riassuntiva dei valori ottenuti per le diverse metriche utilizzate, nell'ordine: Recall, Specificity, Precision, F1-Score, Accuracy, AUC e AUC della curva Precision-Recall (Test 1). In grassetto i valori massimi ottenuti per ogni metrica.	76
3.5	Confronto delle performance ottenute dividendo le proiezioni e lavorando solamente con le proiezioni CC (a) o con le MLO (b). In grassetto i valori massimi ottenuti per ogni metrica.	76
3.6	Confusion Matrix e tabella riassuntiva dei valori ottenuti per le diverse metriche utilizzate (Test 2).	78
3.7	Curve di performance a confronto per i quattro metodi di classificazione e in base ai Test eseguiti.	79
3.8	Test 1: ROC sui sub-fold del 5 Fold Cross-Vaidation	80
3.9	Confronto dei valori medi di AUC, Accuracy e Recall ottenuti su immagini 2D sintetiche.	80
3.10	Test 2: ROC sui sub-fold del 5 Fold Cross-Vaidation	81
3.11	Confronto dei valori medi di AUC, Accuracy e Recall ottenuti per il Test 2.	81

3.12	Feature selection per i due Test effettuati	82
3.13	Confusion Matrix dei quattro metodi su immagini di tomosintesi e tabella riassuntiva dei valori ottenuti per le diverse metriche utilizzate. In grassetto i valori massimi ottenuti per ogni metrica.	84
3.14	Confronto delle performance ottenute su immagini di Tomosintesi dividendo le proiezioni e lavorando solamente con le proiezioni CC (a) o con le MLO (b). In grassetto i valori massimi ottenuti per ogni metrica.	85
3.15	Curve di performance a confronto per i quattro metodi di classificazione su immagini di Tomosintesi.	87
3.16	Confronto delle strategie di calcolo dei parametri mediati su 3, 7 o 9 slices o estratti solo dalla key-slice. In grassetto i valori massimi ottenuti per ogni metrica.	87
3.17	ROC sui sub-fold del 5 Fold Cross-Validation	88
3.18	Confronto dei valori medi di AUC, Accuracy e Recall.	88
3.19	Features selection per la classificazione delle immagini BTO.	89
3.20	Andamento dell'Accuracy e della Loss in fase di allenamento della CNN.	91
3.21	Test Multimodale : Confusion Matrix dei quattro metodi.	92
3.22	Confronto delle performance ottenute su immagini di Tomosintesi dividendo le proiezioni e lavorando solamente con le proiezioni CC (a) o con le MLO (b).	92
3.23	Curve di performance a confronto dei quattro metodi per il Test Multimodale.	93
3.24	Test multimodale: Confronto dei valori medi di AUC, Accuracy e Recall.	94
3.25	Feature Selection per la classificazione delle immagini in modalità multimodale.	94
3.26	Risultati ottenuti in termini di AUC, Accuracy e Recall con il miglior classificatore, SVM, sul training e sul validation mettendo a confronto le performance con immagini 2D sintetiche e con le tomosintesi.	95

3.27	Confronto tra la curva AUC ottenuta sulle performance di 101 radiologi [61] su immagini DM e la curva ROC dell'SVM sul Validation set.	97
------	--	----

Acronimi

AI Artificial Intelligence

AUC Area Under Curve

BTO Breast Tomosynthesis Object

CC Cranio Caudale

DM Digital Mammography

FFDM Full Field Digital Mammography

MLO Medio Laterale Obliquo

SM Synthetic Mammography

Capitolo 1

Introduzione

Nel seguente capitolo si fa riferimento all'anatomia e alle principali patologie a carico della mammella, all'approccio clinico e ai metodi normalmente adottati per la detection e la diagnosi del carcinoma mammario indispensabili alla comprensione dei metodi utilizzati e delle finalità dello studio. La descrizione comprende anche dettagli relativi alle tecniche di imaging coinvolte, mammografia e tomosintesi, e sulle tipologie di immagini utilizzate. Infine, una panoramica sulle applicazioni che coinvolgono le immagini di tomosintesi, la caratterizzazione e la classificazione di lesioni in letteratura, permette di inquadrare gli obiettivi e le scelte fatte nell'ambito delle metodologie.

1.1 Nozioni di anatomia

La ghiandola mammaria

La mammella è un organo pari costituito da una principale componente ghiandolare, la ghiandola mammaria, da tessuto adiposo e tessuto di sostegno. È situata nella zona anterosuperiore del torace e occupa la regione anteriore del muscolo pettorale dal quale è separata da uno strato di tessuto connettivo lasso, o spazio retromammario. È delimitata superiormente dalla seconda o terza costola e inferiormente dalla sesta o settima a seconda della conformazione, mentre si considera come limite mediale la linea margino-sternale e come laterale la linea

ascellare media. Essa presenta caratteristiche estremamente variabili per forma, dimensione e densità del tessuto che dipendono da una serie di fattori, tra i quali l'età e lo stato ormonale che ne comportano continue modifiche, ma anche la fisionomia del soggetto, lo sviluppo della componente ghiandolare e la quantità di tessuto adiposo (Figura 1.1).

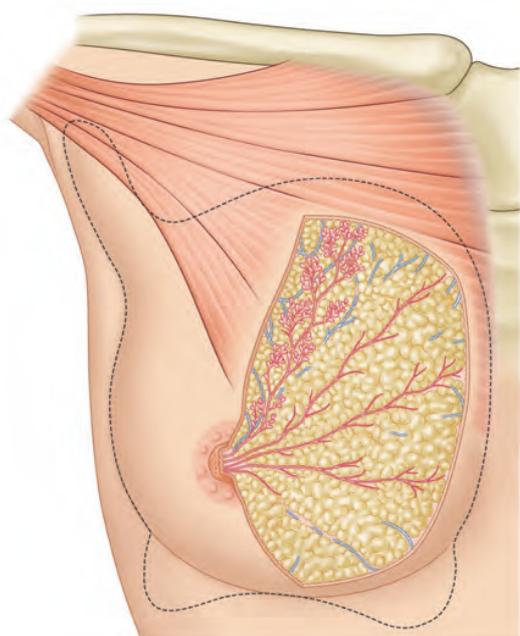


Figura 1.1: Anatomia della mammella e rappresentazione della struttura interna [1].

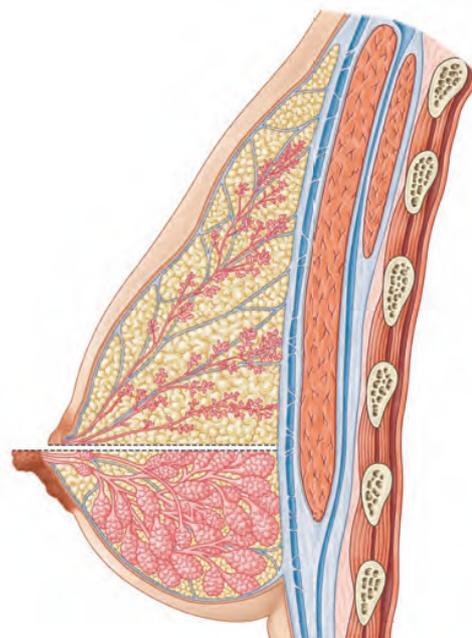


Figura 1.2: Rappresentazione del cambiamento di conformazione della mammella a riposo e durante l'allattamento [1].

Nel soggetto sano, il rivestimento cutaneo della mammella si presenta regolare e privo di avvallamenti o retrazioni. Il colorito, normalmente uniforme e pallido, costituisce un ulteriore elemento per l'analisi morfologica in quanto la presenza di rossori, edema della cute o ulcere può essere ricondotta a patologie anche di natura maligna [2]. Al centro si trova il complesso areola-capezzolo, ossia un'area pigmentata costituita esternamente da cute e internamente da tessuto sottocutaneo e

muscolatura liscia che ne permette la contrazione. È inoltre ricca di ghiandole sebacee modificate e adibite alla lubrificazione della zona durante l'allattamento. Il capezzolo è posizionato al centro dell'areola e, esattamente come accade per l'intero organo, anch'esso può avere caratteristiche decisamente diverse a seconda del soggetto per quanto concerne il posizionamento, la dimensione e la sporgenza oltre a cambiare la sua conformazione nelle varie fasi della vita della donna.

La ghiandola mammaria è esocrina ed è una ghiandola sudoripara modificata. La componente parenchimale è costituita da lobi, di norma 15-20 nella donna adulta, che comprendono dotti e lobuli escretori avvolti da tessuto connettivo. L'unità funzionale della ghiandola è rappresentata dalle unità lobulari duttali terminali (TDLU), dalle quali dipartono i dotti secondari che convergono nei dotti galattofori. A loro volta, questi ultimi confluiscono nel capezzolo, sulla superficie del quale si osserva da una decina ad una dozzina di orifizi, i pori lattiferi, che ne costituiscono i punti di sbocco indipendenti. La componente stromale della ghiandola invece è costituita principalmente da tessuto connettivo. Lo stroma intralobulare risulta privo di fibre elastiche e avvolge i dotti, i lobuli e i TLDU proteggendoli, mentre lo stroma interlobulare circonda i dotti maggiori e comprende sia tessuto connettivo che adiposo. Il tessuto presenta anche strutture più definite che prendono il nome di legamenti sospensori della mammella, il cui fine principale è il sostegno della struttura.

La componente ghiandolare della mammella si sviluppa completamente durante la pubertà: l'azione ormonale degli estrogeni porta alla formazione dei lobuli e, contemporaneamente, si ha la ramificazione dei dotti e lo sviluppo del tessuto connettivo a sostegno del tessuto ghiandolare che, in questa fase della crescita, risulta essere molto denso. Durante tutto il periodo fertile, il tessuto subisce ulteriori cambiamenti legati alla variazione di concentrazione di estrogeni e progesterone che, rispettivamente, inducono e limitano la generazione di epitelio dei dotti e dello stroma [1]. È nuovamente l'azione ormonale a stimolare l'ipertrofia della ghiandola (Figura 1.2) già a partire dal primo trimestre di gravidanza e fino alla fase dell'allattamento, con conseguente aumento

volumetrico e cambiamento della conformazione del complesso areola-capuzzolo: a livello costitutivo, la componente ghiandolare predomina su quella stromale al contrario di ciò che accade in condizioni normali, si ha un aumento della vascolarizzazione e la crescita del sistema di dotti su azione degli estrogeni. Si ha infine il fenomeno della lattazione.

Vascolarizzazione e drenaggio linfatico

La vascolarizzazione della mammella è garantita da numerosi rami arteriosi, venosi e linfatici. L'arteria ascellare alimenta l'arteria toracica superiore, la toracica laterale, la toracoacromiale e la sottoscapolare, mentre altri vasi si diramano dall'arteria toracica interna e dalle arterie intercostali, in particolare la seconda e la quarta. Il sangue deossigenato percorre le vene mammarie interne e laterali fino alla vena ascellare, alla toracica interna e alle intercostali che decorrono parallelamente ai vasi arteriosi. Il drenaggio linfatico è attribuibile a tre principali gruppi di vasi e linfonodi: i nodi ascellari, i parasternali e gli intercostali, tutti in posizione omolaterale. I linfonodi ascellari sono responsabili del 75% del drenaggio totale della mammella, si trovano lungo la vena ascellare e di norma si contano da 30 a 60 unità linfonodali strutturate in tre sottolivelli, come riportato in Figura 1.4.

- Del primo livello fanno parte i linfonodi anteriori o pettorali, che drenano la linfa dalla zona della mammella e dalla pettorale, i linfonodi posteriori o sottoscapolari e i laterali o brachiali.
- Il secondo livello comprende i nodi del gruppo centrale, localizzati nel tessuto adiposo della zona ascellare e che drenano la linfa dai nodi anteriori, posteriori e laterali verso gli apicali, e parte degli infraclavicolari che collegano i nodi centrali e il dotto linfatico destro. Entrambi i gruppi sono analizzati nell'ottica della diagnosi differenziale di patologie maligne potenzialmente metastatiche o mastiti e infezioni.

Inoltre, i linfonodi ascellari rappresentano la sede più probabile per la diffusione di metastasi del tumore alla mammella: si parla di linfonodi sentinella, analizzati in ambito diagnostico soprattutto per la stadiazione

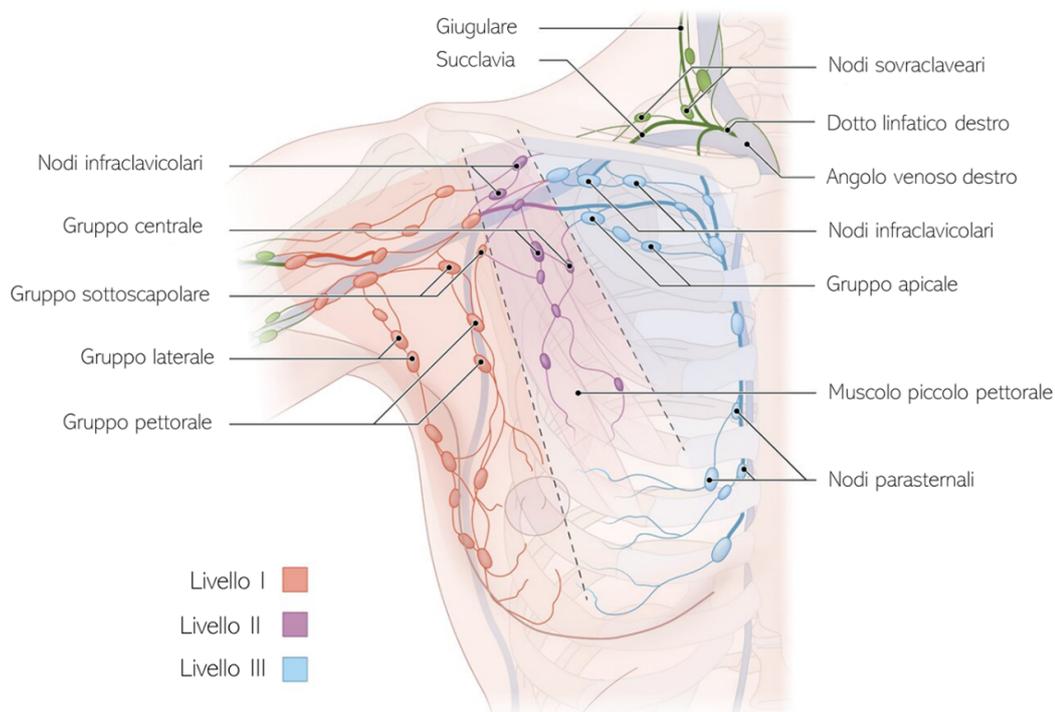


Figura 1.3: Rappresentazione e indicazione della nomenclatura dei principali vasi e nodi del sistema linfatico nella regione toracica divisi nei tre livelli di appartenenza [3].

del tumore. I linfonodi parasternali, invece, sono situati lungo le arterie toraciche interne, sono compresi nel terzo livello e drenano la regione mediale della mammella, mentre i nodi intercostali posteriori sono responsabili del 5% del drenaggio residuo [4].

1.1.1 Anatomia chirurgica

Immaginando di tracciare due assi perpendicolari passanti per il centro del complesso areola-capezzolo, rispettivamente in direzione medio-laterale e cranio-caudale, si ottiene la suddivisione della mammella in quadranti normalmente utilizzata in ambito diagnostico, chirurgico e in ambito tecnico per il corretto posizionamento del soggetto poiché questo comporta la corretta esecuzione dell'esame di screening. Da questa

suddivisione si ottengono due quadranti esterni, indicati come Q1 e Q3 o come quadrante superiore esterno (QSE) e quadrante inferiore esterno (QIE), e due quadranti medialti, Q2 e Q4 ai quali ci si riferisce anche come quadrante superiore e inferiore interno (QSI e QII). Oltre ai quattro quadranti radiali, si considerano normalmente un quadrante centrale che comprende il complesso areola-capezzolo e il parenchima retroareolare, o Q5, e un ulteriore settore in posizione craniale rispetto al quadrante superiore esterno, in corrispondenza del prolungamento ascellare, o Q6.

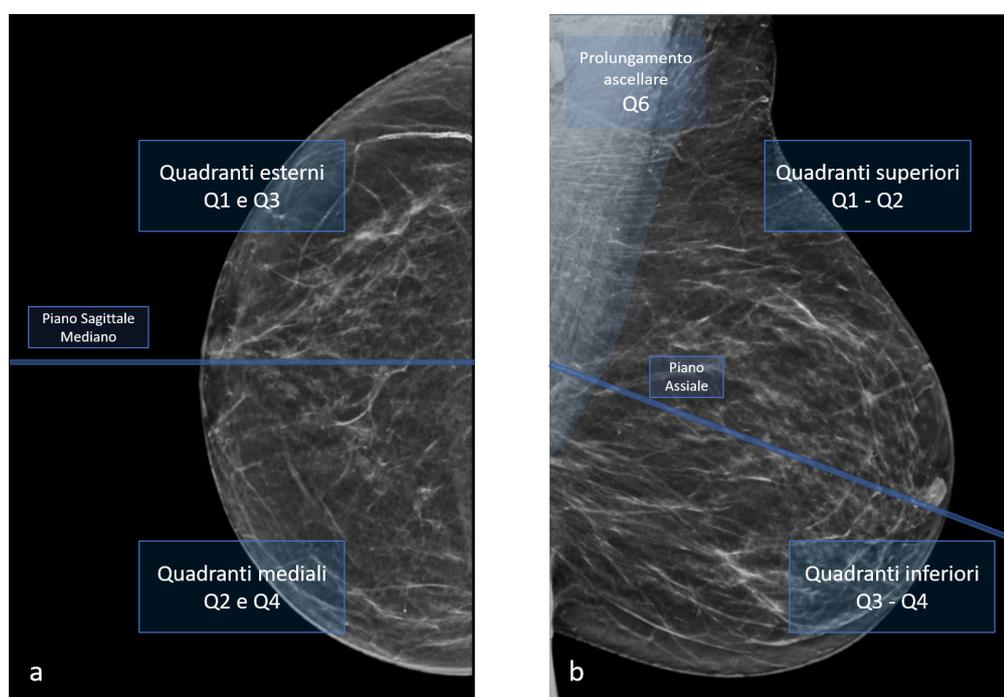


Figura 1.4: Mammografia destra in proiezione cranio caudale (a) e mammografia sinistra in proiezione medio laterale obliqua (b) con indicazione dei quadranti visibili.

La divisione in settori risulta utile in ambito diagnostico in quanto risulta più semplice localizzare in modo univoco la lesione, soprattutto in caso di necessità di ulteriori accertamenti, esami diagnostici o interventi chirurgici e nel follow-up. Anche in ambito tecnico vengono utilizzati per fornire le indicazioni necessarie al corretto posizionamento della

paziente. In proiezione cranio caudale, i quadranti Q1 e Q2 così come Q3 e Q4 risultano sovrapposti. Al contrario, in proiezione medio laterale obliqua, si ha nuovamente sovrapposizione dei quadranti ma con una diversa angolazione, ed è possibile visualizzare più facilmente il Q6.

1.2 Il carcinoma mammario

Il cancro alla mammella è la neoplasia più diffusa nelle donne, nonché una delle principali cause di decesso e, in generale, il più diagnosticato anche nei soggetti prevalenti. Nonostante in molti casi si manifesti con assenza di sintomi ben identificabili, ad oggi, le terapie e le tecniche chirurgiche a disposizione per la cura e la rimozione del tumore risultano efficaci anche grazie alla capacità, nella maggior parte dei casi, di diagnosticare il tumore in fase iniziale.

1.2.1 Incidenza e fattori di rischio

Secondo i dati raccolti dall'AIRTUM [5] nel periodo che va dal 2008 al 2016, il tumore della mammella nella donna è la neoplasia con incidenza maggiore, oltre ad essere la più frequentemente diagnosticata in tutte le classi di età. Nello stesso arco temporale, si registra, in generale, un aumento nel numero di casi di carcinoma mammario di circa lo 0,3% annuo, più significativo nei soggetti sotto i 50 anni per i quali si ha un incremento annuo dell'1,6%. Nel 2018, in Italia, su un totale di 1.825.283 donne sottoposte all'esame di screening, sono stati diagnosticate 1036 lesioni benigne e 8045 lesioni maligne, di cui 1069 tumori non infiltranti, con un'incidenza dello 0,059% sul totale degli esami, e quasi 2500 tumori invasivi ≤ 10 mm con lo 0,137% di incidenza. Dalle linee guida dell'AIOM, nel 2019 si registra una percentuale di diagnosi di tumore alla mammella pari al 40% nelle pazienti di età inferiore ai 49 anni, 35% nella fascia tra i 50 e i 69 anni e 22% al di sopra dei 70 anni. Infine, dai dati ISTAT, nel 2017 il tumore della mammella ha costituito il 16,1%, nonché la prima causa di tutte le morti oncologiche per le donne [5].

	2012-2013	2014-2015	2016-2017	2018
Totale inviti	5312947	5920807	6482301	3364979
Adesioni (% sul numero di inviti)	3044984 (57%)	3305361 (56%)	3605948 (56%)	1825283 (54%)
Esami	2975296	2975296	3541039	182285
Esami di approfondimento (% sul numero di esami)	172377 (5.8%)	191574 (5.81%)	215687 (6.09%)	115057 (6.31%)
Biopsie benigne	1582	1874	1985	1036
Carcinomi diagnosticati allo screening	13811	15606	16392	8045

Figura 1.5: Analisi a livello nazionale dell'attività di screening mammografico 2012-2018 [6].

Nel 2020, in Italia, si stima che questa patologia colpisca circa una donna su nove, con una percentuale del 13% di possibilità di ammalarsi durante il corso della vita. In particolare, nella fascia di età inferiore ai 50 anni, una donna su 45 rischia di contrarre la malattia.

Di fronte a questi numeri è chiaro come i programmi di screening mammografico siano stati e continuino ad essere indispensabili, soprattutto nell'ottica di un possibile aumento del tasso di incidenza nei prossimi anni. In base al rapporto dell'osservatorio nazionale dello screening mammografico del 2019, il numero di donne con invito a effettuare un esame di screening è in aumento ma il rate di adesione resta pressappoco costante al di sopra del 50%. Sul numero totale di esami effettuati, la percentuale di pazienti richiamate per un ulteriore accertamento nel 2018 si attesta al 6,3% e in leggero aumento [6]. Tutti i valori sono riportati nella tabella di Figura 1.5. Inoltre, la sensibilizzazione sul tema, la maggiore consapevolezza, il miglioramento delle tecniche di screening e terapeutiche a disposizione e l'aumento del numero di programmi di screening oncologico a livello nazionale oltre allo screening spontaneo, hanno portato ad un aumento del tasso di sopravvivenza: nel 2019 il tasso di sopravvivenza netta a 5 anni dalla diagnosi del tumore alla mammella era dell'87%. Nel 2017 si contavano 3.304.648 soggetti prevalenti, ossia in vita in seguito alla diagnosi di un tumore, tra i quali 834.000 donne alle quali era stato diagnosticato un carcinoma mammario.

Non sempre le neoplasie mammarie si manifestano con sintomi univocamente riconoscibili, al contrario possono essere asintomatiche o ancora presentare caratteristiche comuni in caso di benignità e malignità, quindi vengono comunemente diagnosticate a seguito di un esame di screening, sulla base di un'ispezione mammaria e in base all'anamnesi della paziente che comprende l'analisi della familiarità, della situazione clinica, delle terapie pregresse o di eventuali esami precedenti oltre alla valutazione dei fattori di rischio.

Esistono sufficienti evidenze dell'influenza di alcuni stili di vita o altri fattori esterni sulla comparsa di neoplasie, e per quanto riguarda il tumore al seno, si fa riferimento principalmente ai seguenti fattori di rischio:

- Il consumo di alcol è considerato dall'AIRC come “sicuramente cancerogeno per l'uomo”. Esiste una correlazione tra l'assunzione importante di alcol e lo sviluppo di neoplasie ma è determinante, l'associazione ad altri fattori di rischio.
- Lo stile di vita sedentario, la dieta e i fattori metabolici: l'attività fisica, soprattutto per le donne in post menopausa, è un fattore determinante per la protezione dall'insorgenza di patologie, compreso il carcinoma mammario [6, 7].
- L'assunzione di specifici farmaci può incidere sulla possibile insorgenza del tumore, in particolare esistono evidenze significative riguardo l'assunzione di contraccettivi orali, che contengono estrogeni e progesterone, ma soprattutto la terapia ormonale sostitutiva o HRT [5].
- L'essersi sottoposte a radioterapia in passato, oppure l'aver già sviluppato neoplasie mammarie.
- L'età avanzata e la predisposizione, principalmente sulla base di fattori di metabolismo, dietetici, ormonali, riproduttivi e di familiarità: la probabilità per una donna con uno o più parenti prossimi che abbiamo sviluppato la patologia di ammalarsi è significativamente più elevata. Altri dettagli valutati possono essere l'età della paziente alla prima gravidanza, oppure la densità del tessuto in quanto

la maggior densità è spesso correlata all'aumentata probabilità di sviluppare il tumore [8].

- Alcune mutazioni genetiche, considerate come fattori di rischio non modificabili in quanto ereditari o comunque presenti alla nascita. In particolare, si considera che le mutazioni dei geni BRCA 1 e BRCA 2 siano in grado di aumentare la probabilità di insorgenza del carcinoma mammario fino all'80% e di causare circa il 20% dei casi di carcinoma ereditario. [9, 10, 11].

Alcuni dei fattori elencati vengono valutati nella fase di anamnesi della paziente e influiscono sulla diagnosi ma, come già citato, non tutti sono modificabili e non vi è correlazione assoluta tra la malattia e uno qualsiasi degli aspetti elencati.

1.2.2 Masse e opacità maligne e benigne

Spesso le neoplasie si sviluppano nei dotti, attraverso i quali crescono e si diffondono, ma possono avere origine da diversi tipi di cellule, tra cui lobulari, tubulari, cribriformi o mucinose [12, 13]. Si distinguono due principali tipologie:

- Tumori in situ: sono tumori non invasivi, non coinvolgono altri tessuti e non producono metastasi. Spesso, soprattutto se sono in fase iniziale, sono di piccole dimensioni. I tumori intraduttali e intralobulari (LCIS) ne rappresentano il principale esempio.
- Tumori infiltranti: sono forme invasive che tendono ad espandersi coinvolgendo diversi tessuti fino ad evolvere in forme metastatiche. Si sviluppano più facilmente se le cellule tumorali cominciano a proliferare in prossimità del tessuto all'esterno dei dotti e dei lobuli. Forme comuni sono il carcinoma lobulare infiltrante (ILC) e il duttale infiltrante.

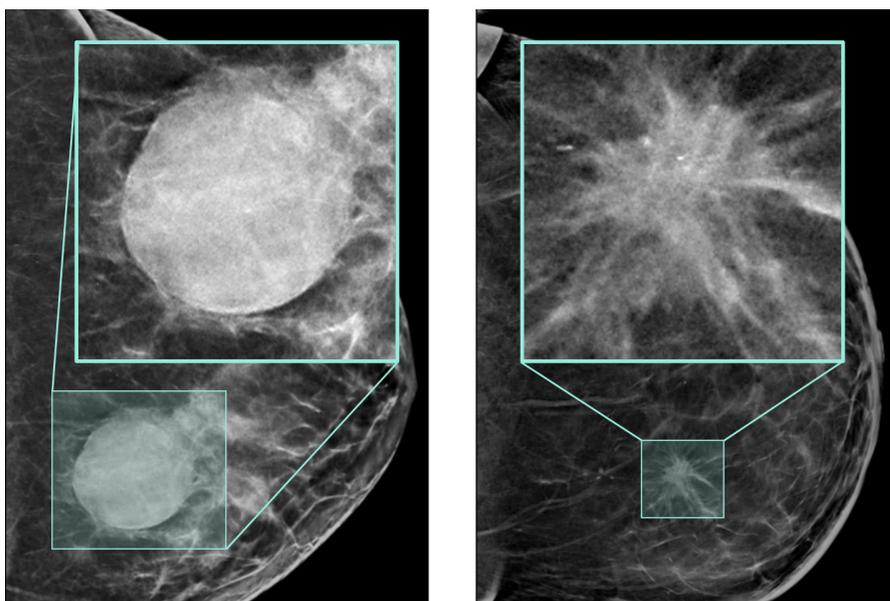


Figura 1.6: Esempi di opacità di natura maligna osservate su immagini di mammografia

Le lesioni di natura benigna possono avere le stesse origini e sviluppare caratteristiche molto simili. Nella categoria delle lesioni epiteliali benigne rientrano:

- Patologie infiammatorie della mammella: tra queste, le steatonecrosi sono cisti derivanti dallo stato di necrosi di una porzione di tessuto adiposo non più irrorato dai vasi sanguigni e possono formarsi in seguito ad interventi chirurgici.
- Lesioni non proliferative: comprendono cisti a contenuto liquido o solido, le quali presentano spesso un contorno circolare o comunque molto regolare e netto dovuto allo strato di epitelio piatto che le racchiude e dimensioni estremamente variabili, così come la conformazione, anche a causa della loro tendenza a crescere. Anche le microcalcificazioni, le adenosi e le fibrosi rientrano nella stessa categoria. [14].
- Lesioni stromali: i fibroadenomi (FAD) sono lesioni stromali che hanno origine nello stroma intralobulare. Sono spesso palpabili, a

contenuto solido e possono presentare una componente calcifica più o meno estesa.

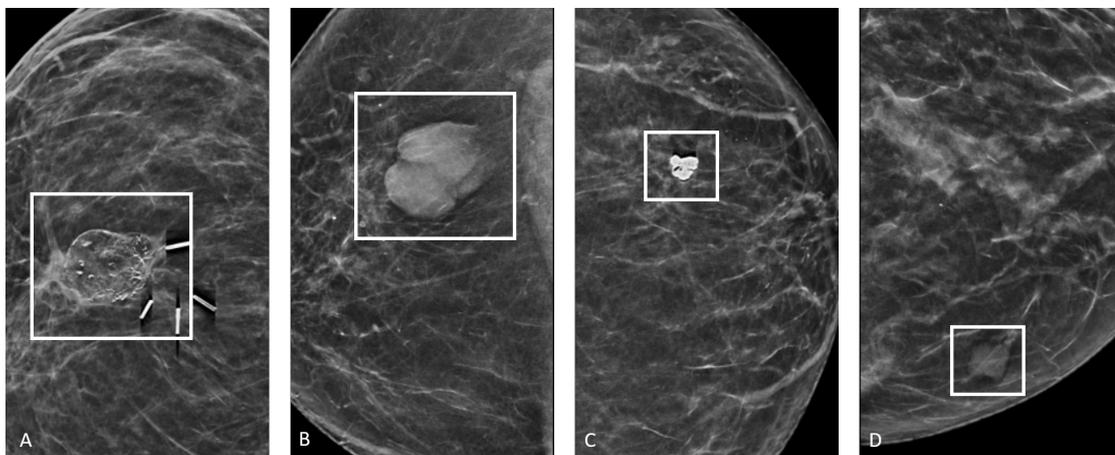


Figura 1.7: Esempi di opacità di natura benigna: (A) Steatonecrosi con clips chirurgiche visibili; (B) Cisti settata; (C) FAD calcifico; (D) FAD.

Le patologie indicate sono definite *opacità nodulari*, e sono il segno più comune di patologia mammaria [15]. Il termine *massa*, invece, viene utilizzato per definire una lesione palpabile a prescindere dalle dimensioni. Le opacità osservabili nelle immagini mammografiche possono avere forme diverse, contorni più o meno netti e intensità variabile. Ad esempio, le cisti a contenuto liquido si manifestano come opacità con intensità minore rispetto alle cisti a contenuto lipidico, mentre i FAD hanno forme più regolari se a contenuto lipidico o più irregolare e dal colore molto chiaro e omogeneo se calcifiche [16]. In generale, se le opacità benigne sono caratterizzate da forme più tondeggianti e contorni più definiti, le maligne si presentano come zone ad alta intensità con contorni poco netti e molto irregolari. Sono considerati indici di malignità le spicule, ossia irregolarità del contorno della lesione che si protendono verso il tessuto circostante e che rappresentano le infiltrazioni del tumore, e le microcalcificazioni, punti ad altissima intensità all'interno della regione di interesse. Un altro segno caratteristico può essere l'aumento

di spessore e di intensità del bordo della mammella, per un mancato drenaggio efficiente dei liquidi. In molti casi risulta complesso diagnosticare una lesione come francamente maligna in quanto le proprietà valutabili visivamente possono essere estremamente variabili e talvolta molto simili ai tratti distintivi di una lesione benigna.

1.2.3 Esame obiettivo e ispezione mammaria

L'esame obiettivo fornisce una prima valutazione dello stato di salute della paziente e può essere eseguito in seguito a riportata sintomatologia o come visita di controllo nel caso di screening volontario. Di norma, la visita è preceduta dall'anamnesi: si tratta di una fase di raccolta dei dati anamnestici in base alla natura delle quali, si fa distinzione tra:

- Anamnesi personale: comprende tutto ciò che riguarda l'aspetto ormonale e riproduttivo della paziente, oltre all'eventuale assunzione di farmaci o l'uso di terapie ormonali. Rientrano in questa categoria anche le informazioni sullo stile di vita e altri fattori di rischio.
- Anamnesi patologica remota: tutto ciò che è riportato nella cartella clinica, quindi patologie della mammella ed esiti degli esami diagnostici fatti.
- Anamnesi patologica prossima: tutto ciò che può essere di rilevanza per la diagnosi oltre alla sintomatologia, ossia tutti i dettagli notati dalla paziente e gli esiti degli esami diagnostici eseguiti per accertamento prima della visita.

L'anamnesi costituisce una componente fondamentale nella diagnosi del carcinoma mammario, nonostante la correlazione con l'effettivo sviluppo di una neoplasia sia di per sé molto bassa. Durante l'ispezione mammaria si analizza l'eventuale sintomatologia riportata dalla paziente e si valutano eventuali anomalie nel seno che possono in alcuni casi essere indizio dell'insorgenza del tumore: asimmetria marcata del seno, alterazioni del profilo mammario, colorito della pelle, presenza di eritema o edema cutaneo, presenza di ulcere o alterazioni dell'epitelio. In questa

fase si individuano le lesioni palpabili, quali noduli solidi o liquidi, e se ne valutano alcune caratteristiche principali, tra cui la dimensione e la mobilità. Rientra in questa fase anche il controllo dei linfonodi nel cavo ascellare e nelle fosse sovraclavari dei quali si verifica la mobilità e la piccola dimensione tipiche nel soggetto sano.

1.2.4 Diagnostica per immagini

Le tecniche di imaging oggi utilizzate per la diagnosi del carcinoma alla mammella sono diverse anche per via dell'eterogeneità che caratterizza le lesioni. La mammografia costituisce la tecnica più diffusa, nonché quella che per più tempo ha rappresentato il metodo di elezione per l'esecuzione dell'esame di screening, ma recentemente la tomosintesi ha preso piede grazie al perfezionamento della tecnologia e alla sua fondamentale capacità di eliminare l'effetto di sovrapposizione dei tessuti e facilitare in molti casi la visualizzazione e la comprensione. La mammografia sintetica o SM è invece una metodologia molto più recente che nasce con lo scopo principale di sostituire la mammografia digitale (DM) introducendo numerosi vantaggi.

Mammografia

L'esame di screening mammografico è un esame radiologico mirato all'indagine del tessuto mammario per analizzarne le varie componenti e riuscire ad individuare eventuali anomalie [2]. Si utilizza il mammografo, un dispositivo dedicato che comprende un tubo radiogeno mobile per la produzione dei raggi X e un dispositivo di compressione. I raggi X sono radiazioni che, avendo lunghezza d'onda compresa tra 100 pm e 10 pm, sono in grado di penetrare all'interno dei tessuti con minima attenuazione dovuta all'assorbimento. In particolare, il tessuto neoplastico presenta un coefficiente di assorbimento molto simile a quello del tessuto fibroso, ma leggermente maggiore ad energie più basse: per questo motivo si lavora con valori di energia del fascio inferiori ai 50 keV.

Con il dispositivo di compressione la mammella viene compressa a contatto con il rivelatore, al fine di diminuirne lo spessore e di conseguenza l'attenuazione del fascio. Essendo i raggi X radiazioni ionizzanti, è fondamentale introdurre il concetto di dose, ossia la quantità di radiazioni ricevute dal soggetto: in particolare, si definisce dose assorbita il rapporto tra l'energia totale trasferita dalla radiazione e l'unità di materia.

La mammella viene posizionata secondo specifiche indicazioni, al fine della corretta visualizzazione del volume in tutte le sue parti, evitando la formazione di pieghe e mantenendo il muscolo pettorale disteso. L'esame di screening prevede l'acquisizione in due proiezioni per massimizzare la visualizzazione di eventuali anomalie: si acquisisce in proiezione *cranio caudale*, indicata come CC, e in proiezione *medio laterale obliqua* o MLO, per entrambi i seni ottenendo quattro immagini in totale. Nella mammografia di secondo livello, ossia nel caso in cui venga eseguita in fase di accertamento, si può scegliere di acquisire solamente una proiezione medio laterale o ML del seno che presenta l'anomalia al fine di ridurre la dose somministrata alla paziente.

Il principale limite della mammografia è la sovrapposizione dei tessuti, che spesso preclude l'individuazione di elementi sospetti. Inoltre, la tecnica non permette di analizzare in modo esaustivo alcune tipologie di lesioni ed è quindi necessario ricorrere all'utilizzo di tecniche di imaging complementari quali la risonanza magnetica o l'ecografia al fine della risoluzione dei dubbi diagnostici o per approfondimenti sulla conformazione e sulle caratteristiche della lesione.

Tomosintesi

La tomosintesi mammaria è una tecnica relativamente recente che è in grado di fornire una riproduzione quasi tridimensionale della mammella. È stata inizialmente approvata dalla FDA come tecnica da affiancare alla mammografia tradizionale e, ad oggi, rimane comunque necessaria l'acquisizione di un'immagine 2D insieme alla tomosintesi. Essa infatti non è considerata, ad oggi, una metodica valida per lo screening. Il principio di funzionamento si basa sull'acquisizione, a diversi angoli

di inclinazione del tubo radiogeno, di un numero finito di proiezioni bidimensionali del seno [17]. Successivamente, le proiezioni vengono utilizzate per ricostruire il volume attraverso specifici algoritmi, tra i più frequenti il filtered backprojection e gli algoritmi di ricostruzione iterativi [18]. La dose somministrata è maggiore rispetto a quella della mammografia, ma comunque al di sotto dei limiti massimi prescritti [19].

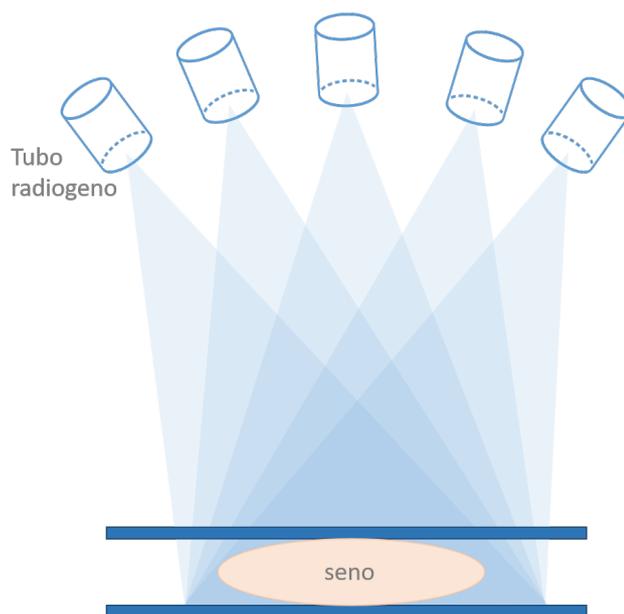


Figura 1.8: Schema del principio di funzionamento della tomosintesi.

Il sistema di acquisizione Hologic comprende un rivelatore in Selenio amorfo, un filtro in Alluminio e un anodo in Tungsteno. Il tubo a raggi X viene fatto ruotare lungo un arco attorno al volume in esame di un angolo di 30° : questa ampiezza viene scelta considerando che il valore incide sulla risoluzione, esattamente come il numero di proiezioni considerate, il che è fondamentale affinché anche le strutture con dimensioni ridotte siano correttamente visualizzate [20]. Esattamente come in mammografia, anche in questo caso la procedura è ripetuta in proiezione cranio caudale e medio laterale obliqua per entrambi i seni.

Il vantaggio principale della tecnica è sicuramente la riduzione dell'effetto di sovrapposizione dei tessuti, in quanto l'intero volume viene

analizzato su piani distinti a diverse profondità. Inoltre, l'alta risoluzione nel piano perpendicolare al verso di acquisizione permette di visualizzare bene la lesione e valutarne le variazioni di conformazione e dimensione al variare della profondità, con la possibilità di individuando una slice principale o key-slice dove la lesione ha diametro maggiore o si vede più nitidamente. Al contrario, alcuni limiti della tecnica sono la bassa risoluzione nel piano dello spessore, quindi parallelamente alla direzione di acquisizione, nonché l'aumento di dose rispetto a quella necessaria per l'esame standard. L'immagine può presentare una serie di artefatti tipici legati alla tecnica di acquisizione, tra questi sicuramente gli artefatti da movimento, ma anche legati all'algoritmo di ricostruzione. Un altro punto critico è costituito dai tempi di lettura delle immagini, che risultano nettamente superiori a quelli necessari in mammografia [21].

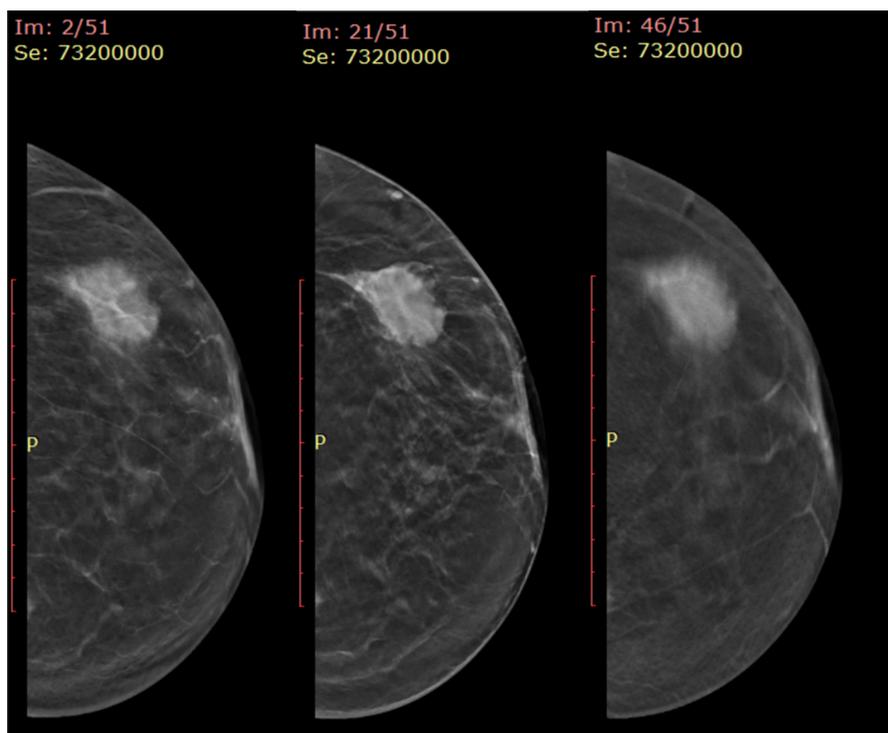


Figura 1.9: Visualizzazione delle singole slices dell'immagine BTO su un visualizzatore DICOM : confronto tra le singole slices e cambiamento nella visualizzazione della lesione.

Software *C-View* e mammografia 2D sintetica

La mammografia 2D sintetica è una ricostruzione bidimensionale dell'immagine di tomosintesi. Nasce nell'ottica di sottoporre le pazienti ad un'unica acquisizione evitando loro di ricevere una doppia dose, ma comunque fornire al medico uno strumento che sia il più possibile equivalente alla mammografia digitale e che, contemporaneamente riesca a migliorarne alcune caratteristiche per una migliore visualizzazione. L'immagine può essere ricostruita a prescindere dalla proiezione e utilizzata come esame complementare alla tomosintesi.

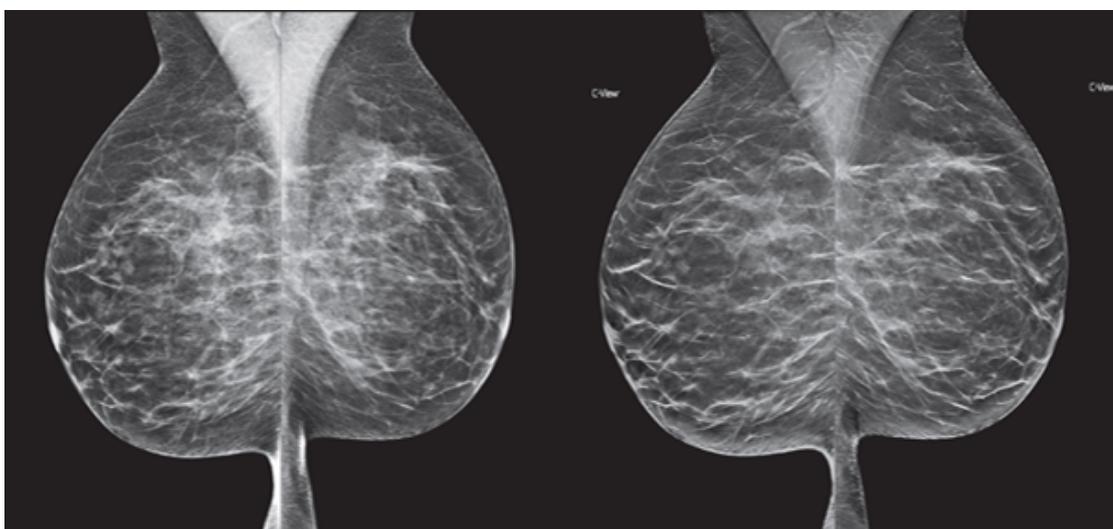


Figura 1.10: Confronto tra immagine mammografica FFDM (Full Field Digital Mammography), a sinistra, e C-View, a destra [22].

L'algoritmo del software *C-View* sviluppato da Hologic si pone come obiettivo quello di ricreare un'immagine 2D sintetica a partire dalle slices della tomosintesi, basando la propria efficacia anche sull'alta risoluzione, pari a circa $100 \mu\text{m}$. Si tratta di una sovrapposizione delle singole slices con enhancing mirato delle strutture di interesse, tra le quali simil-lesioni, microcalcificazioni e strutture lineari che possono essere individuate tramite filtraggio passa alto, con la risultante attenuazione del tessuto fibro-ghiandolare e adiposo a favore di possibili anomalie. In questo modo l'immagine risulta volutamente diversa dalla

mammografia standard (Figura 1.10) poiché cerca di superarne i limiti principali, e per questo presenta una serie di vantaggi [22]. Oltre alla

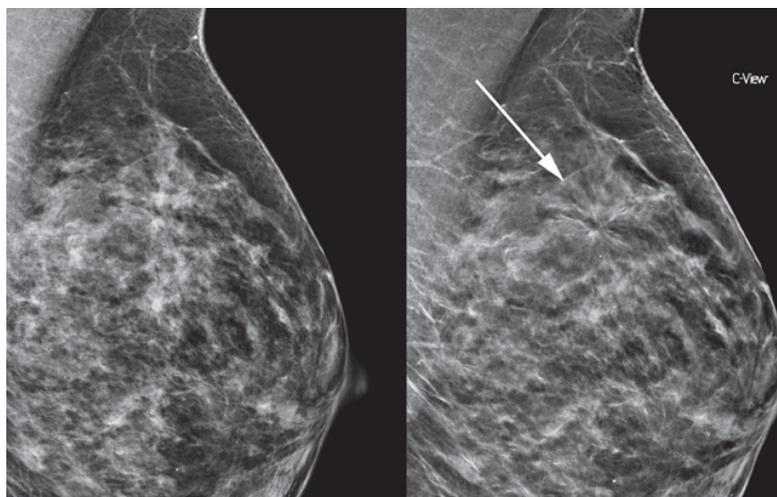


Figura 1.11: Visualizzazione di un tumore duttale infiltrante in DM e SM: evidenza della migliore rappresentazione delle masse spiculate nell'immagine sintetica [22].

fondamentale prerogativa di limitare la dose di radiazioni, la mammografia sintetica fornisce una panoramica della proiezione, facilitando il confronto bilaterale fondamentale per l'individuazione di asimmetrie. Inoltre, si ha una riduzione dei tempi di analisi dell'immagine e di confronto con eventuali esami precedenti, con conseguente aumentato potere diagnostico. La mammografia sintetica in abbinamento alla tomosintesi fornisce risultati confrontabili o migliori anche in termini di recall se comparati a quelli ottenuti con l'abbinamento di tomosintesi e FFDM [23, 24, 25, 26]. In generale, risulta più utile per la detection di microcalcificazioni e la visualizzazione dei bordi spiculati (Figura 1.11), mentre la visualizzazione di masse e simmetrie rimane paragonabile a quella ottenibile con la mammografia digitale abbinata alla tomosintesi. I principali limiti della tecnica sono strettamente connessi a quelli della tomosintesi o vengono introdotti dall'algoritmo utilizzato per generare l'immagine. Tra questi gli artefatti da movimento, difetti legati all'algoritmo di ricostruzione dell'immagine 3D, l'errata accentuazione di strutture fisiologiche e la generazione di artefatti caratteristici come le

pseudocalcificazioni, fino alla difficoltà nella definizione degli effettivi margini di alcune lesioni per effetto della sovrapposizione [27].

Classificazione BI-RADS

La classificazione BI-RADS, acronimo di Breast Imaging - Reporting and Data System, è un ausilio alla pianificazione diagnostico-terapeutica [2].

Classe	Indicazioni
BI-RADS 0	Esame incompleto. Necessità di ulteriori accertamenti, visualizzazione di esami precedenti o esami aggiuntivi.
BI-RADS 1	Esame negativo. Nessuna anomalia e aspetto della ghiandola nella norma.
BI-RADS 2	Esame negativo. Si evidenzia la presenza di patologie francamente benigne.
BI-RADS 3	Esito dubbio. La probabilità che l'anomalia osservata sia maligna è di norma al di sotto del 2%.
BI-RADS 4	Esito dubbio. Osservazione di un'anomalia sospetta con probabilità significativa di malignità. In questo caso, si procede con una biopsia e, in caso di benignità, si esegue comunque un esame aggiuntivo a distanza di un determinato periodo di tempo per scongiurare la possibilità di evoluzione della lesione in patologia maligna.
BI-RADS 5	Esito dubbio. L'anomalia osservata è molto probabilmente maligna, con una percentuale al di sopra del 95%. Si esegue comunque un'esame istopatologico per conferma.
BI-RADS 6	Esito francamente maligno e comprovato dalla Biopsia.

Tabella 1.1: Categorie BI-RADS e breve descrizione in base alle linee guida oncologiche.

Le classi considerate sono riportate in tabella 1.1. In base alla valutazione dell'esame di screening, si assegna una categoria che indichi la presenza di lesioni sospette o meno e l'eventuale necessità di ulteriori controlli. Le categorie più critiche sono quelle intermedie e principalmente la categoria 3 in quanto sebbene risulti molto più probabile che l'esito della biopsia sia benigno.

1.2.5 Diagnosi basata sull'esame istologico

L'esame istologico viene eseguito in caso di lesione o massa sospetta, anche sulla base del livello assegnato dalla classificazione BI-RADS, e permette inoltre di tipizzare un nodulo neoplastico o di delineare le caratteristiche biologiche del tessuto al fine della definizione di una terapia. La procedura più semplice ed efficace nel caso di noduli palpabili e la biopsia tru-cut con assistenza ecografica, mentre la VAB è la procedura da scegliere nel caso di noduli più profondi. È il metodo più diretto per confermare la natura benigna o maligna di una lesione, tuttavia allunga i tempi di elaborazione della diagnosi ed è una procedura che sottopone la paziente ad uno stato di stress importante, per questo idealmente andrebbe eseguita solo nei casi in cui la probabilità di malignità è elevata. Il rapporto dell'ONS del 2019 sottolinea come, nonostante il rapporto tra le diagnosi benigne e maligne sul totale delle biopsie eseguite continui a rimanere al di sotto del valore massimo accettabile senza subire variazioni significative, ad oggi, il numero di biopsie che avrebbero potuto essere evitate è ancora alto e legato al numero di falsi positivi. La percentuale di diagnosi di carcinoma mammario sul totale delle biopsie eseguite è pari a circa il 30% in Europa mentre scende al 20% negli Stati Uniti. Il dato si riconferma in Italia, con circa il 27% di diagnosi di carcinoma su 1000 biopsie in totale, 65% di lesioni benigne e il 6% di lesioni osservate su mammografia di categoria BI-RADS 3.

1.3 Immagini di Tomosintesi e Machine Learning : stato dell'arte

Nonostante la tomosintesi sia una tecnica di imaging ormai largamente utilizzata, le applicazioni che si basano su di essa per la detection e la classificazione automatica di lesioni e anomalie sono molto più recenti e, in generale, meno diffuse rispetto a quelle basate su mammografia. Da un'analisi della letteratura, si evince come ci sia stato, soprattutto negli ultimi anni, un rapido incremento nell'utilizzo di tecniche di Machine learning su immagini di tomosintesi, mentre risulta più recente l'utilizzo di tecniche di Deep learning, in concomitanza con la diffusione e il successo della reti convoluzionali [21]. Infine, l'utilizzo di immagini di mammografia 2D sintetica è ancora limitato per quanto riguarda lo sviluppo di sistemi automatici, ma è stata largamente indagata la possibilità di utilizzarle in sostituzione alla mammografia tradizionale in ambito diagnostico.

I sistemi automatici mirati alla detection di lesioni utilizzano diverse metodologie per identificare una regione sospetta, assegnando uno score di malignità o definendone il contorno. Nel caso di applicazioni a solo scopo diagnostico, invece, la lesione viene identificata preventivamente da uno o più radiologi che ne forniscano la segmentazione manuale o la localizzazione. Tra i principali obiettivi di queste applicazioni, vi è sicuramente la negativizzazione dei falsi sospetti nell'ottica di ridurre il numero di biopsie in pazienti che non presentano effettivamente lesioni maligne: secondo le statistiche già riportate, vi è in effetti un'incidenza di 5 donne con diagnosi di un carcinoma mammario su 1000 che eseguono l'esame di screening, ma all'incirca 100 verranno comunque sottoposte ad ulteriori accertamenti e questo è sicuramente un dato da migliorare. Contemporaneamente, i sistemi diagnostici basati su machine learning si pongono l'obiettivo di:

- Ridurre i tempi di lettura delle immagini di tomosintesi e per l'elaborazione della diagnosi se utilizzati come second reader o come aiuto alla decisione.

- Aumentare la detection delle micro-calcificazioni, le quali risultano meno visibili che in mammografia
- Sfruttare l'informazione 3D per migliorare la detection: la bassa risoluzione sul piano dello spessore rende comunque l'estrazione di informazioni dall'immagine più complessa

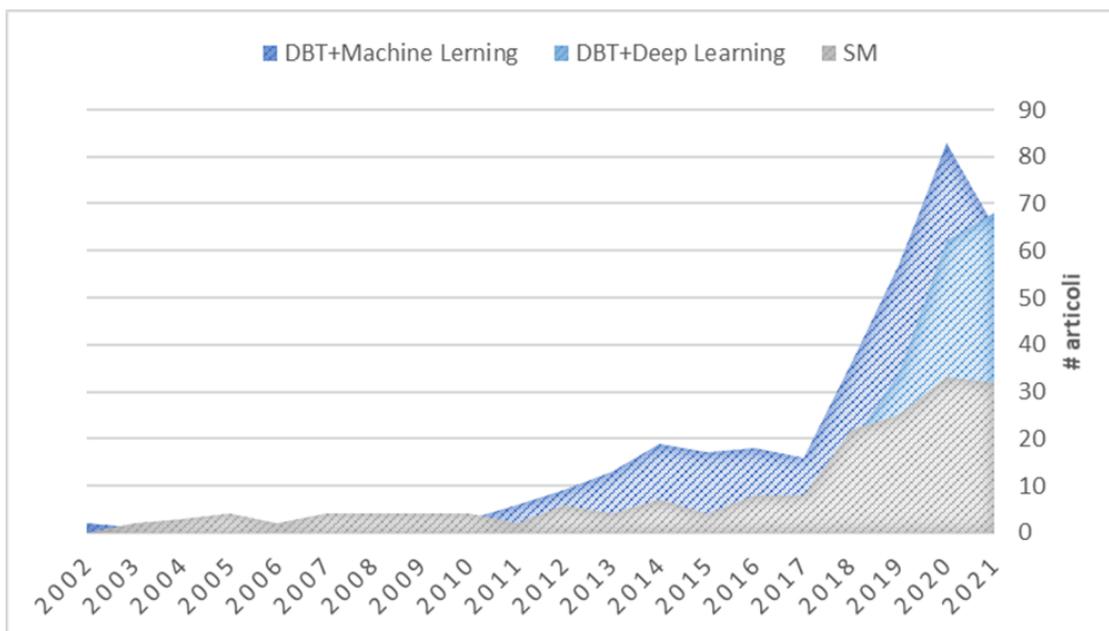


Figura 1.12: Indice del numero di articoli in letteratura per macro-argomento dal 2001 a oggi in base ai dati di NCBI.

Tra le principali criticità da affrontare, vi è la ridotta disponibilità di dataset open source di immagini di tomosintesi che siano bilanciati e che abbiano inoltre numerosità, risoluzione e informazioni adeguate al problema di classificazione. Il problema dello sbilanciamento è strettamente legato all'incidenza del carcinoma mammario, mentre la scarsa disponibilità è probabilmente dovuta al fatto che la tecnica è ancora relativamente recente. Nella maggior parte degli studi in letteratura, le immagini vengono raccolte in seguito a ricerche retrospettive in collaborazione con ospedali o studi clinici dove si effettua l'attività di screening,

e numerosità e casistica dipendono dal livello di attività dei singoli centri. In altri casi si può utilizzare la conoscenza già estratta dalle immagini di mammografia tradizionale per adattare sistemi esistenti alle nuove immagini, tramite la tecnica del Transfer Learning. Un altro limite è rappresentato dalla difficoltà di avere un sistema che riconosca bene sia le masse che le calcificazioni, in quanto presentano caratteristiche molto diverse.

Ad oggi esistono sistemi e CAD con performance equiparabili a quelle di un radiologo che distingue tra patologie maligne e benigne basandosi sulla sola osservazione dell'immagine di mammografia o tomosintesi. Tuttavia, la capacità di discriminazione del radiologo è di norma basata su una serie di informazioni al di là dell'immagine in sé, ed è soggetta a forte variabilità intra e inter-operatore che deve essere considerata al fine della valutazione del sistema. Tenendo conto di queste considerazioni, si evince che ad oggi i risultati raggiunti sia su immagini di tomosintesi che di mammografia non sarebbero abbastanza per migliorare effettivamente le performance dell'attività di screening, ma con un dataset adeguatamente ampio, i risultati potrebbero essere ulteriormente migliorati tramite l'integrazione di diverse informazioni, tra cui dati anamnestici della paziente e caratteristiche degli esami precedenti, per ottenere diagnosi più accurate e precoci.

Capitolo 2

Materiali e Metodi

Nel seguente capitolo sono illustrate tutte le fasi del lavoro svolto. In particolare, vengono descritti i dati a disposizione, le strategie scelte per la caratterizzazione delle lesioni, i test effettuati e le metodologie utilizzate per classificare le lesioni al fine di indagare nel modo più esaustivo possibile le criticità e le potenzialità del progetto.

2.1 Dataset

Lo studio è incentrato sull'analisi di immagini di mammografia 2D sintetica C-View e immagini di Tomosintesi acquisite con il Selenia Dimensions Mammography System (Hologic Inc.). Con l'approvazione e la collaborazione del reparto di Senologia delle Molinette e il supporto tecnico di Tecnologie Avanzate, nel periodo tra dicembre 2020 e aprile 2021 sono stati selezionati, esportati e anonimizzati 82 esami eseguiti tra il 25/03/2014 e il 20/4/2021, a seguito di una ricerca retrospettiva dei dati presenti nel PACS (Picture Archiving and Communication System).

Caratteristiche	Classe	Numero (percentuale)
Età	30 - 40	5 (6%)
	41 - 50	17 (21%)
	51 - 60	23 (28%)
	61 - 70	16 (20%)
	71 - 80	15 (18%)
	81 - 87	6 (7%)
Diagnosi	Benigne	43 (49%)
	Maligne	44 (51%)
Localizzazione	Unilaterale	79 (96%)
	Bilaterale	3 (4%)
Proiezioni	CC	87 (50%)
	MLO	87 (50%)

Tabella 2.1: Descrizione del Dataset

I singoli casi sono stati selezionati sulla base di alcuni criteri di inclusione:

- Pazienti che avessero eseguito un esame di clinica che comprendesse Tomosintesi e mammografia sintetica;
- Casistica riconducibile ad una sola tipologia, ossia masse e opacità ben visibili e osservabili sia in proiezione cranio caudale che in medio laterale obliqua. Sono state quindi escluse distorsioni e micro-calcificazioni;
- Diagnosi comprovata dall'istologico o dalla diagnosi della radiologia;

Il range di età delle pazienti va dai 30 agli 87 anni con una media di 60 anni. Tutti i casi comprendono almeno entrambe le proiezioni, CC e

MLO, del seno con anomalia o entrambe le proiezioni per entrambi i seni, ma ai fini del lavoro svolto sono state considerate soltanto le due proiezioni del seno che presentava la lesione. Il numero di lesioni è pari ad 87, di cui 43 lesioni maligne e 44 benigne, il che rende il Dataset bilanciato. Per ognuna di esse è stata fornita la classe e la localizzazione in termini di quadrante e lateralità. Tutte le caratteristiche sono schematizzate nella tabella 2.1. Le lesioni hanno un diametro che va da 0,4 a 16 cm, con una media di 2,2 cm.

Il lavoro svolto può essere diviso in due fasi distinte sulla base del numero di dati a disposizione. Per la prima fase del progetto, infatti, il Dataset a disposizione contava circa la metà dei casi. Con il successivo aumento della casistica sono stati eseguiti ulteriori test e alcune prove legate all'influenza della numerosità del Dataset sui risultati.

2.2 Estrazione delle ROI

Una volta importate le immagini, i file sono stati riorganizzati per essere facilmente consultabili. Le singole lesioni sono state segmentate manualmente da due senologhe esperte sulla base della diagnosi già fatta. Al fine di sfruttare l'alta risoluzione delle immagini e per facilitare la procedura di segmentazione, sono stati utilizzati i monitor diagnostici Hologic SecureView presenti in reparto.

I contorni sono stati tracciati con gli strumenti del software di visualizzazione già presente sui monitor, oppure con l'utilizzo di RadiAnt [28], un software di visualizzazione di immagini DICOM molto comune che permette, oltre alla visualizzazione dell'immagine diagnostica, anche l'annotazione a mano libera e l'esportazione dell'immagine in jpeg completa di annotazioni. I contorni tracciati su monitor diagnostico risultavano in toni di grigio e, perché fossero più visibili e più facilmente riproducibili, sono stati colorati di rosso (Figura 2.1).

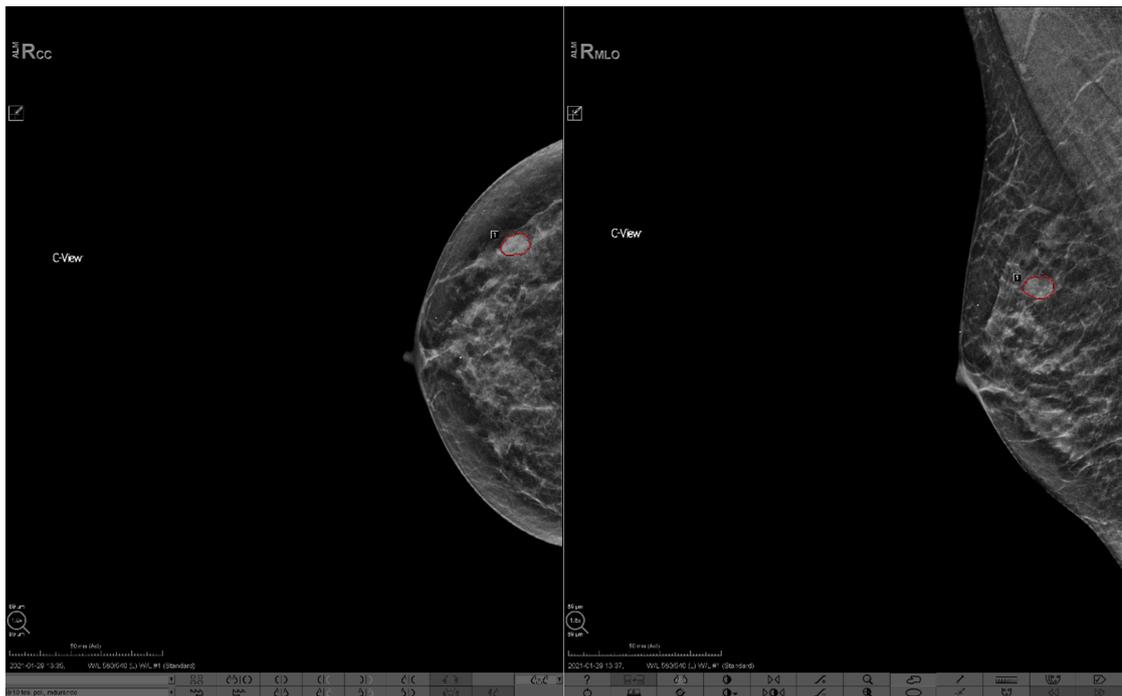


Figura 2.1: Esempio di segmentazione manuale su monitor diagnostico *SecureView* in entrambe le proiezioni.

Il criterio scelto per la segmentazione è stato seguire la forma dell'opacità, con l'obiettivo di includere solo la lesione. Una volta ottenuti, dove fattibile, sono stati esportati direttamente dalle immagini tramite individuazione dei pixel che ne costituivano la forma, oppure sono stati replicati manualmente tramite l'utilizzo del Toolbox di segmentazione di MATLAB. Infine, le ROI sono state individuate come Bounding Box dei contorni manuali. Il lavoro di segmentazione è stato svolto interamente sulle immagini 2D sintetiche originali. Per quanto riguarda l'individuazione e la segmentazione delle lesione sulle immagini di Tomosintesi, come prima cosa è stata individuata, per ogni immagine, la slice principale o "key-slice" sulla quale la lesione è nitida e ha diametro massimo. Successivamente, volendo esportare i contorni manuali già fatti sulla mammografia 2D sintetica, è stato verificato che la slice selezionata presentasse lo stesso orientamento e la stessa scala dell'immagine 2D sintetica per mezzo di un algoritmo di registrazione basato sull'intensità dei pixel. (Figura 2.2). Il risultato era prevedibile essendo

la C-View ottenuta dall'immagine di Tomosintesi, ma i risultati hanno dato ulteriore conferma della fattibilità dell'esportazione (Figura 2.3).

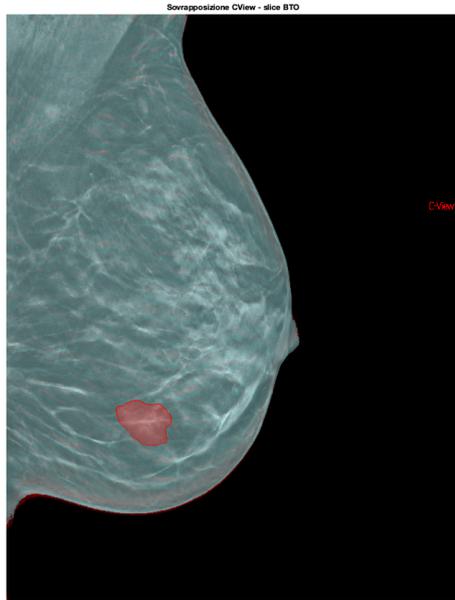


Figura 2.2: Sovrapposizione immagine 2D sintetica e Key-slice BTO

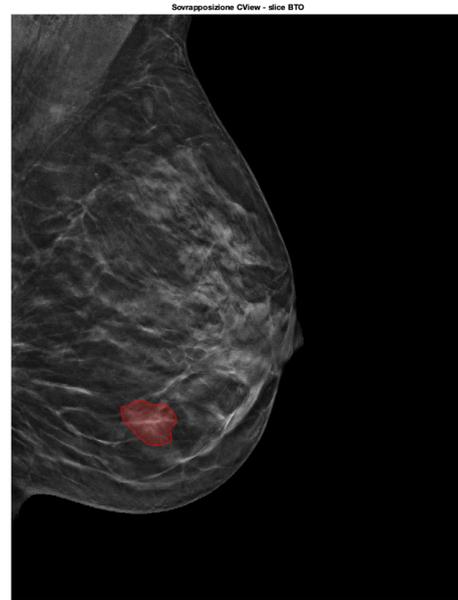


Figura 2.3: Esportazione del contorno della lesione sulla Key-slice BTO

2.2.1 Pre processing

In fase di pre-processing e features extraction, sono stati applicati diversi filtri. Il filtraggio di un'immagine consiste nell'eseguire una convoluzione bidimensionale tra l'immagine stessa e un kernel, ossia una matrice di dimensioni note inferiori a quelle dell'immagine i cui valori variano in base alla tipologia di filtro. I filtri possono essere distinti in:

- Filtri passa basso o filtri di smoothing: matematicamente eseguono un'operazione che equivale alla media di un determinato numero di punti dell'immagine. Un noto filtro di smoothing è il filtro Gaussiano la cui implementazione è assimilabile ad una media pesata dei

pixel ed è tipicamente utilizzato per rimuovere il rumore gaussiano sovrapposto all'immagine. È caratterizzato dalla dimensione e dalla deviazione standard.

- Filtri passa alto o derivativi: matematicamente eseguono una derivata discreta, sono direzionali e vengono utilizzati per accentuare le discontinuità di un immagine. Un filtro derivativo comune è il filtro di Sobel che, per il modo in cui viene implementato, evidenzia maggiormente le discontinuità dell'immagine. Il gradiente di Sobel ha la seguente implementazione:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Le due tipologie di immagini sono state pre-processate in modo diverso sulla base delle strategie di preprocessing più comuni in letteratura. Essendo la C-View già processata con un algoritmo proprietario, si è scelto di non apportare ulteriori modifiche e gli unici step eseguiti comprendono:

1. Conversione in formato double dell'immagine per avere tutti i valori compresi nel range che va da 0 a 1;
2. Rimozione dello sfondo: in numerosi casi all'interno del dataset, le lesioni si trovano in prossimità del bordo della mammella e, nella fase di caratterizzazione del tessuto esterno alla lesione, lo sfondo nero introduce un bias nel calcolo dei parametri quindi deve essere rimosso. Essendo nell'immagine mammografica completamente nero, è sufficiente eseguire un global thresholding con soglia a 0 (Figura 2.4);
3. Correzione della maschera: per preservare il tessuto, data la possibilità che comprenda alcuni pixel neri, la maschera è stata corretta rimuovendo eventuali buchi

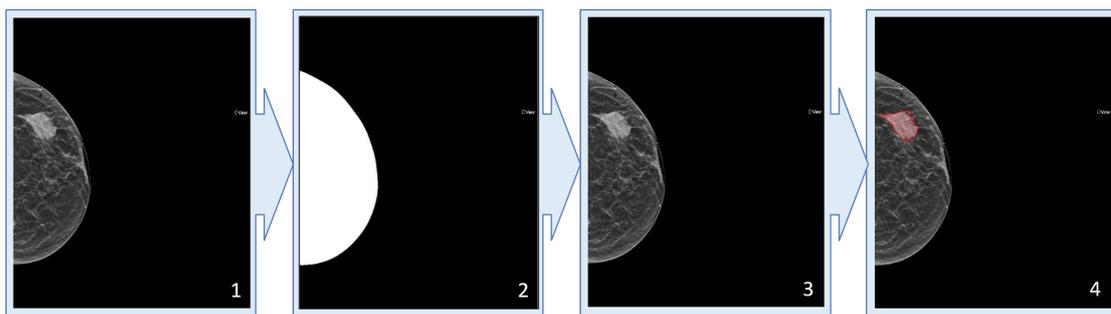


Figura 2.4: Step di pre-processing delle immagini di mammografia 2D sintetica.

Gli stessi step di pre-processing sono stati seguiti per le immagini BTO, ma in questo caso l'immagine è stata ulteriormente filtrata con un filtro gaussiano di dimensione 3 e deviazione standard pari a 0.8 e, in seguito, è stata sottratta all'originale per ottenere l'effetto di sharpening (Figura 2.5) [29].

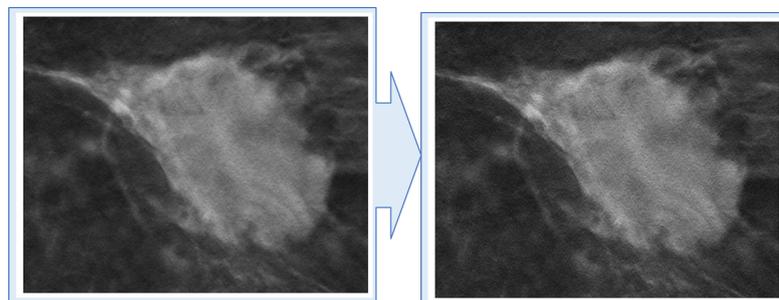


Figura 2.5: Effetto dello sharpening sull'immagine della lesione sulla slice BTO

Per entrambe le tipologie di immagini pre-processate come descritto, a partire dai contorni manuali, sono state calcolate le maschere di segmentazione che isolassero la lesione dal tessuto circostante. Per diminuire il peso computazionale, dall'immagine totale sono state ricavate delle ROI di dimensione pari alla Bounding Box dei contorni manuali maggiorata di un fattore del 10% (Figura 2.6).

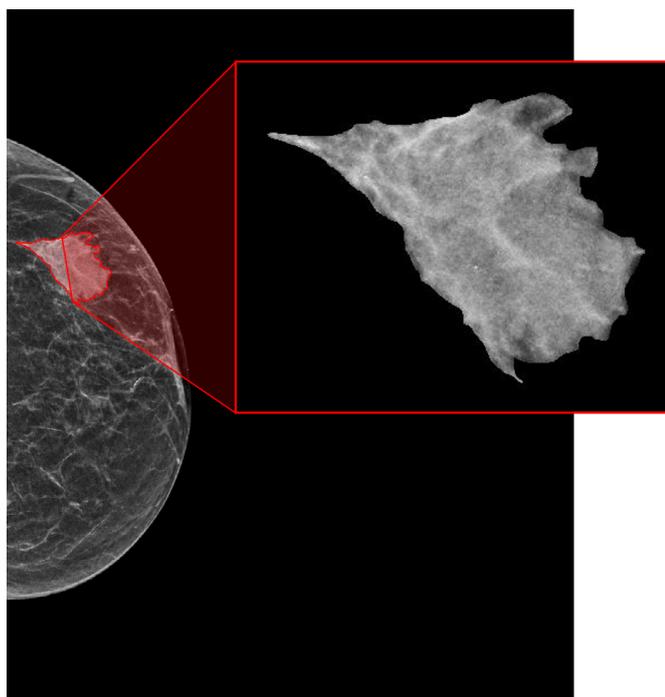


Figura 2.6: Schema di estrazione delle ROI dalle immagini C-View

2.3 Estrazione delle features

Una volta identificate e isolate le regioni di interesse, sono stati implementati gli operatori di tessitura e descrittori di forma riportati nei successivi paragrafi. L'obiettivo è riuscire a caratterizzare il tessuto estraendo informazioni che possano distinguere le due tipologie di lesioni al fine della classificazione. Dagli studi su immagini di mammografia e tomosintesi emerge l'importanza di caratterizzare non soltanto la regione identificabile come lesione ma anche la porzione di tessuto prossima ad essa [30, 31]. Per quanto riguarda il tessuto appartenente alla lesione, è stata utilizzata la maschera precedentemente ottenuta per isolarlo dal tessuto circostante in modo che non incidesse sul calcolo dei parametri. Per la caratterizzazione del tessuto esterno, invece, la stessa maschera è stata dilatata in modo che comprendesse una porzione maggiore proporzionalmente alla dimensione della lesione, implementando di fatto un padding proporzionale in quanto in grado di restituire risultati migliori

[32, 33]. Come prima cosa è stata calcolata la lunghezza dell'asse maggiore dell'area circoscritta al contorno manuale. Considerando il 10% di questo valore come raggio di un elemento strutturale a disco, tramite dilatazione è stata ottenuta la nuova maschera con forma simile e area maggiore.

Il valore da utilizzare per il padding proporzionale è stato scelto in funzione del vasto range di dimensione delle lesioni all'interno del Dataset. Il valore scelto permette da una parte di considerare una porzione di tessuto sufficiente per un'analisi significativa, dall'altra evita che la regione di tessuto esterno sia troppo ampia e che, soprattutto per quanto riguarda le opacità di diametro maggiore, si perda l'informazione dovuta alla prossimità al margine della lesione. Inoltre, nella maggior parte dei casi, si riesce a ottenere un'area completamente contenuta nel tessuto evitando di includere lo sfondo nero che non ha contribuito all'analisi.

La maschera del tessuto esterno si ottiene per operazione logica XOR tra maschera iniziale e dilatata (Figura 2.7). I parametri calcolati sul tessuto esterno sono stati rapportati a quelli calcolati nella regione interna, ottenendo un'informazione indipendente dall'intensità dei pixel della singola immagine ma significativa dal punto di vista della differenza tra opacità e tessuto circostante.

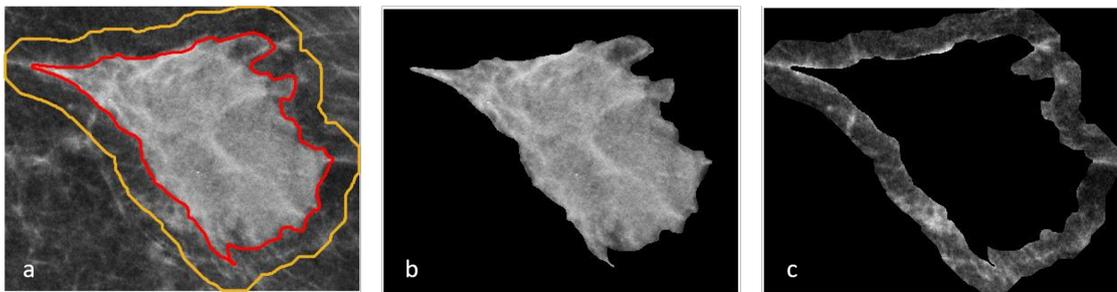


Figura 2.7: (a) Visualizzazione del contorno manuale, in rosso, e del contorno della maschera dilatata in seguito al padding proporzionale, in arancione. (b) Maschera del tessuto interno identificabile come lesione. (c) Maschera della porzione di tessuto esterno

I descrittori di forma, invece, sono parametri legati alla geometria e

calcolati sui contorni manuali originali.

2.3.1 Operatori di tessitura

La tessitura di un'immagine è la ripetizione di un pattern di variazioni locali dell'intensità dei pixel e di conseguenza può essere definita solo in relazione ad una regione e non ai singoli pixel [34], i quali hanno un valore di intensità assoluto. L'analisi di tessitura dell'immagine permette di estrarre informazioni relative sia al valore assoluto dei pixel appartenenti ad una regione di interesse che alla loro distribuzione nello spazio. Il livello di dettaglio dell'analisi dipende dalla dimensione degli elementi di tessitura, che viene scelta in base alla risoluzione e alla dimensione dell'immagine.

Tessitura del prim'ordine

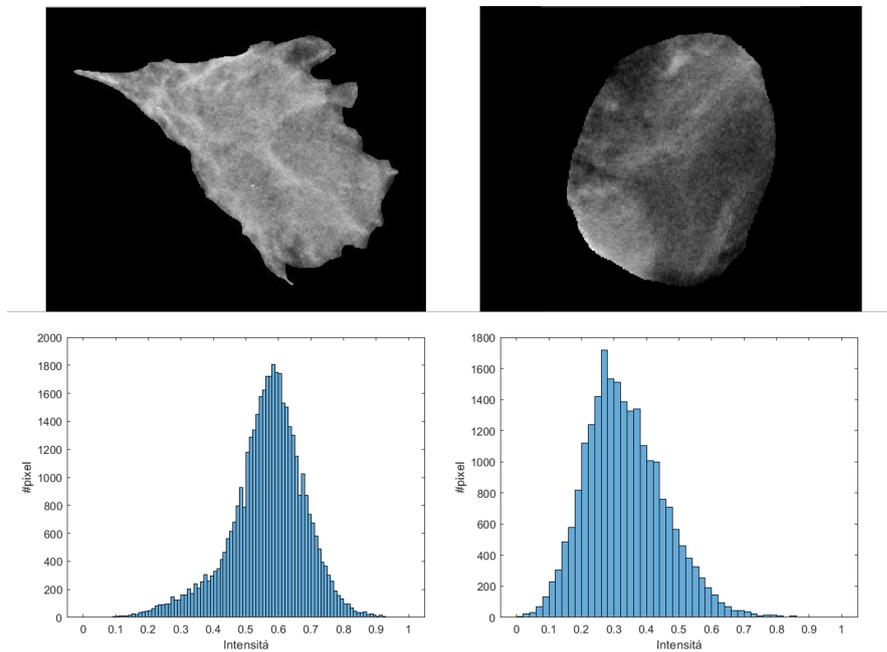


Figura 2.8: A sinistra l'immagine in grayscale, a destra l'istogramma delle luminosità che descrive la distribuzione dei valori di livelli di grigio.

Gli operatori di tessitura del prim'ordine sono operatori statistici basati sull'istogramma delle luminosità dell'immagine e descrivono quindi la distribuzione dei valori correlati ai singoli pixel. Non forniscono alcuna informazione a livello morfologico, ma danno una prima stima della luminosità e dell'omogeneità dell'immagine, che sono caratteristiche deducibili anche dall'osservazione dell'istogramma così come il contrasto (Figura 2.8). I parametri calcolati sono media, varianza, skewness, kurtosis ed entropia, e sono riportati in tabella 2.2 con le relative espressioni matematiche.

Features I ordine	Formula
Media	$\frac{1}{N} \sum_{i=1}^N A_i$
Varianza	$\frac{1}{N-1} \sum_{i=1}^N A_i - \nu^2$
Skewness	$\frac{E(x-\nu)^3}{\sigma^3}$
Kurtosis	$\frac{E(x-\nu)^4}{\sigma^4}$
Entropia	$-\sum hist * \log_2 hist$

Tabella 2.2: Statistica descrittiva del prim'ordine

Descrittori di Haralick e Galloway

A differenza della tessitura del prim'ordine, i descrittori di Haralick e Galloway forniscono una valutazione della dipendenza spaziale tra i valori associati ai singoli pixel. I descrittori di Haralick vengono calcolati sulla matrice delle co-occorrenze o GLCM: si tratta di una matrice quadrata e simmetrica il cui numero di righe i , pari al numero di colonne j , rappresenta l'intero range di possibili valori che i pixel dell'immagine possono assumere discretizzato su un numero finito di livelli. In ogni cella viene riportata la probabilità congiunta che due pixel a determinata distanza D abbiano livelli di grigio pari a $L(i)$ diverso da $L(j)$ (Figura 2.9). Al diminuire del numero di celle con un numero alto di occorrenze, aumenta l'omogeneità dell'immagine e viceversa. Allo stesso modo possono essere valutate l'entità delle

variazioni locali, la presenza di pattern e la complessità dell'immagine. L'intervallo di valori assunti dai pixel è stato discretizzato su 128 livelli e la distanza tra pixel contati come occorrenze è stata impostata pari a 1. Oltre alla $R(0,1)$, ossia la matrice con inclinazione pari a 0° rispetto all'immagine, sono state calcolate le matrici GLCM per tre ulteriori angoli di inclinazione pari a 45° , 90° e 135° (Figura 2.10). I parametri sono stati calcolati sulla matrice risultante dalla media delle quattro ottenute e sono riportati in tabella 2.3.

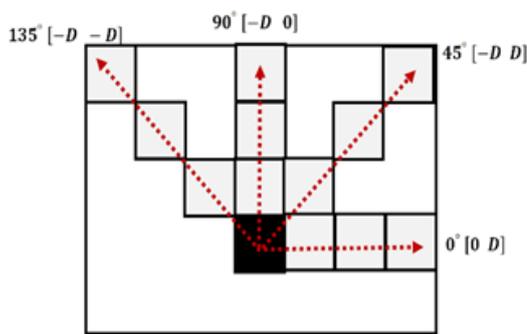


Figura 2.9: Schema di calcolo della matrice GLCM nelle quattro direzioni utilizzate e a distanza generica D

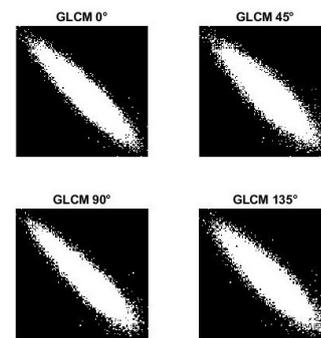


Figura 2.10: Rappresentazione delle matrici GLCM calcolate su una ROI di esempio

I descrittori di Galloway sono calcolati invece sulla matrice delle run o GLRLM: una run è un gruppo di pixel in cui si osserva consecutivamente la stessa intensità di grigio e la matrice che le descrive avrà tante righe quanti i valori che i pixel dell'immagine possono assumere e tante colonne quanto più lunga sarà la run di lunghezza massima. Ad un'immagine omogenea corrisponderà una matrice con valori nulli tranne in corrispondenza del valore di intensità di grigio dell'immagine, mentre i valori saranno più disomogenei in caso contrario. Anche in questo caso, per il calcolo della matrice, sono stati considerati un numero di livelli pari a 128 e distanza pari a 1. I parametri di Galloway descrivono la distribuzione dei valori nella matrice, caratterizzandola in base alla presenza di run più o meno lunghe e a quali valori di intensità

di grigio corrispondono. Per il calcolo della matrice delle run e dei parametri, è stato implementato un algoritmo custom in MATLAB sulla base dell'algoritmo Open Source Gray Level Run Length Image Statistics [35].

Features	Formula
Contrasto	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i - j ^2 * P(i, j)$
Correlazione	$\frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\nu_x)*(j-\nu_y)*P(i, j)}{\sigma_x \sigma_y}$
Energia	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)^2$
Omogeneità	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1+(i-j)^2} P(i, j)$
Simmetria	$1 - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i - j P(i, j)$
Entropia II	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) * \log_2 P(i, j)$
Max Probabilità	$MAX(P(i, j))$
Short Run Emphasis	$\frac{\sum_{i=i}^N \sum_{j=i}^M R(i, j) / j^2}{\sum_{i=i}^N \sum_{j=i}^M R(i, j)}$
Long Run Emphasis	$\frac{\sum_{i=i}^N \sum_{j=i}^M R(i, j) * j^2}{\sum_{i=i}^N \sum_{j=i}^M R(i, j)}$
Gray Level Non-uniformity	$\frac{\sum_{i=i}^N (\sum_{j=i}^M R(i, j))^2}{\sum_{i=i}^N \sum_{j=i}^M R(i, j)}$
Run Length Non-uniformity	$\frac{\sum_{i=i}^M (\sum_{j=i}^N R(i, j))^2}{\sum_{i=i}^N \sum_{j=i}^M R(i, j)}$
Run percentage	$\frac{\sum_{i=i}^N \sum_{j=i}^M R(i, j)}{N * M}$

Tabella 2.3: Statistica descrittiva del second'ordine

I descrittori di Galloway sono stati calcolati anche sull'immagine filtrata con un filtro di Sobel applicato separatamente in direzione X e Y per evidenziare le discontinuità orizzontali e verticali [36]. Le due immagini filtrate nelle due direzioni sono state sommate e sull'immagine risultante è stata calcolata la matrice GLRLM (Figura 2.11).

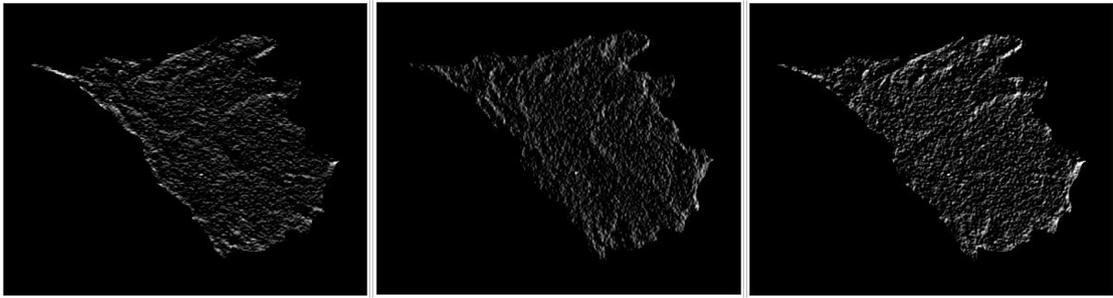


Figura 2.11: Risultato del filtraggio con filtro di Sobel lungo X e lungo Y. L'ultima immagine è la composizione delle due precedenti, ed è quella che è stata utilizzata per il calcolo dei parametri

Local Binary Patterns

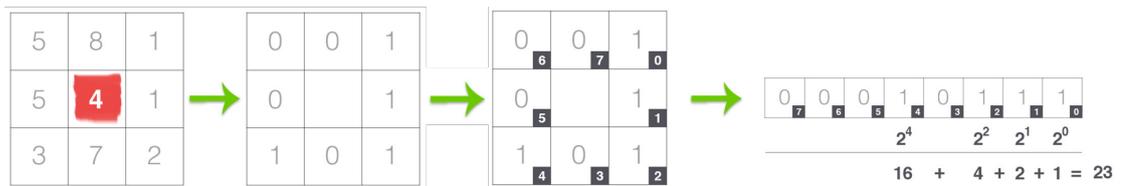


Figura 2.12: Schema del calcolo dei Local binary patterns per valori generici dei pixel [37].

I Local Binary Patterns (LBP) sono stati introdotti da Ojala et al. [38] e permettono di analizzare il contrasto locale dell'immagine. Sono stati utilizzati con successo per la classificazione automatica del livello di densità del seno [39] e in generale richiedono un ridotto costo computazionale, sono utili per l'analisi di tessitura e, per la loro implementazione, sono inoltre indipendenti dalla rotazione. L'immagine in esame viene suddivisa in un determinato numero di vicinati: per ogni pixel viene analizzato il vicinato assegnando 1 se il loro valore è inferiore a quello del pixel centrale, altrimenti 0. In un ordine specifico, gli 0 e 1 vengono letti in sequenza a formare un numero binario che può essere convertito in decimale [37]. I valori ottenuti vengono inseriti in un istogramma con un numero di colonne pari al massimo numero di combinazioni ottenibili con la dimensione del vicinato scelta (Figura

2.12). I LBP sono stati calcolati impostando un numero massimo di vicini pari a 8 e i valori ottenuti sono stati mediati per ottenere un unico parametro.

Laws Energy Matrix

Un'ulteriore metodo per estrarre informazioni sulla texture di un'immagine è filtrarla nel dominio dello spazio con kernel che attenuino separatamente alcune microstrutture della tessitura. Il metodo sviluppato da K. I. Laws prevede il filtraggio dell'immagine con filtri progettati e combinati al fine di ottenere diverse misure energetiche di tessitura: ogni filtro determina una risposta impulsiva che esalta componenti diverse della tessitura, tra cui discontinuità, punti isolati, increspature e ondulazioni [36]. I kernel dei filtri vengono implementati a partire dai seguenti vettori:

$$\begin{aligned} L5 &= [1 \ 4 \ 6 \ 4 \ 1] \\ E5 &= [-1 \ -2 \ 0 \ 2 \ 1] \\ S5 &= [-1 \ 0 \ 2 \ 0 \ -1] \\ R5 &= [1 \ -4 \ 6 \ -4 \ 1] \\ W5 &= [-1 \ 2 \ 0 \ -2 \ -1] \end{aligned}$$

I vettori vengono combinati tramite prodotto esterno e, in questo modo, si ottengono 25 kernel diversi ma tutti di dimensione 5 x 5 che uniscono gli effetti dei filtri derivativi e di smoothing. Al fine di ridurre la dimensione della matrice di parametri ottenuta, è stata eseguita una combinazione dei filtri direzionali in modo da poter estrarre features che non dipendessero dalla direzione del filtraggio. In questo modo, in totale, sono state ottenute 14 immagini diverse (Figura 2.13). Per ognuna delle immagini è stata calcolata la TEM ossia la matrice di texture-energy e sono stati estratti ulteriori quattro parametri che comprendono media, varianza, skewness e kurtosis.

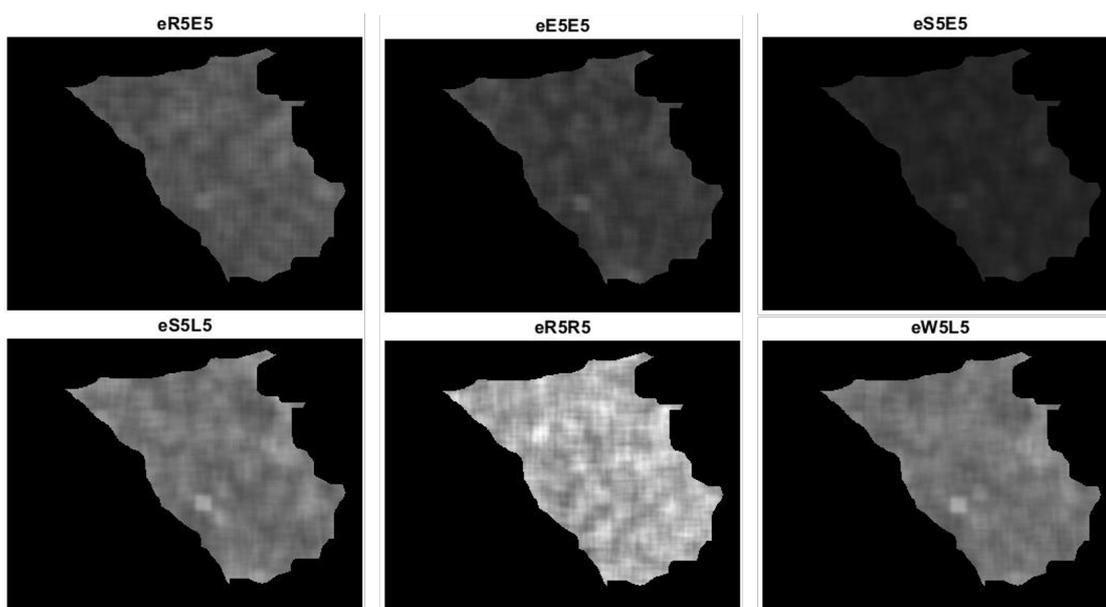


Figura 2.13: Esempi di immagini ottenute in seguito al filtraggio con i filtri di Laws, dalle quali sono stati estratti i parametri descritti. Per ogni immagine è riportato il nome del filtro utilizzato come combinazione dei vettori originari.

Oltre a rappresentare un metodo di analisi di tessitura più complesso, ha dimostrato dare buoni risultati nella diagnosi automatica di lesioni su immagini di mammografia [40].

2.3.2 Metodi Transform-based

I metodi basati sulle trasformate sono tecniche di analisi di tessitura che prevedono di lavorare in uno spazio diverso dove alcune caratteristiche di tessitura sono più visibili e significative. Un approccio comune è il passaggio allo spazio delle frequenze.

Filtro di Gabor

Il filtro di Gabor, che è strettamente correlato alla wavelet di Gabor, è un filtro lineare bidimensionale con risposta all'impulso pari ad una sinusoidale modulata con una gaussiana. L'espressione della funzione è

la seguente:

$$g(x, y, \phi, \theta) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(i(2\pi x'/\lambda + \phi))$$

La definizione del filtro dipende da x e y , ossia le coordinate dei punti dell'immagine, dall'orientazione θ e dall'ampiezza della sinusoidale, o scala, ϕ , i quali sono parametri settabili in base al tipo di applicazione (Figura 2.14). La localizzazione spaziale è determinata dalla funzione gaussiana, mentre con la componente sinusoidale si analizzano le componenti del segnale in frequenza.

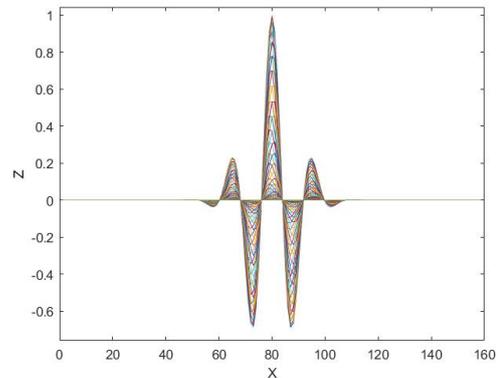
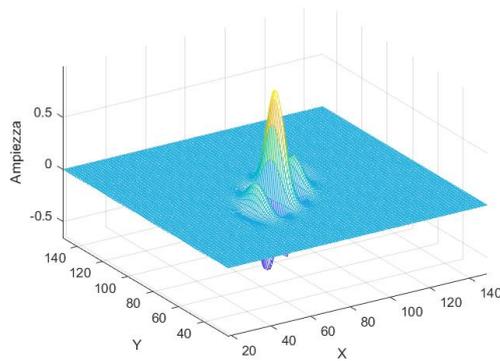


Figura 2.14: Rappresentazione del filtro di Gabor implementato con orientazione apri a 45° e lunghezza d'onda pari a 10. Sono visibili sia la componente gaussiana che la sinusoidale

Figura 2.15: Rappresentazione dello stesso filtro ma in 2D: l'onda ha forma simile alle wavelet, con ampiezza attenuata dalla funzione di Gauss.

Il filtraggio avviene per convoluzione bidimensionale tra l'immagine in grayscale in input e il kernel complesso e il risultato, che appartiene allo spazio di Gabor, viene calcolato in termini di modulo e fase. Viene utilizzato largamente per l'analisi di tessitura in quanto è in grado di individuare lo specifico contenuto in frequenza dell'immagine nella regione di interesse e può essere utilizzato per la detection di determinate

strutture. Comunemente si applicano diversi filtri con orientamento e lunghezza d'onda variabili.

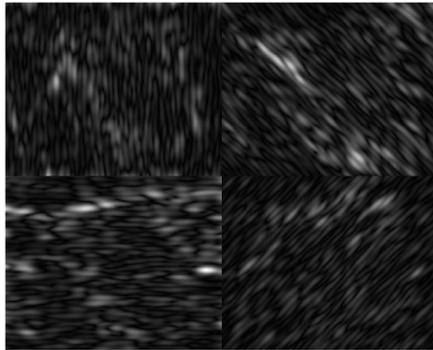


Figura 2.16: Risultato del filtraggio dell'immagine lungo le quattro direzioni scelte e lunghezza d'onda pari a 10.

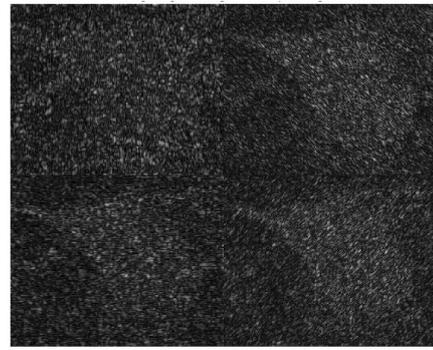


Figura 2.17: Risultato del filtraggio nelle stesse direzioni ma con lunghezza d'onda pari a 5

Come emerge dalla letteratura [41, 42], il filtro di Gabor può essere utilizzato per l'individuazione delle spicule che sono considerate indici di malignità se visibili nell'immagine mammografica e sono molto comuni, in particolare, nelle forme infiltranti. In mammografia, si manifestano come strutture allungate ad alta intensità e sono individuabili facilmente applicando il filtro di Gabor in diverse direzioni. Si è scelto quindi di generare una banca di quattro filtri per filtrare l'immagine in quattro direzioni e calcolare, nella stessa direzione di applicazione del filtro, la matrice GLCM. Gli angoli di inclinazione utilizzati sono 0° , 45° , 90° e 135° , mentre la lunghezza d'onda scelta è pari a 10 (Figura 2.16).

Il valore della lunghezza d'onda viene normalmente scelto nell'intervallo compreso tra 2 e un quinto della dimensione dell'immagine di input. In questo caso è stato scelto anche sulla base dell'osservazione dei parametri e delle immagini ottenuto in seguito al filtraggio: al diminuire della lunghezza d'onda, il modulo della risposta al filtro risulta fortemente influenzato anche dalle minime variazioni di intensità tipiche del tessuto e, in parte, dalla componente di rumore presente nell'immagine (Figura 2.16); al contrario, all'aumentare del valore di lunghezza d'onda,

la capacità di discriminazione dei parametri calcolati diminuisce in modo significativo soprattutto nel caso di lesioni con dimensioni ridotte. Il valore scelto rappresenta un buon compromesso (Figura 2.17). Essendo l'orientamento della lesione e delle spicule un fattore non determinante nella caratterizzazione delle lesioni, i parametri di Haralick sono stati calcolati sulle quattro matrici e successivamente mediati. Allo stesso modo si è deciso di procedere per i Local Binary Patterns, che sono quindi stati calcolati sulle quattro direzioni e mediati.

Trasformata discreta di Wavelet

La trasformata di Wavelet permette di estrarre informazione da un segnale in frequenza, localizzando nel tempo le singole frequenze. Si basa sulla convoluzione del segnale in esame con una funzione, definita wavelet, la quale è caratterizzata da una frequenza, o scala, e dalla localizzazione ossia l'istante di tempo in cui è centrata [43]. Allo stesso modo, la trasformata di Wavelet può essere applicata ad un'immagine e, in questo caso, si parla di trasformata discreta, la cui espressione è $T = \int_{-\inf}^{\inf} x(t) d_{(m,n)}(t) dt$. Essa utilizza un numero finito di wavelet che assumono valori di scala e localizzazione in un range definito e discreto. Per convoluzione bidimensionale con l'immagine si ottiene un'immagine sotto campionata ricostruita per mezzo di coefficienti di dettaglio e di approssimazione. Questa tecnica è utilizzata per la detection e per l'estrazione di features nell'analisi di tessitura. In questo caso, in particolare, l'informazione può essere utile alla localizzazione di microcalcificazioni, tipiche delle lesioni maligne. L'immagine è stata quindi processata tramite trasformata discreta bidimensionale, utilizzando una funzione biortogonale discreta. Come suggerito in letteratura [44], i coefficienti di dettaglio ottenuti nelle tre direzioni (x, y e xy) sono stati mediati al fine della riduzione della dimensionalità della matrice e ne è stata calcolata inoltre l'energia in direzione Y e in diagonale.

2.3.3 Descrittori di forma

Un elemento fondamentale per la caratterizzazione delle lesioni è la descrizione morfologica. Se la dimensione della lesione non è considerata indicativa di malignità ma più spesso correlata alla precocità della diagnosi, alcune caratteristiche del contorno sono spesso prese in considerazione come indici di benignità e malignità. A livello clinico, un'opacità che presenti una forma irregolare e un contorno non netto, interamente o in parte, suggerisce la tendenza della lesione all'infiltrazione nel tessuto adiacente e quindi la probabile natura maligna. Al contrario, una lesione benigna tende a crescere ed espandersi comprimendo e spostando i tessuti adiacenti e, per questo motivo, presenta comunemente un contorno più netto. In molti casi, ovviamente, questa discriminazione non è attuabile, in altri è semplicemente complicata da stabilire ad occhio nudo. Al fine di quantificare il livello di irregolarità dei contorni manuali, sono stati calcolati una serie di parametri che ne descrivessero le caratteristiche citate. Per calcolare i descrittori di forma è stata analizzata la maschera della lesione. Come caratteristiche fondamentali della ROI sono stati calcolati l'area e il perimetro in pixel, in quanto questi valori sono alla base del calcolo di quelli più complessi. Oltre ad area e perimetro sono stati implementati alcuni parametri geometrici fondamentali:

- Perimetro e area dell'ellisse equivalente, ossia l'ellisse che presenta lo stesso valore di secondo momento angolare della regione in esame (Figura 2.20).
- Lunghezza dell'asse minore e dell'asse maggiore in pixel: il valore si riferisce alla lunghezza degli assi dell'ellisse equivalente. La seconda misura è già stata utilizzata per il padding proporzionale.
- Involuppo convesso: è il poligono convesso con area minore che può contenere la regione di interesse (Figura 2.20).

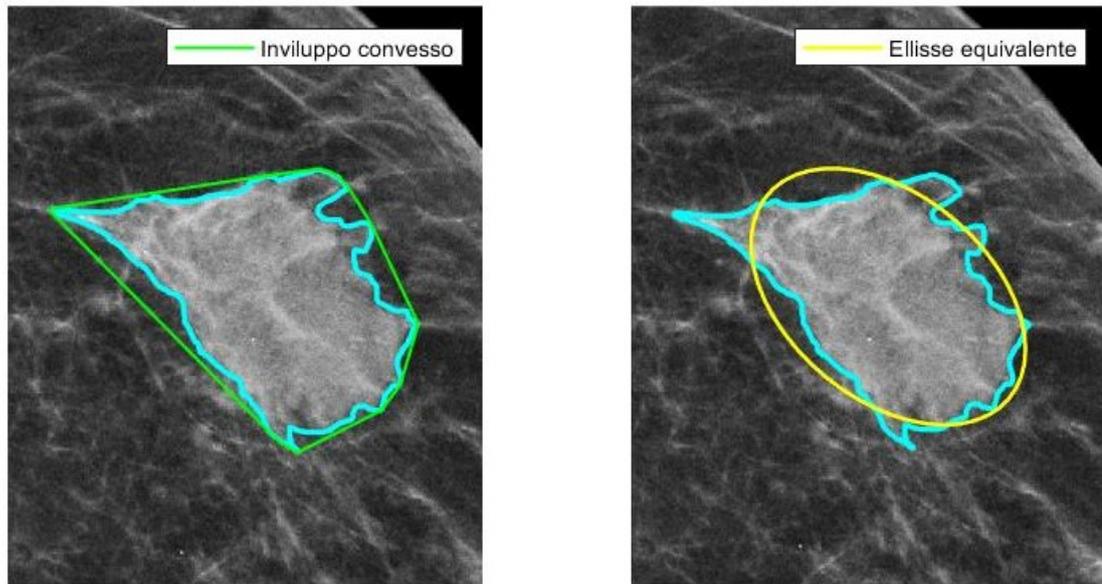


Figura 2.18: Rappresentazione dell'ellisse equivalente, a destra, e dell'inviluppo complesso, a sinistra, per una lesione del Dataset. Per confronto è stato riportato sempre il contorno manuale in ciano.

Sulla base delle misure di area e perimetro della maschera della lesione, dell'ellisse e dell'inviluppo convesso, sono stati calcolati ulteriori dieci parametri, alcuni dei quali basati sulla letteratura [45]:

- Eccentricità : la misura di eccentricità viene calcolata sull'ellisse equivalente e può avere un valore compreso tra 0 e 1. È il rapporto tra la distanza dei fuochi dell'ellisse e la lunghezza dell'asse maggiore e fornisce informazioni sulla forma della lesione;
- Circolarità : si calcola come

$$C = \frac{4\pi A}{2p^2}$$

e misura il grado di approssimazione con cui la regione può essere paragonata ad una circonferenza;

- Solidity: è il rapporto tra l'area della regione di interesse e l'area dell'inviluppo convesso. Di fatto è un'intersezione tra l'insieme dei

pixel contenuti nella regione e quelli contenuti nel poligono quindi fornisce informazione su quanto la forma della lesione si discosta da quella del poligono convesso che la racchiude;

- **Compattezza:** sono state calcolate due misure di compattezza, utilizzando le seguenti formule:

$$C1 = \frac{2\sqrt{A\pi}}{2p}$$

$$C2 = \frac{A}{A_{ellisse}}$$

La prima è legata alle sole caratteristiche del contorno manuale, la seconda è simile alla misura di solidity ma rapporta l'area della regione a quella dell'ellisse equivalente.

- **Fattore di forma:** è un descrittore di forma della lesione ma non dipende dalle dimensioni, in quanto area e perimetro vengono rapportati:

$$F = \frac{4\pi A}{2p^2}$$

- **Convessità:** è il rapporto tra il perimetro dell'involucro convesso e il contorno manuale. In questo caso si ha un'indicazione del livello di complessità del margine della lesione, che se si discosta di molto dal poligono convesso, determina un alto valore di convessità:

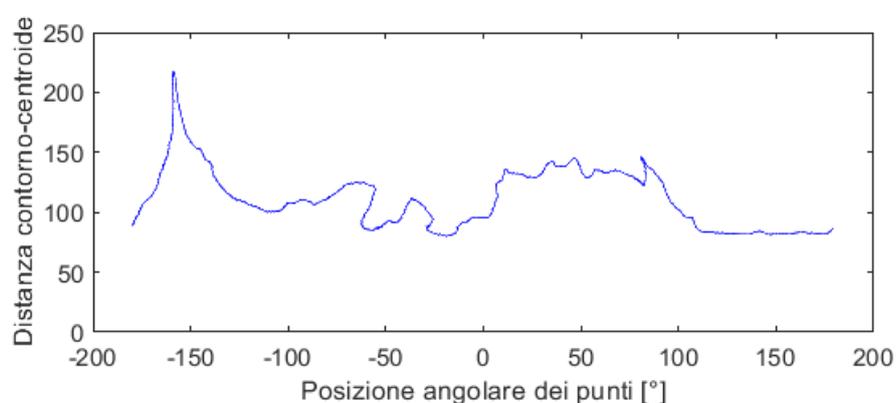
$$Cv = \frac{2p(Hull)}{2p}$$

- **Elongazione :** calcolata come

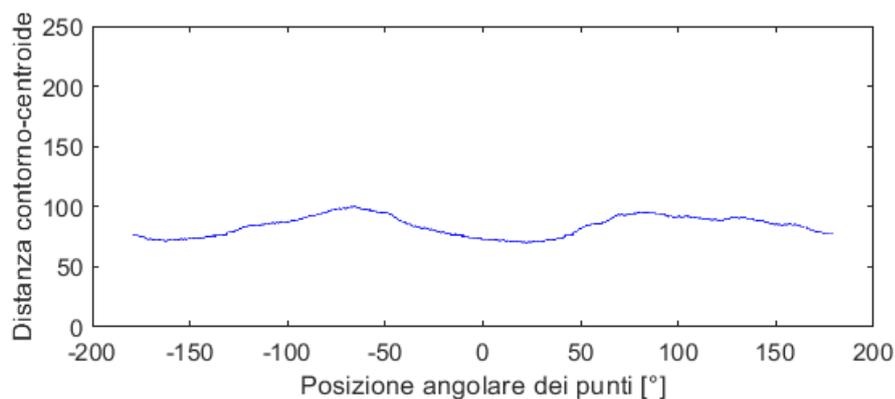
$$E = \frac{A}{2r_{max}^2}$$

Da informazione sulla forma della lesione, discriminando tra forme allungate o più compatte

- L2SRatio : è il rapporto tra asse minore e asse maggiore dell'ellisse equivalente e descrive quindi la forma generica della lesione
- Circonferenza ellittica normalizzata: calcolata come rapporto tra il perimetro della regione di interesse e quello dell'ellisse equivalente. E' un ulteriore indicatore di irregolarità del margine della lesione ed esattamente come per la convessità, aumenta se esso si discosta molto dall'ellisse equivalente (Figura 2.19).



(a) Profilo di una lesione maligna



(b) Profilo di una lesione benigna

Figura 2.19: I grafici riportano due profili di lesioni a confronto, tracciati in base alla distanza dal centroide della regione e alla posizione angolare dei punti del contorno.

2.4 Feature selection

In totale, il numero di features estratte è pari a 219 per ogni immagine. La scelta di calcolare un numero di parametri relativamente alto è legata alla volontà di riuscire a caratterizzare il tessuto nel modo più esaustivo possibile e per ottenere risultati migliori. Al fine di valutare quali parametri risultano effettivamente significativi al fine della classificazione, è stata implementata la features selection per mezzo del test statistico F-Test basato sull'analisi multivariata delle varianze o ANOVA [46]. Come prima cosa, i parametri calcolati sono stati standardizzati utilizzando lo Z-score, ossia sottraendo a tutti i valori la media della distribuzione e dividendo per la deviazione standard. La standardizzazione non cambia né la forma della distribuzione né l'ordine dei valori, ma facilita il confronto diretto tra valori misurati su scale diverse oppure soggetti ad arbitrarietà ed è propedeutica all'analisi delle varianze, per la quale si assume che le distribuzioni dei valori siano standardizzate. L'analisi delle varianze si basa sul due ipotesi contrastanti:

- H_0 -> le varianze sono uguali
- H_1 -> le varianze sono diverse

Nell'ANOVA, in particolare, le ipotesi considerate sono leggermente diverse e si assume che:

- H_0 -> le medie di tutti i gruppi sono uguali
- H_1 -> almeno una delle medie è diversa dalle altre.

Le due ipotesi si riferiscono alla media della distribuzione dei valori per le singole classi considerate per il problema di classificazione, in questo caso lesioni benigne e maligne. Con l'F-Test si calcola esattamente questa differenza tra gruppi: se non c'è differenza significativa, e di conseguenza il parametro non è significativo al fine della distinzione tra le due classi, il valore di F sarà molto vicino a 1. Come prima cosa si calcola la somma dei quadrati, ossia la deviazione di ogni elemento dalla media. A questo punto si calcolano: la somma totale dei quadrati,

ossia il quadrato delle differenze tra le singole osservazioni e la media della distribuzione, la media dei quadrati tra gruppi, ossia la somma al quadrato delle distanze delle singole medie dalla media di tutto e la media dei quadrati all'interno dei gruppi, ossia la somma al quadrato delle differenze tra le singole osservazioni e la media della distribuzione dei valori calcolati per gli elementi appartenenti alla singola classe. I gradi di libertà si calcolano invece in base al numero di classi e al numero di variabili per gruppo. Il valore di F, infine si calcola come:

$$F = \frac{\text{varianza tra gruppi (Between)}}{\text{varianza singolo gruppo (Within)}}$$

I parametri sono stati ordinati in base ai valori di F ottenuti, in modo che risultassero disposti in ordine decrescente di significatività. A questo punto è stato selezionato un sottogruppo di features da utilizzare in fase di classificazione degli elementi del Dataset. I primi test sono stati eseguiti sui parametri estratti dalle ROI delle immagini C-View e, di seguito, si riporta il risultato della feature selection in questo caso. Il numero di features da prendere in considerazione è stato scelto in base a due motivazioni principali:

1. il valore critico per l'F-Test indica la soglia minima dei possibili valori assumibili al di sotto della quale si accetta l'ipotesi nulla e, di conseguenza, le medie delle distribuzioni saranno uguali e non avrà significatività statistica. Questo valore dipende dal numero di classi e dal numero di elementi per ogni classe e, in questo caso, risulta essere pari a 3.87. Per avere un gruppo di features significative, il valore associato di F deve essere maggiore del valore critico e, con il numero selezionato, rappresentato dalla linea verticale tratteggiata, si riesce a rimanere largamente sopra a questo valore, che corrisponde all'altezza della retta tratteggiata nera;
2. In seguito ad un'analisi delle prestazioni al variare del numero di features considerate in un range compreso tra 5 e 30, si è notato come le prestazioni tendessero a migliorassero all'aumentare della dimensione del gruppo di features, ma con progressivo abbassamento delle prestazioni in fase di validazione;

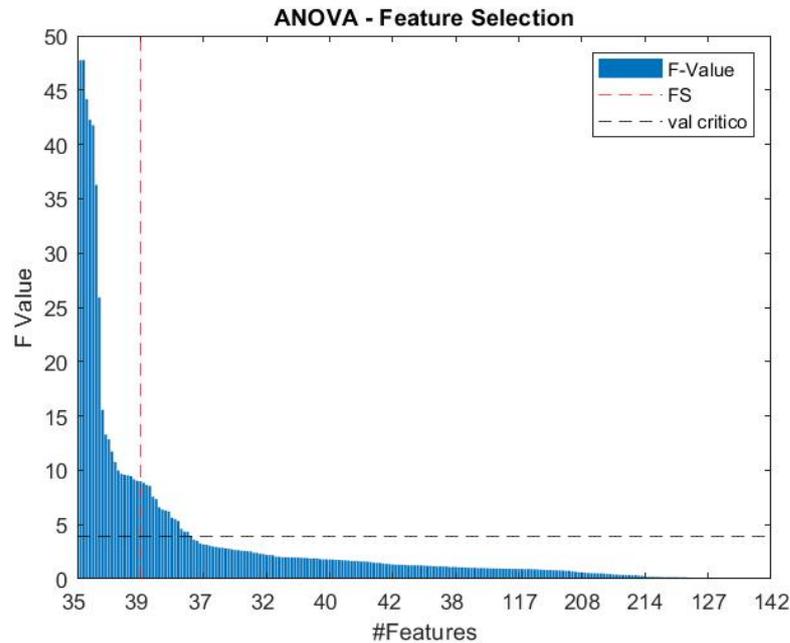


Figura 2.20: Features in ordine decrescente in base al valore di F. Sono riportati inoltre il valore critico come soglia orizzontale e il numero di features selezionato come soglia verticale

Il numero di features selezionato è pari a 20.

2.5 Metodi di classificazione

L'obiettivo principale dello studio è riuscire a classificare correttamente lesioni benigne e maligne sulla base delle caratteristiche che esse presentano nelle immagini di mammografia 2D sintetica e tomosintesi.

Come già detto precedentemente, è importante riuscire a sviluppare un metodo che assicuri un basso livello di falsi positivi. La promettente prospettiva di utilizzo delle mammografie 2D sintetiche in sostituzione alle tradizionali FFDM e la numerosità del Dataset hanno influito sulla scelta dei metodi. Per quanto riguarda il primo fattore, e come già descritto, è stata già provata l'accuratezza della mammografia sintetica in ambito diagnostico e, sebbene ad oggi non venga mai utilizzata

singolarmente per diagnosticare il carcinoma mammario ma sempre come supporto all'analisi della tomosintesi, viene comunque processata in modo da mantenere le stesse caratteristiche della mammografia e migliorandone alcuni limiti e artefatti. Per questo motivo sono stati utilizzati metodi di classificazione implementati in letteratura per la classificazione di immagini di mammografia, anche per valutare l'entità della variazione che si ottiene nelle prestazioni utilizzando immagini diverse. La numerosità del Dataset ha inciso molto sulle scelte fatte e i metodi di classificazione, basati sul machine learning e sul deep learning, sono stati sviluppati in relazione alla casistica a disposizione.

Per ognuno dei metodi proposti, sia per i classificatori che per la rete convoluzionale, verrà fatta una breve introduzione teorica sul funzionamento, al fine di poter descrivere i criteri di scelta dei parametri e le strategie adottate.

2.5.1 Metodi basati sul Machine Learning

Gli algoritmi di machine learning sono strumenti basati su metodi statistici, in grado di identificare in modo automatico pattern significativi presenti in un set di dati. Vengono utilizzati in un vasto range di applicazioni, tra cui lo sviluppo di modelli predittivi, l'elaborazione di immagini e la classificazione di dati in due o più classi. L'implementazione prevede una fase di allenamento, quindi di apprendimento, e una fase di testing dove la conoscenza acquisita viene applicata a nuovi dati. A seconda delle caratteristiche del training set, ossia dell'insieme di dati utilizzato in fase di allenamento, l'apprendimento automatico può essere supervisionato o non supervisionato. Gli algoritmi di machine learning supervisionato sono utilizzati sia per task di regressione che di classificazione: durante la fase di training, viene dato come input al sistema un insieme di elementi strutturati per i quali l'utente ha specificato la classe di appartenenza e dai quali l'algoritmo estrae la conoscenza che poi utilizza nella fase di test [47]. Al contrario, l'apprendimento non supervisionato avviene su dati non strutturati e la suddivisione in classi avviene senza il controllo dell'utente. L'ottimizzazione mira all'equilibrio tra:

- **Overfitting:** il modello si adatta completamente ai dati utilizzati nella fase di training riuscendo a classificarli correttamente ma perdendo totalmente la capacità di generalizzare e quindi utilizzare la conoscenza estratta dai dati di training per classificare correttamente quelli di test
- **Underfitting:** il modello non è in grado di estrarre conoscenza dai dati di training e, di conseguenza, non è in grado di generalizzare

Una volta estratte le features dalle ROI e in seguito al test statistico per decretare il gruppo di features più significative al fine della distinzione tra le due classi, sono stati implementati cinque metodi di classificazione basati sul machine learning, che verranno descritti di seguito, e una CNN custom descritta nella sezione successiva.

Linear Discriminant Analysis

L'LDA è un classificatore lineare che si basa sull'analisi discriminante lineare, un metodo statistico a sua volta strettamente correlato all'analisi delle varianze o ANOVA. Risulta più robusto e capace di generalizzare anche se allenato con un training set ridotto [48] e, di conseguenza, si rivela una scelta ottimale per questa task di classificazione. Si assume che gli elementi appartenenti alle singole classi siano disposti secondo una distribuzione gaussiana, della quale vengono calcolate la media μ e la covarianza σ [49]. Se non si conosce la probabilità di appartenenza a priori ad una delle classi j , si assume che ogni elemento abbia la stessa probabilità di appartenenza ad ognuna di esse. La funzione utilizzata per il calcolo di questa probabilità è denominata funzione discriminante ed è implementata come:

$$d_i(x) = \log(f_i(x)) + \log p_i$$

Tale funzione contiene il termine non lineare

$$\log(f_i(x))$$

il quale, con l'introduzione dell'analisi discriminante lineare che permette di assumere che il valore di covarianza tra tutte le classi coinvolte nel

processo decisionale sia lo stesso, può essere sviluppato, ottenendo [50]:

$$d_k(x) = x^T * (\mu_k) - 1 - 1/2 * \mu_k^T * (\mu_k) - 1 + \log p_k$$

In fase di classificazione, viene calcolato il valore della funzione per ogni classe e ogni elemento sarà assegnato a quella per la quale presenta la maggiore probabilità di appartenenza. La funzione discriminante rappresenta inoltre il margine di separazione (Figura 2.21) tra due classi nel piano, nonché il luogo dei punti x che hanno la stessa probabilità di appartenere a una delle due.

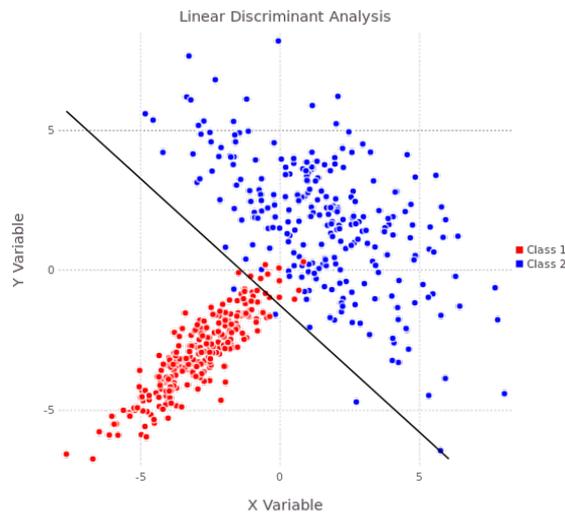


Figura 2.21: Esempio di divisione degli elementi di due classi con l’analisi di discriminazione lineare [50]

In fase di progettazione del classificatore LDA è stato scelto di implementare un modello lineare che usasse le matrici di covarianza calcolate per le singole classi secondo la seguente espressione:

$$\Sigma_{\gamma} = (1 - \gamma)\Sigma + \gamma * \text{diag}(\Sigma)$$

Il parametro gamma regola la regolarizzazione della matrice e, in seguito all’analisi delle prestazioni, è stato impostato pari a 0.

Support Vector Machine

L'SVM è un algoritmo supervisionato utilizzato in diversi studi basati su immagini di mammografia e tomosintesi, sia sulla base di features estratte manualmente dalle immagini che da reti convoluzionali. Inoltre, ha la capacità di sfruttare relazioni non lineari per classificare elementi che non sono linearmente separabili (Figura 2.22). Per questi motivi è stato scelto come metodo di classificazione per questo studio. L'algoritmo è volto alla determinazione dell'equazione dell'iperpiano che meglio riesce a separare gli elementi appartenenti alle due classi in esame, la quale sarà del tipo:

$$w^T * x + b = 0$$

Per farlo, viene risolto un problema di ottimizzazione con l'obiettivo di massimizzare la distanza tra i support vector, ossia gli elementi delle due classi che, nello spazio, si trovano più vicini tra loro e al piano di separazione. Questa distanza è anche denominata margine e il numero di support vector da può essere definito in fase di progettazione poichè incide sulla flessibilità del modello e sulle prestazioni [51]. Oltre che dalla distanza dai support vector, l'ottimizzazione dipende:

- dal kernel, ossia dal tipo di trasformazione che si vuole utilizzare per la separazione delle classi
- dal costo di misclassificazione C , ossia un termine di penalizzazione scelto in base all'importanza di avere un modello più o meno accurato e con margini più o meno ampi.

L'equazione da massimizzare è la seguente:

$$M(w) = \frac{1}{2} * w^T * w + C * \sum \varepsilon_i$$

L'SVM implementato presenta un Kernel lineare e un costo di misclassificazione progettato per mantenere un basso livello di falsi positivi. Al fine di limitare il fenomeno dell'overfitting, è stato imposto un limite sul numero di support vectors considerati tramite il parametro Box Constraint.

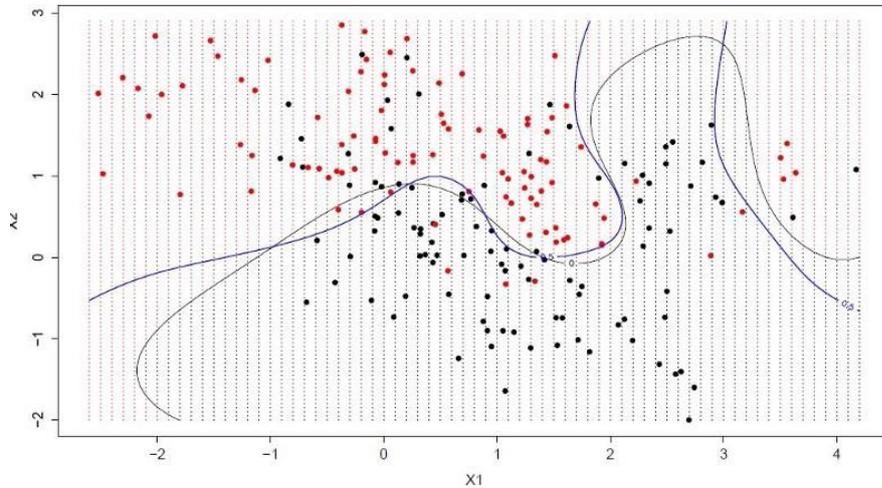


Figura 2.22: Esempio di modello SVM per la classificazione binaria [47].

K-Nearest Neighbours

Il KNN è un classificatore più semplice in quanto utilizza informazione locale, basandosi principalmente sul calcolo della distanza, secondo una specifica metrica, tra gli elementi all'interno del Dataset. L'obiettivo è determinare la posizione dei k punti con distanza minore dall'elemento in esame, dove k è il parametro fondamentale da definire in fase di progettazione. In particolare, k viene scelto in rapporto al numero di elementi ed è buona norma scegliere un valore dispari in quanto riduce la probabilità di assegnazione di un elemento a due classi distinte. Al variare di k cambiano le prestazioni dell'algoritmo, in particolare un k più alto previene l'overfitting.

La ricerca dei k punti più vicini avviene secondo il metodo di ricerca esaustiva: vengono calcolate le distanze tra il punto in esame x e tutti gli altri punti ottenendo una serie di valori che vengono messi in ordine crescente al fine di facilitare la ricerca delle distanze minime. In questo caso, la metrica scelta per il calcolo della distanza è la distanza euclidea standardizzata, che ha la seguente espressione:

$$d_{mn}^2 = (x_n - y_m) * V^{-1} * (x_n - y_m)$$

Dove V è una matrice di coefficienti utilizzati per la standardizzazione. Durante la fase di allenamento, vengono generati k gruppi di elementi, ad ognuno dei quali viene assegnata una delle classi di appartenenza sulla base del label più frequente. La classe dei nuovi elementi verrà assegnata sulla base di quale gruppo risulterà avere distanza minima dal nuovo elemento. È stato implementato un KNN con k pari a 7. Come già citato la metrica di distanza scelta è la standard euclidean ed è stato utilizzato un costo di misclassificazione che sfavorisse i falsi positivi.

Artificial Neural Network

Le Artificial Neural Network o ANN sono algoritmi supervisionati il cui funzionamento è modellato sul metodo di elaborazioni di informazioni tipico della mente umana [47]. Si tratta di un'insieme di unità interconnesse e strutturate in grado, dato un certo numero di input, di dare in output una predizione che, nel caso di un problema di classificazione, è la classe di appartenenza dell'elemento. L'unità fondamentale dell'ANN è il neurone che estrae particolari caratteristiche dai dati in ingresso e le elabora. È caratterizzata da una funzione di attivazione che, spesso, è una funzione sigmoidale del tipo:

$$\frac{e^t}{1 + e^t}$$

Ogni rete neurale è una struttura costituita da (Figura 2.23):

- un determinato numero di neuroni di input pari al numero di variabili in esame
- un output layer
- un numero variabile di hidden layers che, a loro volta, possono includere un numero variabile di neuroni.

La progettazione della rete comprende la definizione di questa struttura in termini di numero di layers e numero di neuroni di cui ogni layer è composto. I neuroni elaborano i valori in input x secondo la seguente

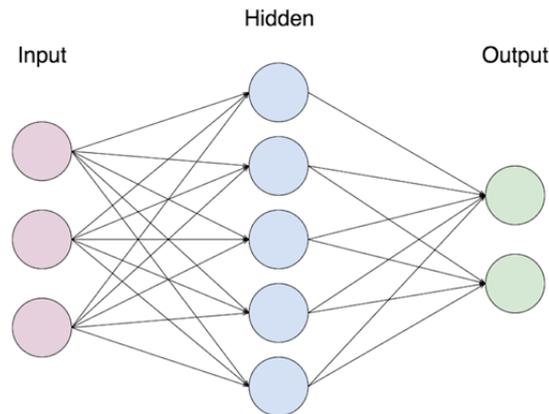


Figura 2.23: Esempio di struttura di una rete ANN [47]

relazione: $z^{[L]} = w^{[L]} * x^{[L-1]} + b^{[L]}$ dove w e b rappresentano due parametri caratteristici del neurone, denominati peso (weight) e bias e il cui valore viene inizializzato in maniera random ma si aggiorna durante la fase di allenamento sulla base della funzione di costo e del learning rate. Essa può essere rappresentata come una superficie convessa nello spazio J in funzione dei valori di peso e bias con un punto minimo o ottimo globale, in corrispondenza del quale si hanno prestazioni ottimali della rete. La cost function è la media della funzione di Loss, la quale misura la qualità delle performance dell'algorithm sul set di training e che può coincidere con l'errore quadratico.

L'output di ogni neurone viene calcolato in base al valore z , calcolato come sopra, e alla funzione di attivazione:

$$a = \sigma(z) = \frac{1}{1 + e^{-z}}$$

L'output layer della rete può essere composto da un solo neurone con funzione di attivazione lineare, che quindi restituirà un valore compreso nel range che va da -infinito a infinito. La rete implementata per questa task di classificazione è una feed forward net che presenta un hidden layer costituito da cinque neuroni ed è stata allenata con l'algorithm di Back Propagation. Il principale vantaggio della rete, nonché il motivo per cui si è deciso di implementarla, è la capacità di modellare un processo non lineare, basando il proprio funzionamento sui dati e non

su uno specifico set di regole di classificazione preimpostate, come accade per i metodi statistici.

Inoltre, in letteratura ci sono esempi di implementazione di reti su immagini di mammografia che hanno dato buoni risultati [1, 14]. Nonostante questo, è importante monitorare il fenomeno dell'overfitting controllabile con la variazione della complessità della rete e del numero di dati nel training set.

Setting dei Parametri

Tutti i metodi descritti sono stati implementati, sia nella prima che nella seconda fase del lavoro, in modo che fossero ottimizzati per dare buone prestazioni sul Dataset a disposizione. In particolare, i parametri sono stati settati nell'ottica di:

- Evitare l'overfitting, per cui si ottengono ottime prestazioni sui dati di training a fronte di prestazioni discrete sui dati di test del sistema a causa della perdita della capacità di generalizzazione, come già descritto
- Assicurare un basso numero di falsi positivi e una buona percentuale di corretti classificati: a fronte della misclassificazione dei dati, i metodi possono essere corretti affinché l'errore sia polarizzato, aumentando di fatto la sensibilità

I parametri sono stati settati in seguito all'ottimizzazione dei modelli eseguita implementando l'ottimizzazione Bayesiana con funzione di acquisizione $a(x)$ del tipo *expected improvement*. L'ottimizzazione Bayesiana ha come obiettivo trovare il minimo globale di una funzione e, per farlo, lavora su una copia definita per punti della funzione, o processo gaussiano. Il procedimento è basato su step successivi in cui viene valutata la funzione nei punti definiti *seed points*, viene aggiornato il modello gaussiano e vengono definiti nuovi punti che massimizzano la funzione di acquisizione $a(x)$. La funzione di acquisizione scelta calcola il rate di miglioramento previsto sulla base della posizione del punto più basso trovato, quindi più vicino al minimo, e il rispettivo valore associato. Sono stati testati i parametri numerici e altri quali

la metrica di calcolo delle distanze nel KNN, la natura della funzione di discriminazione dell'LDA o il tipo di kernel implementato nell'SVM. Nella tabella di figura 2.24 sono riportati i valori selezionati come migliori per l'implementazione dei classificatori sia per i modelli allenati sul Dataset iniziale che comprendeva 48 casi, che sul Dataset aggiornato già descritto precedentemente.

TUNING CLASSIFICATORI									
Immagini CVIEW									
96 IMMAGINI					174 IMMAGINI				
KNN	k = 7	BK 'smallest'	cityblock'		KNN	k = 5	BK 'smallest'	seuclidean'	
LDA	$\gamma = 0$	$\delta = 0.05$	linear'		LDA	$\gamma = 0$	$\delta = 0$	pseudolinear'	
SVM	BC = 10	Sh Period = 100	Nu 0.1	linear'	SVM	BC = 1000	Sh Period = 10000	Nu 0.1	linear'
Immagini BTO									
96 IMMAGINI					174 IMMAGINI				
KNN	k = 6	BK 'smallest'	euclidean'		KNN	k = 7	BK 'smallest'	seuclidean'	
LDA	$\gamma = 0$	$\delta = 0.01$	pseudolinear'		LDA	$\gamma = 0$	$\delta = 0.05$	pseudolinear'	
SVM	BC = 10	Sh Period = 100	Nu 0.1	linear'	SVM	BC = 1000	ShPeriod = 10000	Nu = 0.1	linear'

Figura 2.24: Tabella riassuntiva del setting dei classificatori con le migliori combinazioni di parametri

Inoltre, essendo il problema di classificazione diviso per le due tipologie di immagini a disposizione, sono riportati i risultati del tuning per entrambe le prove e la combinazione finale di parametri è stata scelta come compromesso tra i risultati che, in ogni caso, risultano molto simili. La struttura della rete è stata progettata principalmente sulla base dell'osservazione dei grafici di allenamento e alla stabilità dei risultati ottenuti oltre ai criteri già citati.

2.5.2 Convolutional Neural Network

Le reti convoluzionali sono algoritmi di Deep learning in grado di analizzare le immagini in ingresso in modo molto più approfondito, individuando dipendenze spaziali e caratteristiche significative con un maggiore livello di astrazione al fine della classificazione o della segmentazione di oggetti al loro interno. Lavorano sia con immagini in scala di grigi che RGB e vengono largamente utilizzate con immagini di mammografia e tomosintesi [21, 52, 53]. Presentano un considerevole

numero di vantaggi rispetto agli algoritmi di machine learning: tra questi, sicuramente, la capacità di analizzare l'immagine in modo molto più dettagliato oltre alla possibilità di ridurre il fenomeno di overfitting in quanto l'allenamento viene fatto su diversi set di dati con dimensioni inferiori. Le CNN sono composte da layers con funzioni

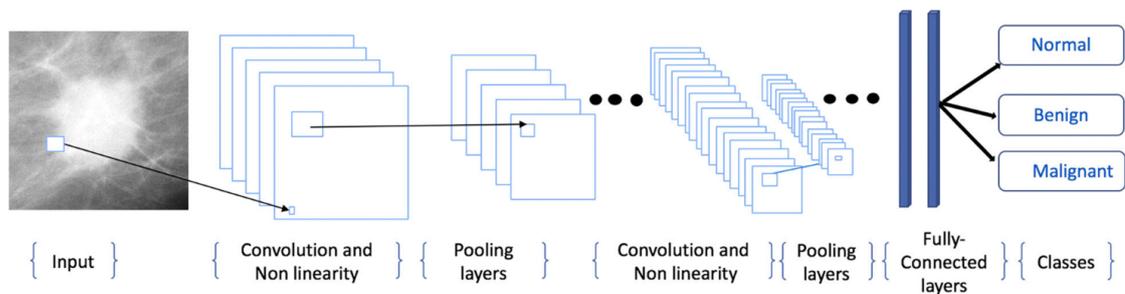


Figura 2.25: Esempio della struttura di una rete convoluzionale [31].

diverse. Le immagini vengono date in input al sistema e vengono elaborate per mezzo della convoluzione con kernel di diversa natura e tipologia. Alcune tipologie di layer utilizzabili sono (Figura 2.25) [54]:

- Layer convoluzionali: elaborano l'immagine per mezzo di un numero variabile di filtri con kernel di dimensioni e natura diversa, che vengono fatti scorrere sull'immagine in esame, secondo un dato valore di stride.
- Pooling layers: riducono la dimensionalità dello spazio delle features, introducendo anche una componente di de-noising dei pesi. Dividono l'immagine in regioni più piccole e per ognuna calcolano e riportano solo il valore massimo, nel caso dei Max pooling layers, oppure la media, nel caso degli Average Pooling layers.
- Fully connected layers: la loro configurazione è molto simile a quella di un hidden layer di una rete neurale standard. Lo scopo principale è la classificazione sulla base di combinazioni lineari delle features estratte ad alto livello e tramite l'utilizzo della tecnica di classificazione Softmax. Connettono l'input, che coincide con l'output dei layer precedenti, con una serie di neuroni

- ReLu layer: svolge un'operazione di filtraggio degli elementi in input per mezzo della funzione ReLU (Rectified linear unit), una funzione di attivazione non lineare

Si è scelto di indagare quali potessero essere le prestazioni raggiunte da una CNN custom di dimensioni relativamente ridotte per valutare le potenzialità dell'applicazione nonostante il limite sul numero di dati a disposizione e, in generale, per stabilire la fattibilità dello studio sia con tecniche di machine learning che di deep learning.

Pre-processing e Data augmentation

Se nei metodi di classificazione basati sul machine learning i dati in input sono costituiti da features estratte manualmente dalle regioni di interesse, nelle reti convoluzionali l'analisi delle immagini e l'estrazione di caratteristiche significative ad alto livello avviene senza l'intervento dell'utente. Per questo motivo, si è scelto di non pre-processare le immagini e di utilizzare come dataset l'insieme delle ROI estratte direttamente dalle immagini di mammografia sintetica originali. È necessaria, tuttavia, una fase di preparazione dei dati in quanto l'input layer della CNN richiede che le immagini abbiano tutte la stessa dimensione e siano quadrate. Le ROI sono state estratte, come già descritto, a partire dalle bounding box delle maschere di ogni lesione dilatate per mezzo del padding proporzionale. Data l'evidente eterogeneità nella dimensione delle lesioni ma, soprattutto, vista la maggioranza di bounding box di forma rettangolare, si è scelto di riscalarle le immagini in modo da ottenere un dataset con dimensioni omogenee. Nel grafico (Figura 2.26) è riportata la distribuzione dei valori del lato maggiore di ogni bounding box, il valore della dimensione media e il valore scelto per il rescaling, nonché la potenza di 2 più prossima al valore medio.

In questo modo si è ottenuto un Dataset composto da 174 ROI quadrate di dimensione 256x256 che sono state suddivise tra Benigne e Maligne. Infine, si è scelto di aumentare la numerosità del Dataset per mezzo di un lieve data augmentation. Si tratta di una metodologia che ha come obiettivi principali ridurre l'overfitting e aumentare la capacità di generalizzazione del modello applicando alle immagini

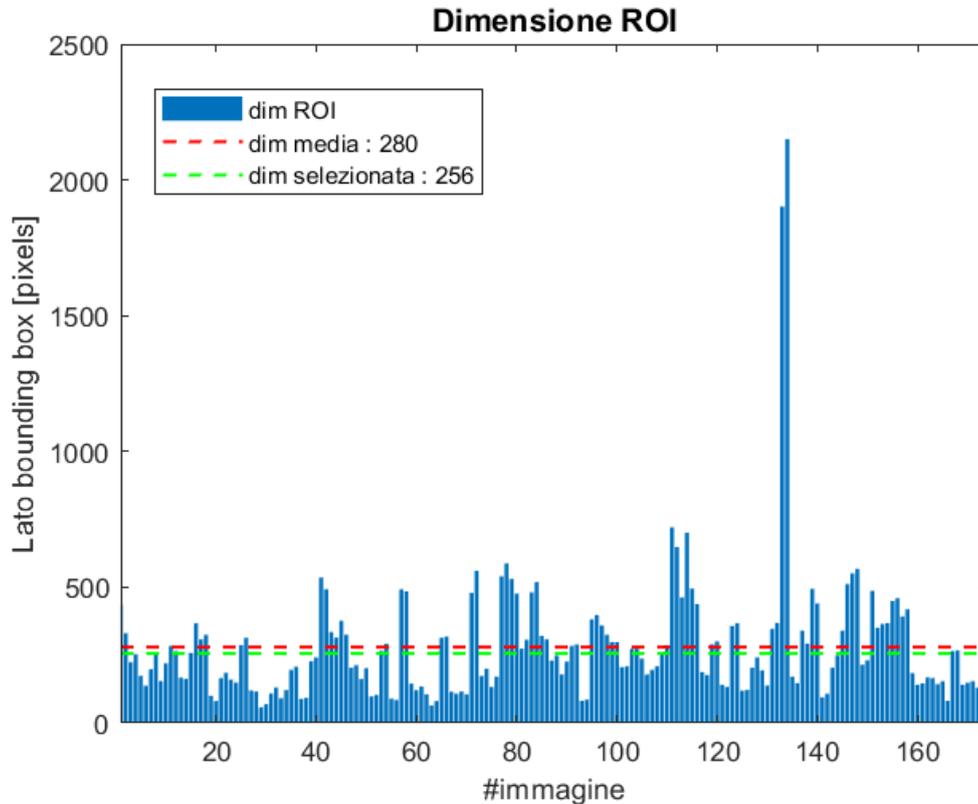


Figura 2.26: Distribuzione delle dimensioni delle ROI del Dataset con indicazione del valore medio e del valore scelto per il rescaling

alcune trasformazioni, tra cui trasformazioni geometriche, modifiche del valore di pixel o sovrapposizione di rumore. In questo modo si riesce ad aumentare artificialmente il numero di immagini nel Dataset introducendo nuovi elementi utilizzabili in fase di training. Da un'analisi della letteratura, si evince come molti studi ricorrano alla tecnica del data augmentation [31], principalmente per riuscire ad ottenere performance migliori con dataset limitati, ma anche per aumentare la robustezza dell'algoritmo allenandolo con un training set più ampio [55]. In questo caso, è stato implementato il data augmentation offline, applicando ad ogni immagine tre trasformazioni geometriche mediante l'utilizzo della funzione `affine2d`, e salvando le immagini ottenute su disco (Figure 2.27). Le trasformazioni sono state estratte in maniera random tra:

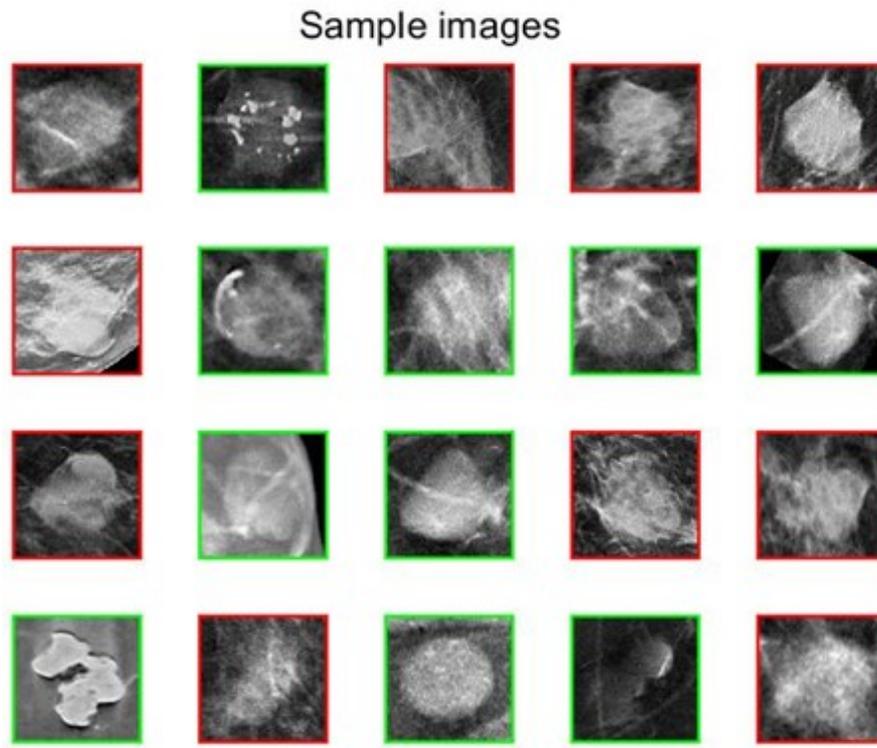


Figura 2.27: Esempi di ROI in seguito al data augmentation, con indicazione della classe di appartenenza: verde = Benigne, rosso = Maligne.

- Rotazione di un angolo pari a 90° , 180° oppure 270° , specificando la matrice di rotazione
- Flip dell'immagine lungo l'asse X o Y, con fattore di scala 'a' pari a 1 e angolo di rotazione 0:

$$T_x = \begin{bmatrix} sc & -ss & 0 \\ a * ss & a * sc & 0 \\ 0 & 0 & 1 \end{bmatrix}; T_y = \begin{bmatrix} a * sc & -a * ss & 0 \\ ss & sc & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Architettura della rete e allenamento

La rete custom implementata si ispira alla struttura delle reti VGG16 e VGG19, entrambe utilizzate in studi che hanno dato risultati promettenti [56, 55]. Per definire l'architettura sono state svolte numerose prove variando sia la struttura che i parametri di allenamento. Viene descritta la CNN che ha dato i migliori risultati sul Validation set, la quale presenta una struttura lineare e abbastanza semplice. È costituita da un Input layer per immagini quadrate di dimensioni $256 \times 256 \times 1$. Seguono tre layer convoluzionali, ognuno dei quali applica alle immagini 4 filtri di dimensione 12×12 . Tra i layer convoluzionali sono stati inseriti, in sequenza:

- Batch Normalization layer con fattore di normalizzazione L2 pari a 0.005
- ReLU layer
- Max Pooling layer con regioni di pooling di dimensione 2 e stride pari a 2

Sono stati aggiunti ulteriori due layers convoluzionali che applicano 8 filtri di dimensione 12×12 , nuovamente seguiti dal blocco di layers citato prima ma con valore di stride pari a 4. Infine, sono stati inseriti un layer di dropout, che imposta alcuni degli elementi in input a 0 con data probabilità, seguito da Fully Connected Layer, Softmax Layer e Classification Layer. Anche in fase di allenamento sono state utilizzate diverse combinazioni di parametri, ma i risultati migliori sono stati ottenuti impostando il learning rate iniziale pari a 0.01 e fattore di drop del learning rate pari a 0.9, di modo che ogni 5 epoche il valore risultasse ridotto del 10%. La dimensione dei mini batch utilizzata è pari a 32 immagini.

2.6 Metodi di validazione

Al fine di validare i metodi di classificazione sviluppati, si è scelto di implementare il K-fold cross validation. Come descritto nella sezione

dedicata alla struttura del Dataset, il numero totale di immagini a disposizione è pari a 174. Tuttavia, questo valore si raggiunge contando separatamente le due proiezioni, cranio caudale e medio laterale obliqua, acquisite per ogni lesione. Si è scelto di suddividere il Dataset includendo sempre entrambe le proiezioni per ogni lesione in modo che, anche qualora presentassero caratteristiche molto simili, questo non si traducesse in una polarizzazione dei risultati. Per rispettare questo criterio, il K-fold è stato implementato sui pazienti e non sulle singole immagini e, in base al numero di elementi a disposizione, si è deciso di implementare un 5-Fold Cross Validation. L'algoritmo viene utilizzato per avere una stima delle capacità del modello di classificare dati nuovi [57] e può essere riassunto nelle seguenti fasi:

1. Creazione di K sub-fold di validation: il Dataset completo viene suddiviso in modo da ottenere K gruppi di elementi. Nella prima fase del progetto, con un numero di dati a disposizione estremamente limitato, si è scelto di creare i sottogruppi manualmente per garantire, all'interno di ognuno di essi, una buona rappresentatività sia in termini di classe di appartenenza delle lesioni che per quanto riguarda la loro dimensione. Nella seconda fase invece, dato l'aumento della casistica, gli elementi sono stati estratti in maniera random, ma sempre mantenendo il bilanciamento delle classi.
2. Ottimizzazione dei modelli di classificazione per ognuno dei sottogruppi: i restanti elementi vengono utilizzati per l'allenamento dei classificatori, i quali vengono successivamente testati con il corrispondente sub-fold di validation.
3. Salvataggio delle prestazioni: le prestazioni di ognuno dei K modelli vengono memorizzate per poter essere valutate successivamente.

La scelta di questo metodo di validazione è strettamente legata alla numerosità del Dataset poiché in questo modo si riduce in parte l'effetto che l'allenamento con training set ridotto può avere sulle prestazioni. Si può valutare inoltre il livello di rappresentatività del Dataset sulla base della variazione nelle prestazioni dei modelli, individuando la presenza

di sottogruppi di training maggiormente in grado di rendere il modello capace di generalizzare oppure la presenza di outliers.

2.6.1 Metriche di validazione

Infine, per quantificare le prestazioni ottenute sia nella task di classificazione di lesioni su immagini di tomosintesi che su immagini di mammografia 2D sintetica, sono state utilizzate diverse metriche descritte di seguito.

Confusion Matrix

La Confusion Matrix rappresenta uno dei metodi di visualizzazione delle performance di un algoritmo più comuni che permette di quantificare egli errori commessi dal classificatore [58]. Sulla diagonale della matrice viene riportato il numero di volte in cui la classe predetta e la classe reale coincidono, mentre nelle restanti caselle viene riportano il numero di elementi classificati erroneamente. La struttura standard di una confusion matrix per un problema di classificazione binaria è quella riportata nell'immagine (Figure 2.28): Le actual classes sono le classi

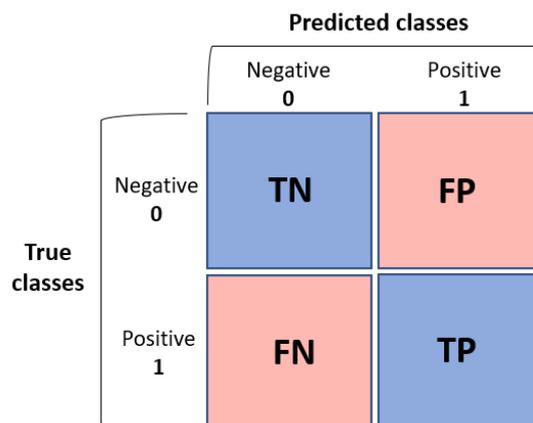


Figura 2.28: Schema di rappresentazione di una Confusion Matrix

reali che, in questo lavoro, sono la classe delle lesioni benigne e quella delle lesioni maligne. Le predicted classes, ossia le classi predette, sono

le classi alle quali il classificatore assegna gli elementi in esame. I valori riportati rappresentano rispettivamente:

- TN e TP: True Negative e True Positive, ossia gli elementi di classe 0/1 correttamente classificati come appartenenti alla classe 0/1
- FN e FP: Falsi Negativi e Falsi Positivi sono gli elementi misclassificati

A partire dai valori descritti, possono essere calcolate diverse metriche di valutazione, tra cui:

- Sensibilità o TPR (True Positive Rate) o RECALL, ossia il numero di istanze positive correttamente classificate come tali sul totale:

$$TPR = (TP) / (TP + FN)$$

- Specificità o TNR (True Negative Rate):

$$TNR = TN / (TN + FP)$$

- Precision o PPV (Positive Predictive Value):

$$PPV = TP / (TP + FP)$$

- F1-Score: si calcola come: $F1 = 2(PR * RECALL) / (PR + RECALL)$ e restituisce un valore alto solo quando sia Precision che Recall hanno valori alti.

Accuracy

L'accuracy è una metrica basata sulla confusion matrix ed è la percentuale di elementi che vengono correttamente classificati dal sistema sul numero totale di elementi in esame. Viene calcolata a partire dai valori della confusion matrix secondo la formula:

$$Accuracy = \frac{(TN + TP)}{(TN + FP + FN + TP)}$$

Fornisce una stima generale delle performance dell'algoritmo e rappresenta un valore significativo specialmente nel caso in cui il dataset sia bilanciato ma non dà informazione sulle performance relative alle singole classi.

Curve ROC e Area Under Curve (AUC)

La curva ROC, acronimo di Receiver Operating Characteristics è una curva di probabilità ed è uno strumento fondamentale per la valutazione dei classificatori binari. Essa descrive le performance del classificatore al variare della soglia di discriminazione impiegata e quantifica la probabilità di ottenere dei veri positivi rispetto a quella di ottenere dei falsi positivi [59]. Viene rappresentata su un piano in cui il valore di probabilità di ottenere dei falsi positivi varia sull'asse x, mentre il valore di sensibilità varia sull'asse y. Perché le prestazioni del classificatore siano considerate buone, la curva ROC deve trovarsi completamente al di sopra della retta di scelta casuale (Figura 2.29) rappresentata dalla bisettrice del piano e che descrive il comportamento di un classificatore che assegnerebbe un elemento ad una classe in modo completamente casuale. Il piano è diviso in una zona di sensibilità alta e una di sensibilità bassa: nella zona di sensibilità bassa, un buon classificatore risponderà a grandi variazioni di sensibilità con minime variazioni di probabilità di ottenere dei falsi positivi e, a parità di probabilità, avrà una soglia decisamente maggiore di quella del classificatore descritto dalla retta di scelta casuale. La curva A in figura, rappresenta invece il comportamento del classificatore ideale poiché è caratterizzato da un valore di probabilità di ottenere falsi positivi nullo e una sensibilità del 100%.

La curva Precision-Recall è una metrica utilizzata quando vi è uno sbilanciamento del Dataset o quando uno dei due errori tra *falsi positivi* e *falsi negativi* risulta essere più importante, come in questo caso. Si riporta sull'asse Y il valore di Precision, mentre sull'asse X si ha la Recall. Il classificatore ideale è rappresentato dalla curva che tocca il punto (1,1) in alto a destra del piano, mentre, come accade per la curva ROC, il classificatore casuale avrà una curva sovrapposta alla bisettrice del piano. La baseline della curva dipende dallo sbilanciamento del Dataset e rappresenta un classificatore che attribuisce tutte le istanze alla classe dei positivi, in questo caso classe 1 (Figura 2.30). L'Area Under Curve è un ulteriore parametro di quantificazione delle performance che si calcola come integrale della curva normalizzato rispetto all'area totale del piano, per questo risulta essere un valore compreso tra 0 e

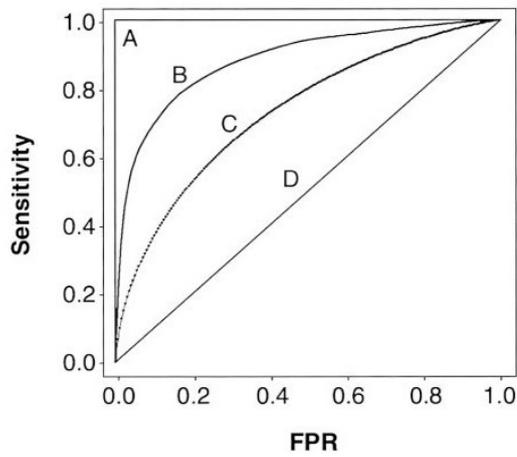


Figura 2.29: Esempi di curve ROC.

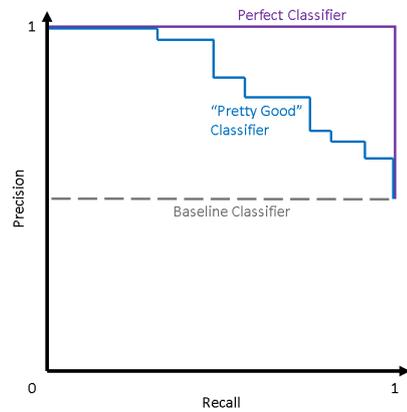


Figura 2.30: Esempi di curve Precision - Recall [60].

1. In particolare, da informazione sulle capacità del classificatore di distinguere tra due classi. Il classificatore ideale avrà un valore di AUC pari al 100% mentre un buon classificatore sarà caratterizzato da un valore di AUC maggiore dell'80%. Infine, si tratta di un buon parametro per la quantificazione delle performance in quanto non dipende dalla soglia di classificazione.

Capitolo 3

Risultati

Nel seguente capitolo si riportano i risultati ottenuti con i metodi di caratterizzazione e classificazione delle lesioni descritti precedentemente. In fase di sviluppo del progetto, è stata indagata l'influenza della numerosità del Dataset sui risultati ottenibili. In seguito, una volta definite le strategie di pre-processing e di features extraction sul Dataset completo, sono state eseguite diverse prove volte a:

- Valutare le performance di classificazione al variare della tipologia di immagine utilizzata, individuando quale permette di ottenere risultati migliori.
- Indagare l'efficacia di diverse strategie per l'utilizzo dell'informazione 2D e 3D della lesione sulle due proiezioni
- Individuare le principali criticità e potenzialità dei metodi utilizzati

Infine, sono stati analizzati i migliori risultati ottenuti confrontandoli direttamente e con la letteratura.

3.1 Effetto dell'aumento della numerosità del Dataset

Il primo test effettuato riguarda l'effettivo impatto che un Dataset con maggiore numerosità ha sulle prestazioni dei diversi metodi di

classificazione implementati, anche in correlazione alla tipologia di dati inclusi. I 48 casi iniziali sono stati utilizzati per avere una prima stima delle performance ottenibili: i classificatori sono stati allenati separatamente con le 96 lesioni su immagini 2D sintetiche e sulle key-slice delle immagini BTO, e sono stati validati con il 5-Fold Cross Validation. I risultati del training e del test, in termini di AUC, Accuracy e Recall, sono stati organizzati per essere confrontati direttamente con quelli ottenuti in seguito sulle immagini del Dataset totale. Gli step di pre-processing, le tecniche di estrazione delle features e i parametri dei classificatori non sono stati modificati al fine di avere un confronto che dipendesse unicamente dal numero di elementi (Figura 3.1).

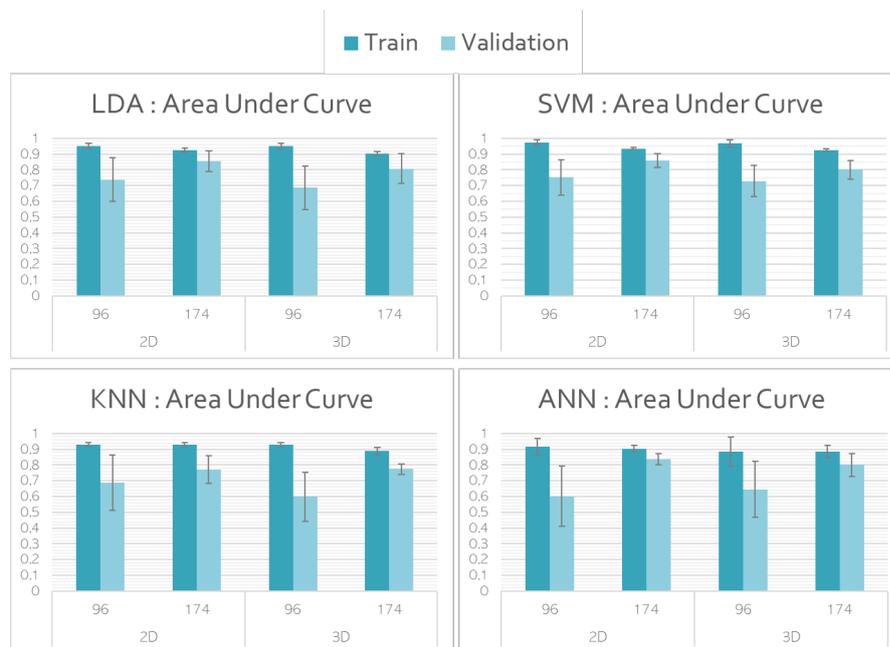


Figura 3.1: Grafici di confronto per i valori medi di AUC al variare del numero di dati, da 96 a 174 immagini, e della tipologia di immagini, indicando con 2D le immagini sintetiche e con 3D le tomosintesi

Il primo grafico mostra i valori di Area Under Curve ottenuti per ogni classificatore al variare del numero di immagini e della tipologia di immagini utilizzate. In particolare, quello riportato è il valore medio ottenuto sui cinque sub-fold di training e validation con relativo valore



Figura 3.2: Grafici di confronto per i valori medi di Accuracy e Recall

di deviazione standard. Allo stesso modo sono stati confrontati i valori di Accuracy e Recall, inseriti in un unico grafico in quanto strettamente correlati (Figura 3.2). Dall'osservazione dei grafici, si può notare come l'aumento di immagini utilizzate per la costruzione dei classificatori, a prescindere dal metodo di classificazione utilizzato, implichi:

- L'aumento delle prestazioni sul set di Validation: in tutti i casi e per entrambe le tipologie di immagini utilizzate, la media delle prestazioni in fase di validazione aumenta. L'incremento può essere più o meno evidente, ma vi sono differenze significative in diversi casi, ad esempio per l'ANN sia in termini di AUC che di accuracy.
- Una visibile riduzione del drop di prestazioni tra Training e Validation suggerisce l'aumentata capacità dei classificatori di generalizzare e quindi, un ridotto overfitting ai dati di training.
- In generale, una riduzione della variabilità dei risultati in funzione delle immagini utilizzate in fase di allenamento e di test, nonostante in alcuni casi il range di variazione risulti ancora ampio.

- La diminuzione delle prestazioni ottenute sul training set: si osservano, in alcuni casi, prestazioni peggiori sui dati di training dovute probabilmente alle caratteristiche dei classificatori scelte per ridurre il più possibile il fenomeno di overfitting sui dati a discapito di prestazioni migliori. Si nota principalmente nell'SVM, sia nei valori di AUC che di Accuracy.

E' stata fatta un'ulteriore analisi della composizione del gruppo di parametri più significativi al fine della classificazione al variare della numerosità (Figura 3.3). L'insieme delle 219 features è stato diviso in sottocategorie sulla base del tipo di immagine su cui sono state calcolate, in particolare immagine originale della lesione o della regione esterna, regione interna filtrata con filtro di Sobel o intera ROI filtrata con i vari filtri di Laws. Per lo stesso principio, i descrittori di forma sono stati considerati separatamente. Nella tabella è riportata la composizione della matrice totale, con indicata la percentuale sul totale di ognuna delle sottocategorie.

Categoria	Numero (Percentuale)
Texture regione interna	31 (14%)
Texture regione esterna	31 (14%)
Laws Energy Texture	140 (64%)
Edge Enhanced Texture	5 (2%)
Descrittori di forma	12 (5%)

Tabella 3.1: Divisione in sotto-categorie delle features calcolate

Considerando le 20 features più significative selezionate dal test ANOVA su tutto il Dataset, è stata quantificata la rappresentatività dei diversi tipi di parametri calcolati. Si può notare come i descrittori di forma, in linea coi risultati attesi, siano significativi anche all'aumentare delle immagini e costituiscano, nonostante siano soltanto 12, rispettivamente un quinto e un terzo delle features utilizzate. La percentuale di parametri di tessitura della regione esterna diminuisce mentre aumenta quella di features che caratterizzano la regione interna alla lesione da 0 a 5%, probabilmente in correlazione alla maggiore rappresentatività

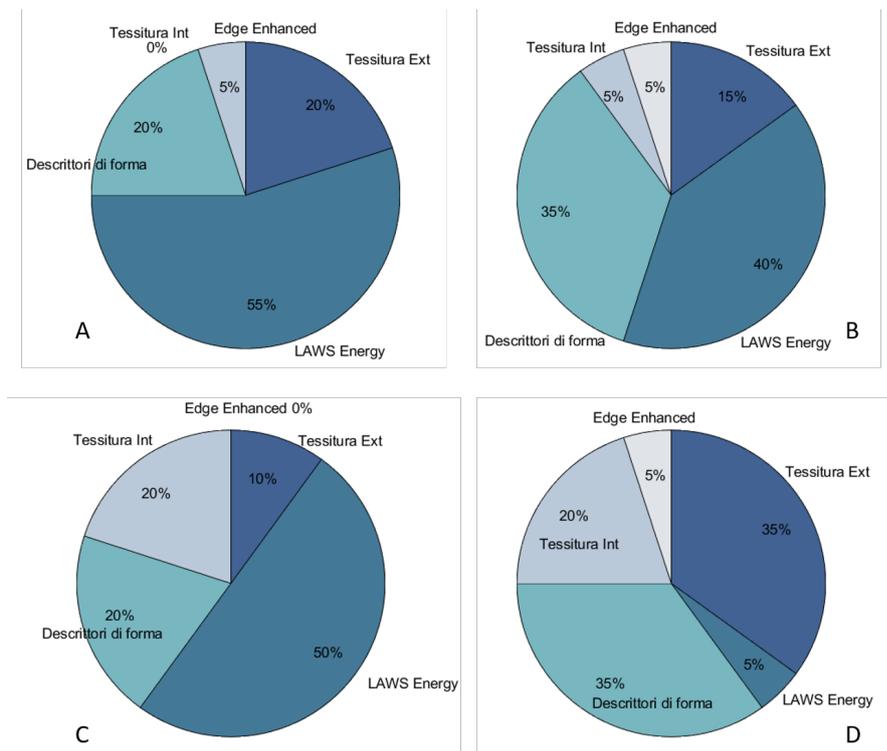


Figura 3.3: Composizione del gruppo di features utilizzate per la classificazione delle lesioni su 96 e 174 immagini 2D sintetiche (A e B) e su 96 e 174 immagini di tomosintesi (C e D).

e all'individuazione di alcuni pattern comuni tra casi simili. Rimane costante il numero di features calcolate sulle immagini con enhancement delle discontinuità e costituisce il 5% sul totale nonostante si tratti di solo 5 parametri sul totale. La complessità delle immagini 2D sintetiche dovuta alla sovrapposizione di più slices spiega l'efficacia della caratterizzazione con la tessitura di Laws e, allo stesso modo, il motivo della differenza con i risultati ottenuti sulle singole slices dell'immagine di tomosintesi. Per lo stesso motivo la tessitura della regione esterna e interna risulta più significativa. Infine, anche nel caso delle immagini BTO si conferma l'efficacia dei descrittori di forma e dei parametri calcolati sulle immagini filtrate con Sobel.

Le variazioni osservate in seguito al raddoppiamento del numero di casi potrebbero non essere generalizzabili, poiché altri fattori possono

influire molto sui valori ottenuti, ma sicuramente evidenzia l'effettiva informazione contenuta in alcuni dei parametri calcolati e suggerisce un possibile trend all'aumentare del Dataset.

3.2 Classificazione di lesioni su immagini 2D sintetiche

In fase di studio delle immagini 2D sintetiche sono state applicate due strategie differenti per la classificazione delle lesioni. Nel primo test sono state utilizzate tutte le 174 immagini all'interno del Dataset, considerando le due proiezioni come casi distinti mentre, nel secondo test, sono state unite le informazioni derivanti da entrambe le proiezioni per classificare gli 87 casi. Per ogni test, i classificatori sono stati inizialmente allenati e testati su tutti gli elementi del Dataset, ottenendo i risultati mostrati tramite Confusion Matrix e curve ROC. In seguito alla validazione con il 5-Fold Cross Validation, le prestazioni per ogni Fold sono state confrontate tramite curve ROC e, in seguito, mediate per ottenere una panoramica dei risultati dei quattro metodi. In tutti i casi sono state analizzate le Features utilizzate per la classificazione.

3.2.1 Risultati sul Training

I primi risultati riportati sono relativi al primo Test, dove la proiezione cranio caudale e medio laterale obliqua di ogni lesione sono considerate separatamente. La motivazione principale di questa scelta è legata alla volontà di massimizzare il numero di elementi a disposizione, ma risulta inoltre interessante indagare il comportamento di un sistema che classifica separatamente le proiezioni poiché, in caso di predizioni concordi, si ha un maggiore livello di confidenza nei risultati. Per avere un'idea generale delle performance, sono state calcolate le Confusion Matrix di ogni classificatore, riportando i valori ottenuti per le diverse metriche di validazione utilizzare. Sono stati inoltre valutati i risultati che si ottengono sulle singole proiezioni (Figura 3.4, 3.5).

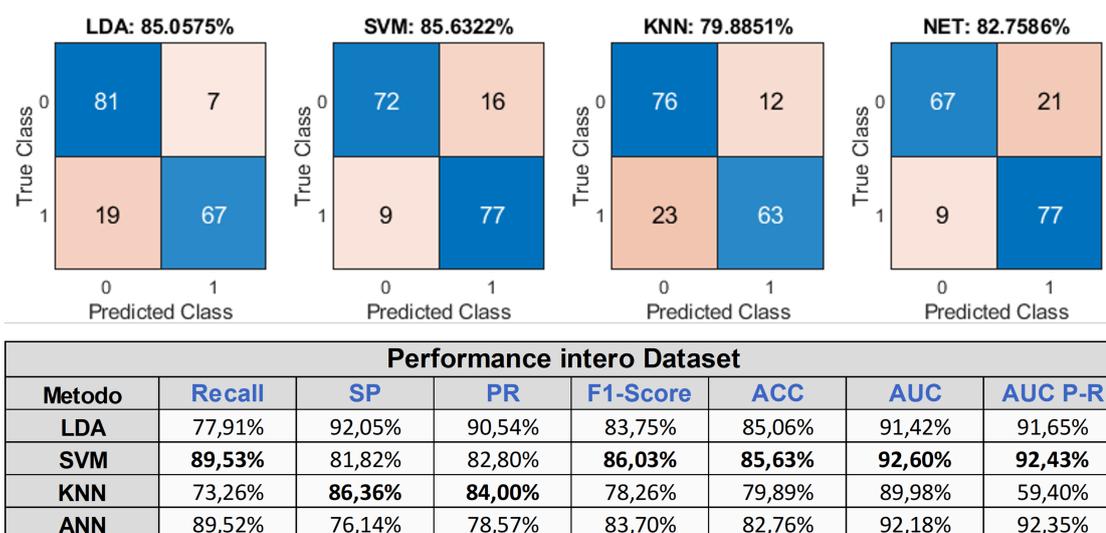


Figura 3.4: Confusion Matrix dei quattro metodi su immagini 2D sintetiche e tabella riassuntiva dei valori ottenuti per le diverse metriche utilizzate, nell'ordine: Recall, Specificity, Precision, F1-Score, Accuracy, AUC e AUC della curva Precision-Recall (Test 1). In grassetto i valori massimi ottenuti per ogni metrica.

Performance immagini proiezione CC							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	69,77%	93,18%	90,91%	78,95%	81,61%	88,48%	67,18%
SVM	88,37%	84,09%	84,44%	86,36%	86,21%	90,80%	67,69%
KNN	72,09%	86,36%	83,78%	77,50%	79,31%	89,43%	67,83%
ANN	86,05%	77,27%	78,72%	82,22%	81,61%	88,21%	67,09%

(a) Performance classificatori sulle sole immagini in proiezione CC

Performance immagini proiezione MLO							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	86,05%	90,91%	90,24%	88,10%	88,51%	94,61%	68,17%
SVM	90,70%	79,55%	81,25%	85,71%	85,06%	94,71%	68,21%
KNN	74,42%	86,36%	84,21%	79,01%	80,46%	90,35%	68,13%
ANN	93,02%	75,00%	78,43%	85,11%	83,91%	95,93%	68,43%

(b) Performance classificatori sulle sole immagini in proiezione MLO

Figura 3.5: Confronto delle performance ottenute dividendo le proiezioni e lavorando solamente con le proiezioni CC (a) o con le MLO (b). In grassetto i valori massimi ottenuti per ogni metrica.

I modelli sono stati in seguito allenati su metà del Dataset, considerando solamente le immagini in proiezione CC o MLO. Dai valori nelle tabelle riportate si nota come, a prescindere dal metodo di classificazione, siano le immagini in proiezione CC ad essere misclassificate più facilmente. Particolarmente marcata è la differenza nei valori di Recall, dove si ha un netto miglioramento per LDA, SVM e ANN. Anche l'AUC risulta maggiore utilizzando le sole proiezioni MLO. Sulla base del risultato ottenuto, è stata valutata la capacità dei metodi di classificare correttamente entrambe le proiezioni della stessa lesione o, in generale, restituire un risultato coerente tra proiezioni. Per farlo, è stata calcolata la percentuale di casi sul totale in cui la classe predetta per le due proiezioni risultava diversa e i valori sono riportati in tabella 3.2. La percentuale migliore è restituita dall'SVM.

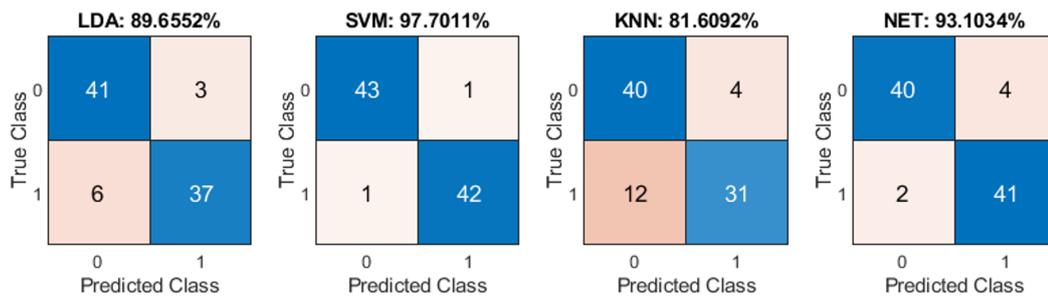
	LDA	SVM	KNN	ANN
% Discordanza tra proiezioni (CC e MLO)	18,39%	14,94%	17,24%	20,69%

Tabella 3.2: Mammografia sintetica: percentuale di casi in cui le due proiezioni vengono assegnate a classi diverse dai diversi metodi di classificazione.

La separazione delle proiezioni in fase di classificazione comporta sicuramente la polarizzazione dei risultati dovuta ad eventuali caratteristiche in comune tra proiezioni. Si tratta infatti di acquisizioni della stessa porzione di tessuto, sebbene la direzione di compressione diversa comporti la sovrapposizione di strutture diverse. Per valutare in che misura questo aspetto incidesse sulle prestazioni, è stato eseguito il secondo test descritto: in fase di classificazione sono stati considerati i migliori parametri tra tutti quelli calcolati su entrambe le proiezioni, sempre sulla base del test ANOVA.

In base alle Confusion Matrix (Figura 3.6) ricalcolate per ogni metodo, le percentuali sul training risultano migliori. È particolarmente evidente il miglioramento nelle performance dell'SVM, seguito dalla rete. Sicuramente la caratterizzazione delle opacità risulta in questo

modo più completa, tuttavia i risultati sono da verificare nella fase di validazione.



Performance intero Dataset							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	86,05%	93,18%	92,50%	89,16%	89,66%	96,93%	94,75%
SVM	97,67%	97,73%	97,67%	97,67%	97,70%	98,52%	96,28%
KNN	72,09%	90,91%	88,57%	79,49%	81,61%	89,06%	60,89%
ANN	95,35%	90,91%	91,11%	93,18%	93,10%	94,03%	90,08%

Figura 3.6: Confusion Matrix e tabella riassuntiva dei valori ottenuti per le diverse metriche utilizzate (Test 2).

L'ultimo confronto tra i classificatori viene fatto tramite curve ROC (Figura 3.15). Il grafico evidenzia come le prestazioni sul training siano paragonabili per LDA, SVM e ANN, leggermente inferiori per il KNN per il quale si nota, dall'andamento della curva, il valore inferiore di Recall: risulta particolarmente evidente nelle curve Precision Recall, dove la curva del KNN si trova sempre al di sotto delle altre. Al fine di confrontarle direttamente, si riportano le curve ottenute per entrambi i test effettuati. Nel caso del secondo test, il ridotto numero di elementi si riflette visibilmente sull'irregolarità delle curve.

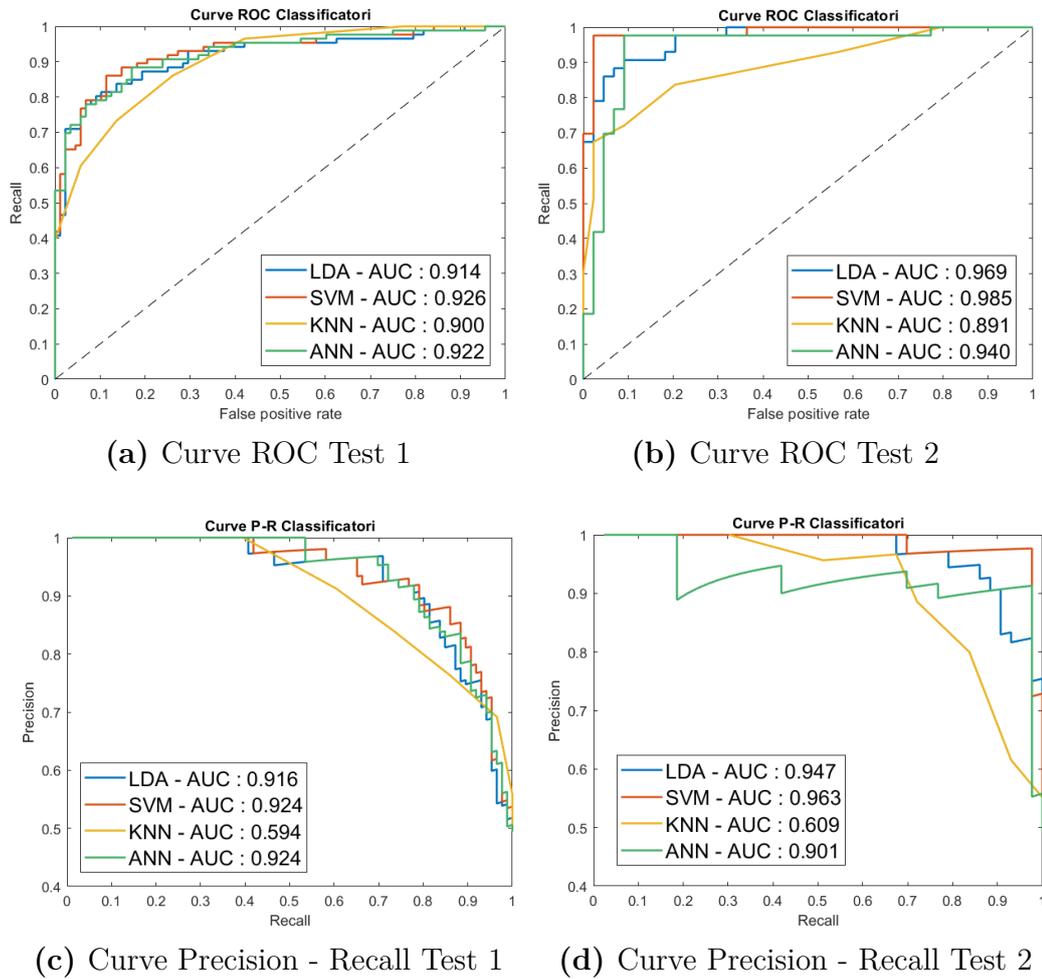


Figura 3.7: Curve di performance a confronto per i quattro metodi di classificazione e in base ai Test eseguiti.

3.2.2 Risultati sul Validation

Anche in questo caso vengono riportati separatamente i risultati ottenuti con le due strategie scelte. In primo luogo, per valutare visivamente la variabilità delle performance al variare delle immagini utilizzate per la validazione, si è deciso di rappresentare le cinque ROC curve ottenute per ognuno dei sub-fold di validation (Figura 3.8). Come riferimento, oltre alle ROC relative ai fold di validation, sono riportate in azzurro anche le curve ROC relative ai fold di training. Come auspicabile, queste

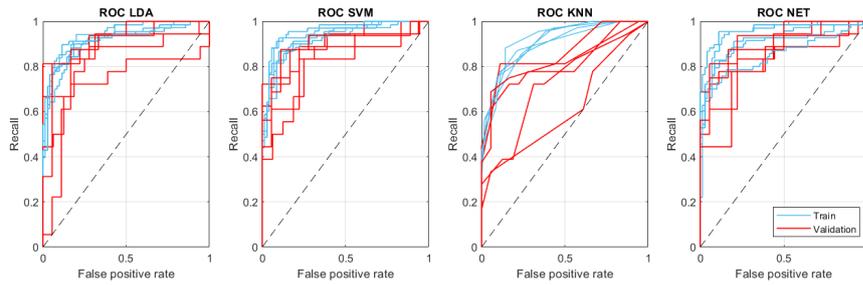


Figura 3.8: Test 1: ROC sui sub-fold del 5 Fold Cross-Vaidation

ultime sono per lo più confrontabili tra loro e con i risultati ottenuti nella fase precedente. Osservando i dati relativi ai set di validazione, tuttavia, si nota un discostamento che suggerisce una dipendenza dai dati confermata dai dati numerici riportati in seguito (Figura 3.9). Nonostante questo, e fatta eccezione per il KNN, i risultati sono ripetibili e parzialmente sovrapposti a quelli di training. Successivamente, al fine di un confronto quantitativo delle performance, sono stati calcolati i valori medi di AUC, Accuracy e Recall e riportati nello stesso grafico. A livello teorico, un buon classificatore ha un valore di AUC pari o

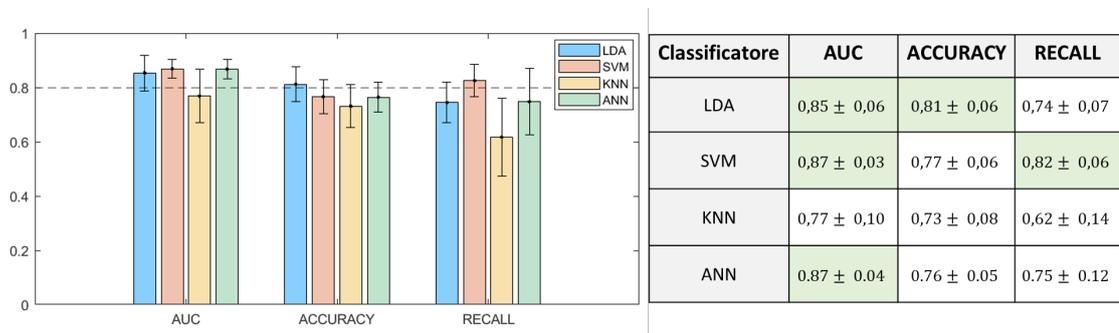


Figura 3.9: Confronto dei valori medi di AUC, Accuracy e Recall ottenuti su immagini 2D sintetiche.

maggiore a 0.8: in base a questa soglia, riportata anche graficamente, sono stati evidenziati in tabella i risultati migliori. La migliore Recall si raggiunge con l'SVM, con il quale si ottiene anche il valore più alto di AUC. In generale la variabilità rimane quasi sempre al di sotto del 10%. Risulta invece decisamente meno accettabile quella sui risultati

che si ottengono nel secondo test (Figura 3.10, 3.11). Lo svantaggio principale nel considerare i singoli casi è che significa dimezzare il Dataset tornando, di fatto, nelle condizioni iniziali: questo incide sia sulle curve ROC che sui valori di deviazione standard riportati, tutti maggiori del 10%. Se da una parte la numerosità ha sicuramente un impatto sui valori ottenuti, nonostante il numero di elementi sia paragonabile le prestazioni sono migliori di quelle iniziali, e questo è strettamente legato alla caratterizzazione più completa fornita ai classificatori. Gli ottimi risultati ottenuti in fase di training sono probabilmente legati all'overfitting dei dati: nessuno dei quattro metodi è in grado di estrarre pattern comuni per tutti i fold e questo è legato alla rappresentatività del Dataset resa artificialmente maggiore considerando le singole proiezioni come casi.

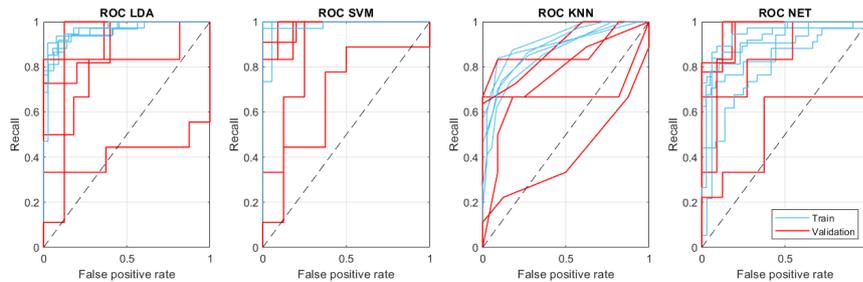


Figura 3.10: Test 2: ROC sui sub-fold del 5 Fold Cross-Validation

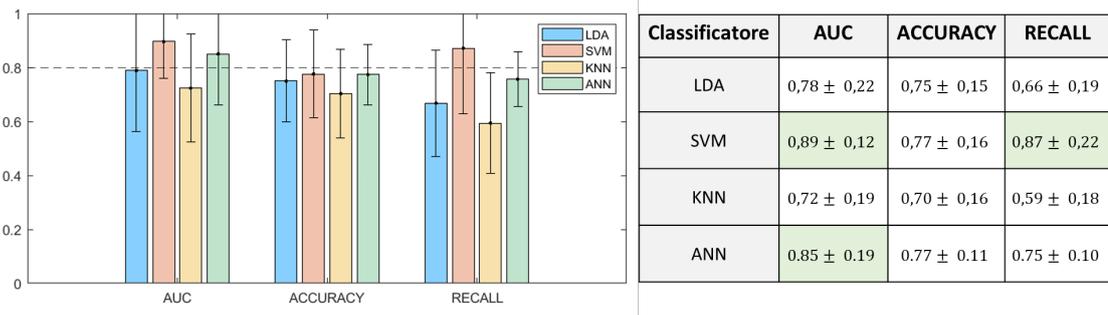


Figura 3.11: Confronto dei valori medi di AUC, Accuracy e Recall ottenuti per il Test 2.

Per ridurre la variabilità, si può considerare un maggiore numero

di parametri a discapito dei massimi risultati ottenibili. Il numero di features è stato variato considerandone da 20 a 30: questo ha portato a migliori risultati dell'SVM ma ad un peggioramento dell'LDA e del KNN. Le 20 features più significative selezionate in base alla strategia di classificazione sono state analizzate e mostrate nei grafici in Figura 3.12. Sull'asse y è riportata la frequenza con la quale le singole features vengono selezionate come più significative per i vari fold mentre, sull'asse x, in ordine di frequenza, sono riportate le singole features con il nome associato in fase di studio.

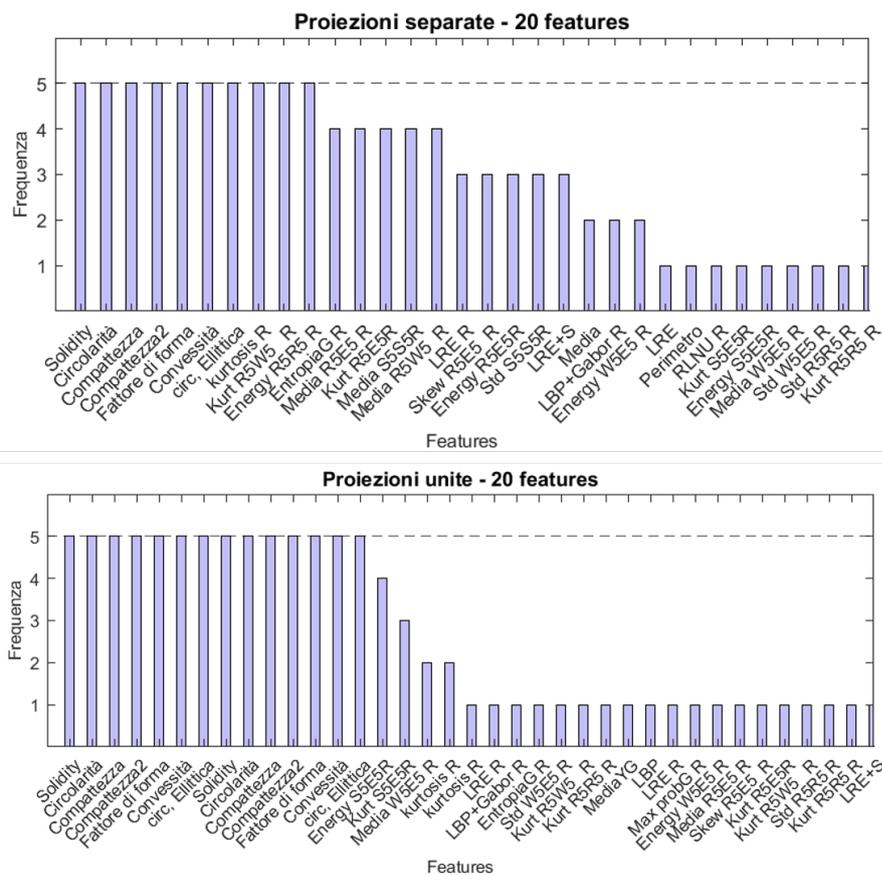


Figura 3.12: Feature selection per i due Test effettuati

I grafici si riferiscono rispettivamente al primo e al secondo test. Si nota subito come nel primo caso ci sia una minoranza di features selezionate per un solo sottogruppo, mentre il numero aumenta nel secondo

test: questa differenza è spiegabile con la poca rappresentatività di tutti i sottogruppi del Dataset, per i quali vengono considerate caratteristiche diverse. I descrittori di forma rientrano sempre nel gruppo selezionato con massima frequenza perché, come già indagato, sono altamente informativi per la discriminazione delle due classi e, come emerge da questi test, lo sono a prescindere dalle proiezioni: questa ipotesi è confermata dal fatto che, quando entrambe vengono considerate per classificare il singolo caso, i descrittori di forma vengono selezionati sia in proiezione CC che in MLO. Seguono una serie di parametri calcolati sulla regione esterna, ad esempio la Kurtosis calcolata sull'immagine di base o Energia e Kurtosis ottenute sull'immagine filtrata con i kernel di Laws. Anche queste risultano molto significative, soprattutto in quanto normalizzate rispetto ai valori calcolati all'interno della regione e quindi indipendenti di fatto dall'intensità dell'immagine.

3.3 Classificazione di lesioni su slices di Tomosintesi

Le stesse prove sono state eseguite sulle immagini di Tomosintesi. Sono stati impiegati inoltre gli stessi metodi e le stesse metriche di validazione con lo scopo di avere un confronto diretto tra le due tipologie di immagini. Come fatto precedentemente con i due test descritti, in questo caso si è scelto di testare i diversi metodi al variare della strategia di calcolo dei parametri in fase di caratterizzazione. I valori sono stati calcolati sulla key-slice oppure su un range di rispettivamente 3, 7 e 9 slices che comprendessero quella principale e successivamente mediati. Si riportano i risultati migliori, ottenuti con la prima strategia di calcolo, oltre ad alcune evidenze del confronto tra le varie strategie.

3.3.1 Risultati sul Training

La scelta della key-slice è un elemento che introduce forte variabilità, in quanto soggetta a variabilità intra e inter-operatore. Tuttavia, nella maggior parte dei casi, da un'osservazione delle immagini del Dataset si

evince come l'opacità mantenga pressappoco la stessa forma, intensità e struttura tra slice adiacenti, con le dovute eccezioni costituite principalmente da lesioni con dimensioni ridotte o con microcalcificazioni. L'andamento delle features è pressappoco costante al variare delle slices. Per ridurre l'impatto di questa scelta si possono mediare i parametri su un certo numero di slices, considerando che questa operazione introdurrà un certo quantitativo di rumore nei valori, il quale potrà incidere sulla classificazione. Fisicamente, l'effetto che la media ha sui parametri è traducibile con l'effetto di sovrapposizione dei tessuti e quindi una parziale perdita del vantaggio delle immagini di tomosintesi.

Per avere una prima stima delle performance sono state calcolate le Confusion Matrix (Figure 3.13,3.14):

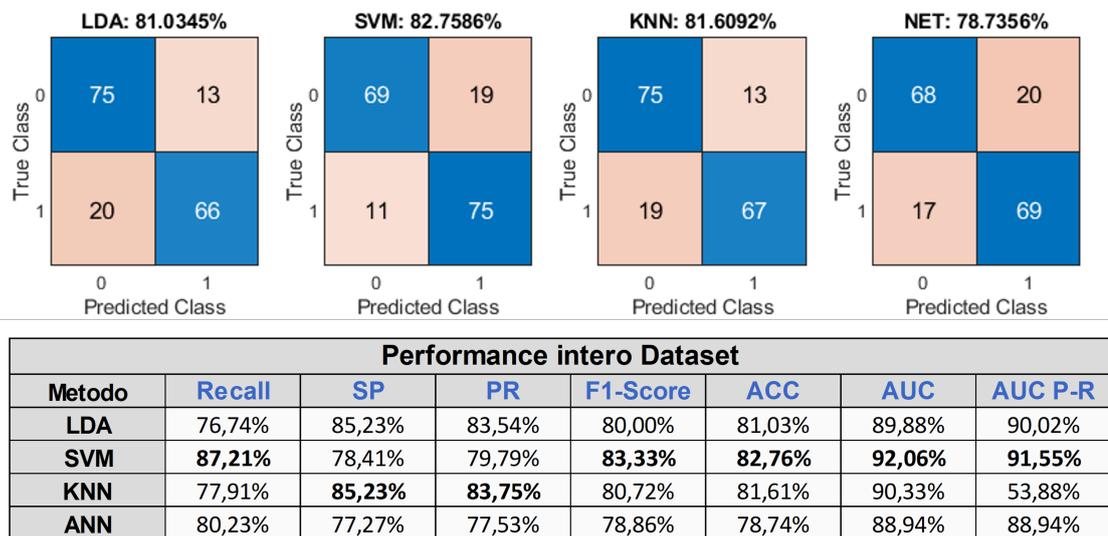


Figura 3.13: Confusion Matrix dei quattro metodi su immagini di tomosintesi e tabella riassuntiva dei valori ottenuti per le diverse metriche utilizzate. In grassetto i valori massimi ottenuti per ogni metrica.

Performance immagini proiezione CC							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	76,74%	86,36%	84,62%	80,49%	81,61%	89,64%	67,45%
SVM	86,05%	84,09%	84,09%	85,06%	85,06%	91,97%	67,87%
KNN	74,42%	84,09%	82,05%	78,05%	79,31%	90,88%	68,50%
ANN	79,07%	75,00%	75,56%	77,27%	77,01%	86,63%	66,84%

(a) Performance classificatori sulle sole immagini in proiezione CC

Performance immagini proiezione MLO							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	76,74%	84,09%	82,50%	79,52%	80,46%	90,33%	67,54%
SVM	88,37%	72,73%	76,00%	81,72%	80,46%	91,91%	67,91%
KNN	81,40%	86,36%	85,37%	83,33%	83,91%	90,09%	67,78%
ANN	81,40%	79,55%	79,55%	80,46%	80,46%	91,49%	67,79%

(b) Performance classificatori sulle sole immagini in proiezione MLO

Figura 3.14: Confronto delle performance ottenute su immagini di Tomosintesi dividendo le proiezioni e lavorando solamente con le proiezioni CC (a) o con le MLO (b). In grassetto i valori massimi ottenuti per ogni metrica.

Le percentuali di accuracy sono inferiori sulle immagini di tomosintesi, probabilmente a causa dell'underfitting dei modelli o per una caratterizzazione non ottimale delle opacità. Come precedentemente descritto, anche in questo caso sono state calcolate le metriche di validazione per i classificatori costruiti su immagini in una sola proiezione. A differenza di quanto accadeva con le immagini 2D sintetiche, in questo caso la differenza di prestazioni è meno marcata, con valori paragonabili nella maggior parte dei casi. L'ANN continua ad avere migliori performance sulle immagini in proiezione MLO, mentre per gli altri metodi i risultati si discostano da quelli ottenuti precedentemente. In ogni caso, la Recall è comunque più bassa sulle proiezioni CC. Si riconferma in ogni caso il trend visto anche sull'intero Dataset, con un abbassamento delle prestazioni generale. Il sistema è meno efficiente anche nella classificazione uniforme delle due proiezioni e le percentuali, ad eccezione dell'LDA, sono tutte maggiori o comparabili al caso precedente (Tabella 3.3).

	LDA	SVM	KNN	ANN
% Discordanza tra proiezioni (CC e MLO)	17,24%	18,39%	27,59%	19,54%

Tabella 3.3: Tomosintesi: percentuale di casi in cui le due proiezioni vengono assegnate a classi diverse dai diversi metodi di classificazione.

Se le prestazioni dei quattro metodi su immagini sintetiche presentavano andamenti leggermente diversi, in questo caso le curve ROC non mostrano sostanziali differenze così come i valori di AUC sono molto simili. Si nota invece che, mentre nella task di classificazione di immagini sintetiche il KNN era associato ad un valore basso di AUC, in questo caso non è il minimo ottenuto ma, nonostante questo, è ancora quello con valore di Recall associata inferiore. Per confronto tra le diverse strategie di calcolo dei parametri, si riportano le curve ROC del classificatore con prestazioni migliori, ossia l'SVM, al variare del numero di slices considerate (Figura 3.16) e i valori delle diverse metriche in tabella. La curva relativa alla strategia scelta si trova completamente al di sopra di quella che si ottiene mediando i parametri su 3 slices, ed è associata a migliori valori di Recall rispetto alle due curve restanti. All'aumentare del numero di slices diminuisce la variabilità dovuta alla scelta della key-slice ma, d'altra parte, aumenta l'effetto di sovrapposizione che se in fase di allenamento porta a risultati anche migliori, in fase di validazione comporta una perdita di capacità di generalizzazione e a risultati peggiori.

3.3.2 Risultati sul Validation

Anche nel caso delle immagini di tomosintesi, le performance sono state confrontate tramite le curve ROC per ognuno dei modelli ottimizzati nel 5-Fold cross validation e sulla base dei valori ottenuti. Le curve sono soggette a variabilità maggiore, rispecchiando il trend già osservato in fase di training, ma nonostante questo i valori di deviazione standard sono comunque tutti inferiori al 10%. La differenza più evidente è il drop delle prestazioni tra training, in azzurro, e validation in rosso che

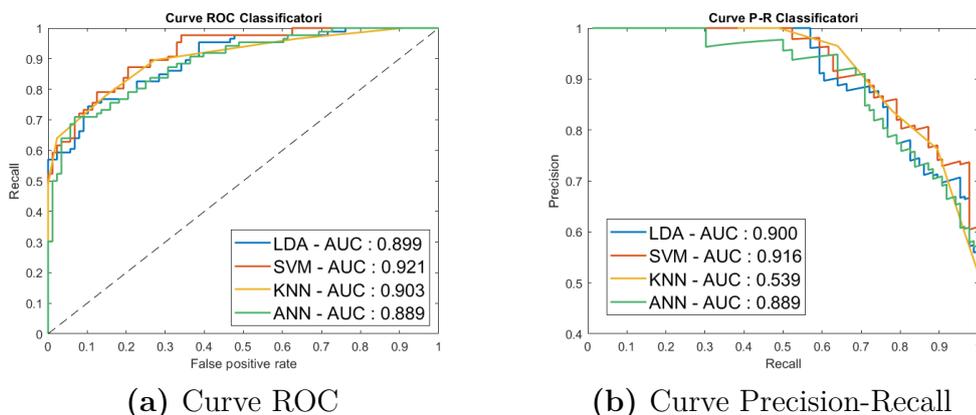
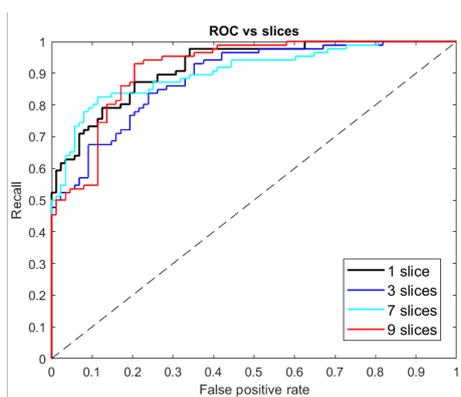


Figura 3.15: Curve di performance a confronto per i quattro metodi di classificazione su immagini di Tomosintesi.



Confronto SVM vs numero slice							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
1 SLICE	87,21%	78,41%	79,79%	83,33%	82,76%	92,06%	91,55%
3 SLICE	80,23%	77,27%	77,53%	78,86%	78,74%	88,70%	88,41%
7 SLICE	83,72%	80,68%	80,90%	82,29%	82,18%	90,43%	90,90%
9 SLICE	87,21%	79,55%	80,65%	83,80%	83,33%	91,69%	90,16%

Figura 3.16: Confronto delle strategie di calcolo dei parametri mediati su 3, 7 o 9 slices o estratti solo dalla key-slice. In grassetto i valori massimi ottenuti per ogni metrica.

suggerisce la ridotta capacità di generalizzazione per i metodi allenati su immagini di tomosintesi.

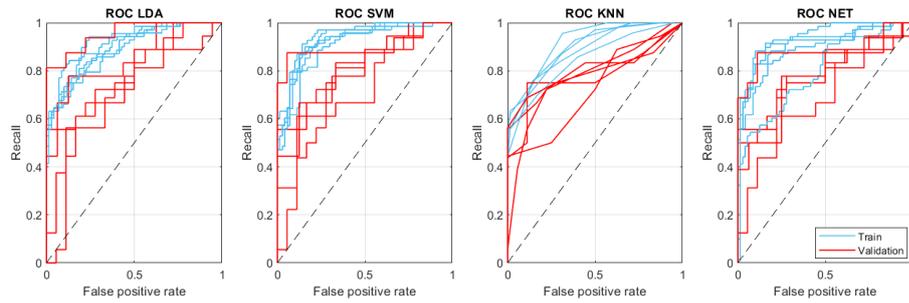


Figura 3.17: ROC sui sub-fold del 5 Fold Cross-Validation

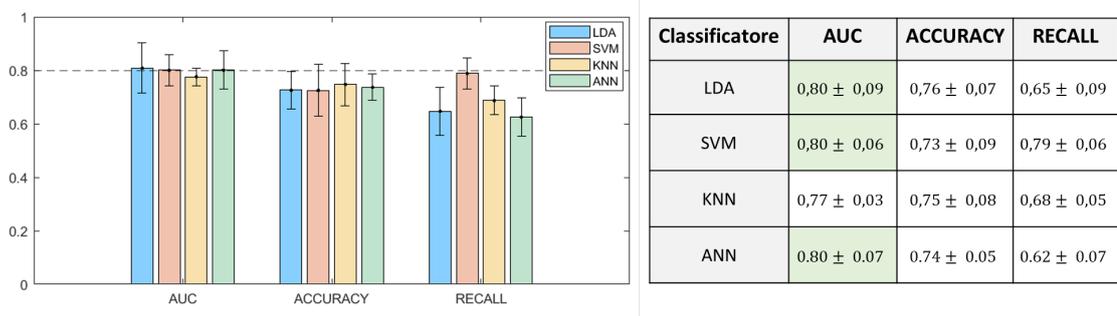


Figura 3.18: Confronto dei valori medi di AUC, Accuracy e Recall.

I valori di AUC sono tutti prossimi all'80% ma inferiori a quelli ottenuti su immagini sintetiche, e in generale si abbassano le prestazioni di tutti i metodi (Figura 3.17,3.18). Il motivo di questo peggioramento potrebbe essere la minore informazione delle singole slices di tomosintesi legata alla mancata sovrapposizione di tessuti che, sebbene renda più comprensibile l'immagine all'occhio dell'esperto, potrebbe non avere lo stesso effetto sugli algoritmi. I risultati subiscono variazioni poco significative al variare del numero di slices considerate per il calcolo dei parametri ma, sulla base dei valori medi di accuracy, recall e AUC, si nota come le prestazioni siano migliori quando viene considerata una sola slice. L'LDA è il classificatore che restituisce le migliori prestazioni in termini di valori medi dei parametri e, nel grafico, si nota un trend decrescente nei valori di accuracy all'aumentare del numero di slices. Particolarmente importante per la scelta è risultato il valore di Recall ottenibile, in generale migliore con la strategia scelta.

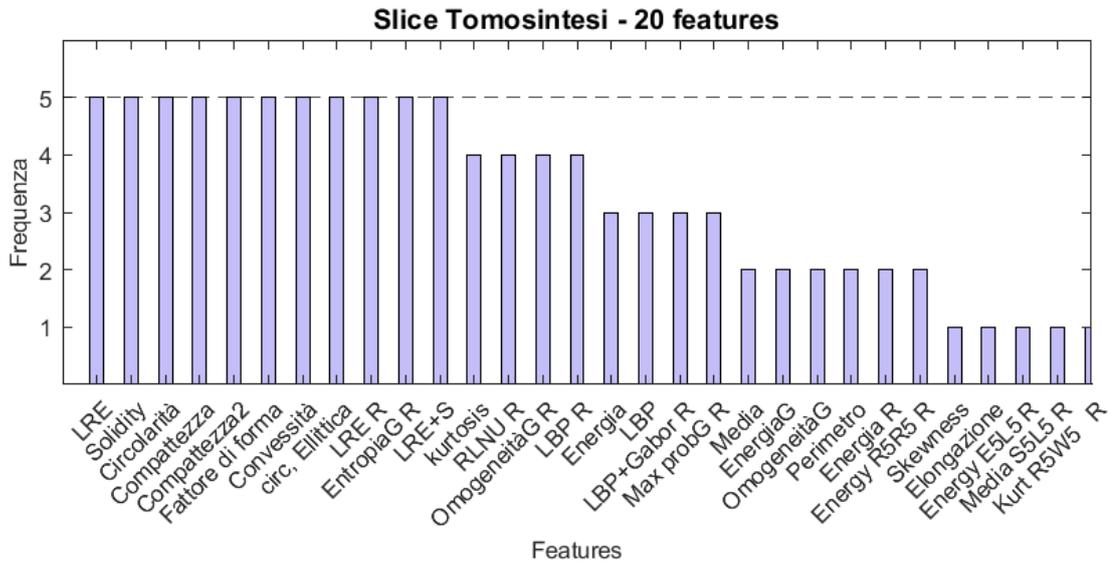


Figura 3.19: Features selection per la classificazione delle immagini BTO.

La Feature selection per le immagini di tomosintesi presenta alcune differenze rispetto a quella delle immagini 2D sintetiche. Se l'importanza dei descrittori di forma al fine della discriminazione tra classi si riconferma (Figura 3.19), i parametri di tessitura di Laws hanno valenza statistica decisamente inferiore. Questo risultato è legato al livello di complessità dell'immagine: la singola slice dell'immagine BTO ha una tessitura meno complessa dell'immagine 2D sintetica, la quale porta tutta l'informazione del tessuto sovrapposto alla lesione. È quindi plausibile che descrittori statistici come l'LRE e l'Entropia risultino selezionati con massima frequenza, seguono la Kurtosis e l'RLNU.

Un'ulteriore osservazione va fatta sul numero di features selezionate per uno solo dei sub-fold di validazione che, rispetto a quanto succedeva per le immagini 2D, è inferiore e suggerisce quindi una migliore caratterizzazione delle immagini e la conseguente migliore rappresentatività dei parametri calcolati.

3.4 Performance CNN

La CNN custom è stata allenata per 40 epoche e l'andamento dell'accuracy su Training e Validation è riportato nei grafici in Figura 3.20. L'andamento sul training è estremamente variabile soprattutto nella prima fase, in cui la curva di Validation segue l'andamento fino a quanto, a causa dell'overfitting, essa mantiene il trend di crescita ma rimane al di sotto di quella di Training. Lo stesso effetto è visibile nelle curve di Loss, dove la curva relativa al Validation non scende mai sotto lo 0.5. In particolare, anche i valori di Loss di partenza sono decisamente elevati. Per ottimizzare la fase di allenamento, il learning rate è stato diminuito gradualmente all'aumentare del numero di epoche per evitare quanto possibile l'overfitting, infatti entrambe le curve salgono molto lentamente. Allo stesso modo la struttura è stata modificata sulla base dell'osservazione delle prestazioni e con l'obiettivo di migliorarle. Nonostante il Data augmentation abbia quadruplicato il Dataset a disposizione passando da 174 a 696 immagini, con la CNN custom non si riescono ad ottenere risultati migliori di quelli riportati in tabella 3.4. I valori sono paragonabili a quelli ottenuti con i metodi di Machine learning, nonostante con i metodi di deep learning si possano raggiungere prestazioni decisamente migliori. Anche l'AUC calcolato dalla curva ROC sul Validation set è pari a 0,826 quindi paragonabile o inferiore ai risultati già ottenuti.

	Training	Validation
Accuracy	0,904	0,746
LOSS	0,341	0,536

Tabella 3.4: Performance della CNN su Training e Validation

Da questo test emergono alcune osservazioni: la numerosità del Dataset presenta anche in questo caso un grosso limite ed è quindi complesso indagare il funzionamento di una rete in modo esaustivo. D'altra parte, anche con riferimento alla letteratura, il numero di dati utilizzato per implementare reti CNN è decisamente maggiore e non

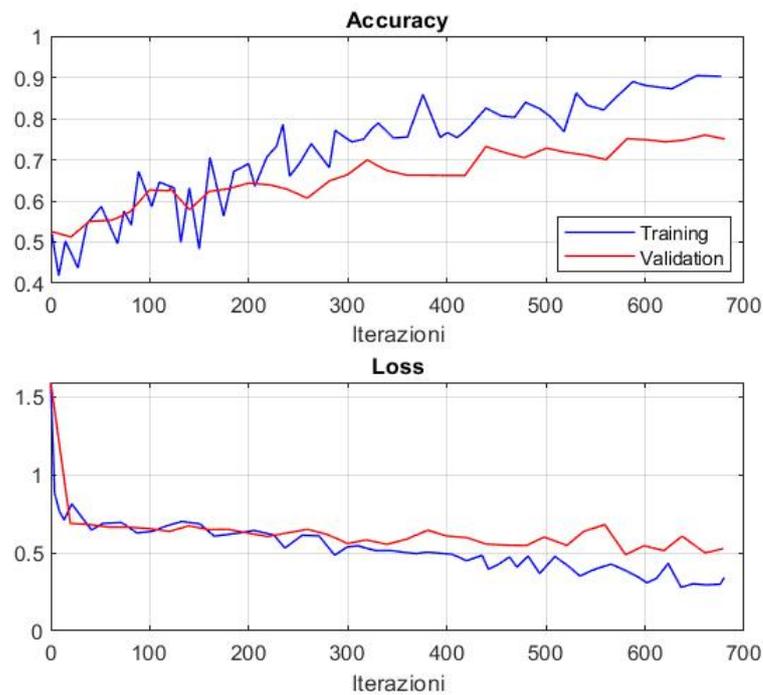


Figura 3.20: Andamento dell'Accuracy e della Loss in fase di allenamento della CNN.

replicabile solo per mezzo del data augmentation che introduce anche una notevole ridondanza di dati.

3.5 Integrazione e confronto delle due modalità

Le due tipologie di immagini presentano vantaggi complementari ed è stato dimostrato in letteratura come l'integrazione dell'informazione 2D e 3D della lesioni porti a risultati migliori dal punto di vista della caratterizzazione [25]. Per questo motivo è stato eseguito un ultimo test di integrazione delle due modalità, al fine di valutare eventuali miglioramenti nelle prestazioni. Per ogni ROI sono stati considerati tutti i parametri calcolati per le due tipologie e i più significativi sono stati utilizzati per la classificazione. A differenza dei casi precedenti, le

performance dei quattro metodi sono molto eterogenee, il che si nota facilmente dalle confusion matrix e dai valori riportati in tabella (Figura 3.21).

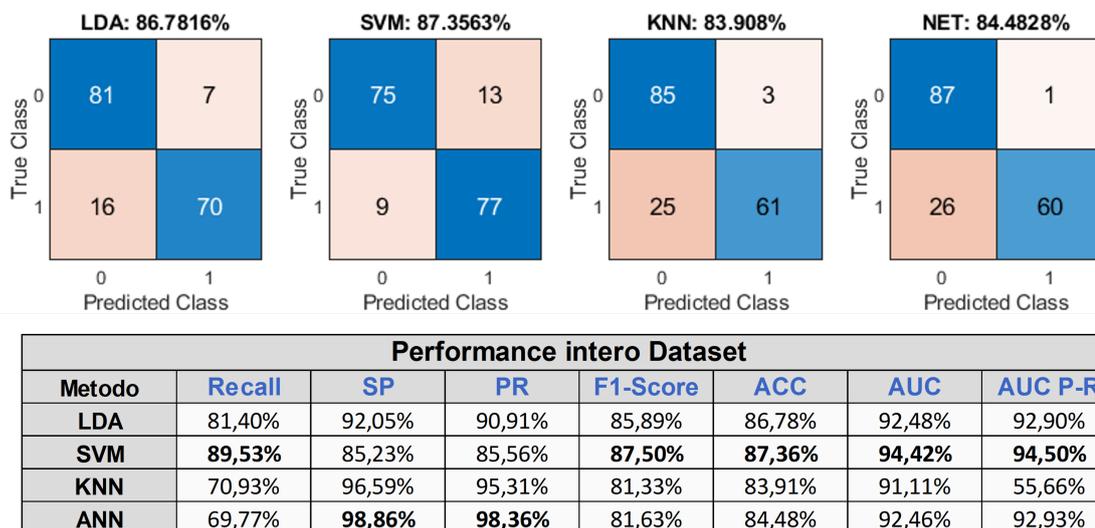


Figura 3.21: Test Multimodale : Confusion Matrix dei quattro metodi.

Performance immagini proiezione CC							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	79,07%	86,36%	85,00%	81,93%	82,76%	90,64%	67,55%
SVM	86,05%	86,36%	86,05%	86,05%	86,21%	92,71%	67,92%
KNN	69,77%	93,18%	90,91%	78,95%	81,61%	90,09%	67,98%
ANN	72,09%	97,73%	96,88%	82,67%	85,06%	93,82%	68,11%

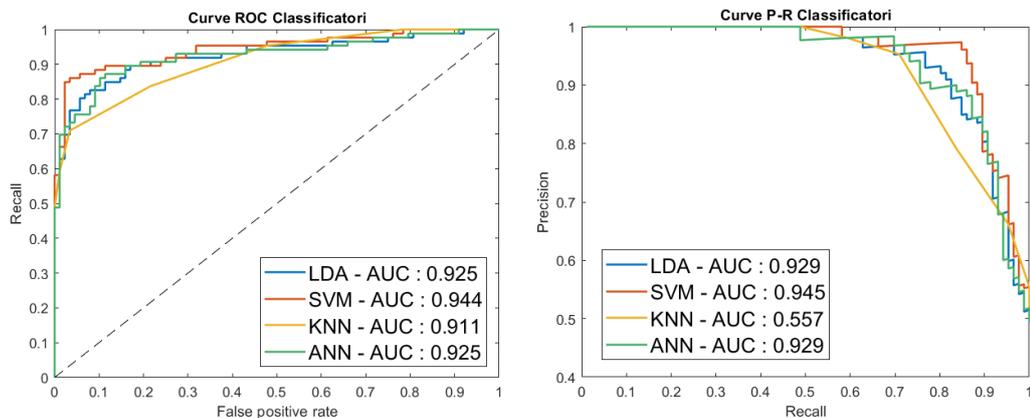
(a) Performance classificatori sulle sole immagini in proiezione CC

Performance immagini proiezione MLO							
Metodo	Recall	SP	PR	F1-Score	ACC	AUC	AUC P-R
LDA	83,72%	97,73%	97,30%	90,00%	90,80%	94,45%	68,06%
SVM	93,02%	84,09%	85,11%	88,89%	88,51%	95,98%	68,30%
KNN	72,09%	100,00%	100,00%	83,78%	86,21%	91,89%	68,27%
ANN	67,44%	100,00%	100,00%	80,56%	83,91%	91,07%	67,47%

(b) Performance classificatori sulle sole immagini in proiezione MLO

Figura 3.22: Confronto delle performance ottenute su immagini di Tomosintesi dividendo le proiezioni e lavorando solamente con le proiezioni CC (a) o con le MLO (b).

Il miglior valore di Recall si ottiene nuovamente con l'SVM, nonché la percentuale di corretti classificati maggiore. Da notare è anche l'ottimo valore di Specificity che si ottiene con l'ANN. Non cambia nemmeno l'incidenza di misclassificazione sulle due proiezioni distinte, anche se la differenza diventa meno significativa, probabilmente in correlazione al livello di dettaglio dei parametri descrittivi utilizzati. Si conferma nuovamente, sulla base dell'osservazione dei migliori risultati ottenibili, come in generale tutti i metodi risultino più performanti quando vengono utilizzate solo le proiezioni MLO. Risulta evidente dall'osservazione delle curve ROC (Figura 3.23) che le prestazioni sono comparabili a quelle raggiunte con le sole immagini C-View.



(a) Curve ROC Test Multimodale (b) Precision-Recall Curve Test Multimodale

Figura 3.23: Curve di performance a confronto dei quattro metodi per il Test Multimodale.

Per quanto riguarda i risultati sul Validation, tuttavia, emerge l'effetto della ridondanza dovuta alla parziale correlazione dei parametri calcolati sulle due immagini, in quanto si tratta comunque della stessa modalità di imaging e, questo, incide sui risultati (Figura 3.24). I valori di AUC sono tutti prossimi all'80% e nonostante il numero di elementi sia lo stesso del test fatto sulle immagini C-View, la variabilità è nettamente maggiore, mentre è paragonabile sui valori di Accuracy.

Anche la massima Recall ottenibile è decisamente inferiore al risultato già visto per le 2D sintetiche.

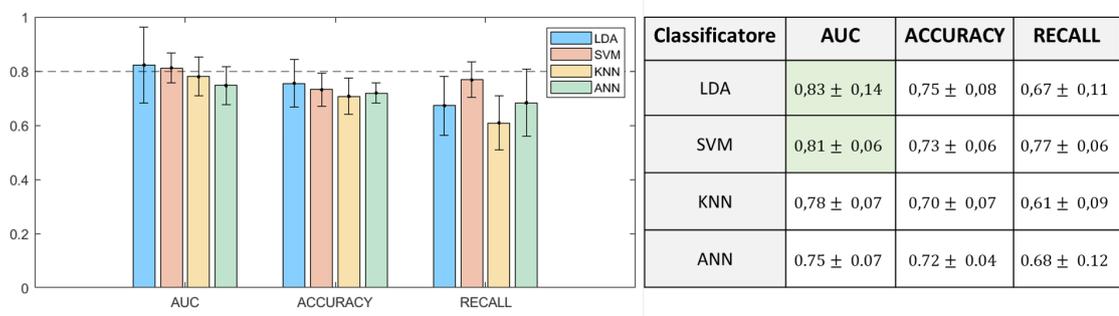


Figura 3.24: Test multimodale: Confronto dei valori medi di AUC, Accuracy e Recall.

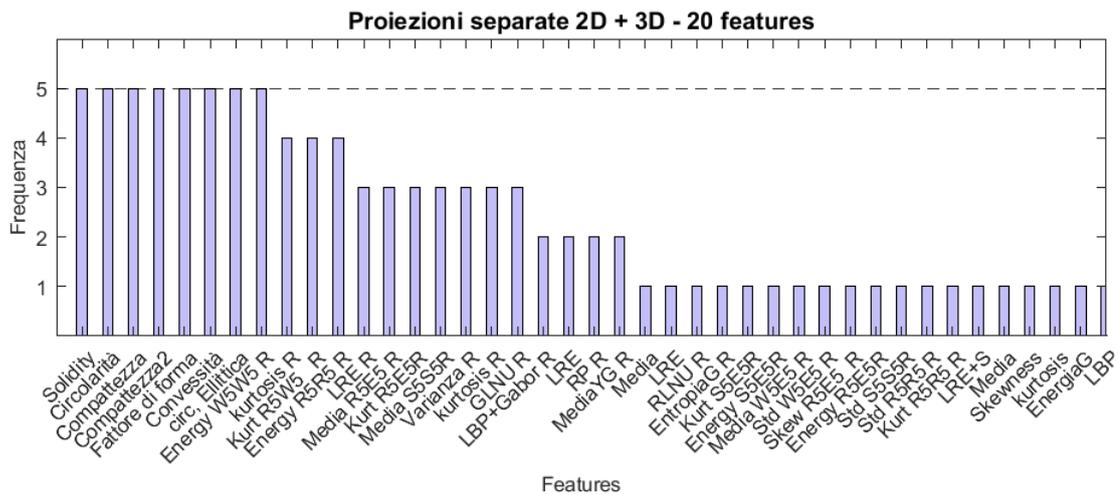


Figura 3.25: Feature Selection per la classificazione delle immagini in modalità multimodale.

In ultima analisi, anche le features significative sono ristrette ai descrittori di forma, i quali sono stati considerati comunque solo una volta. In generale, diminuisce la frequenza di selezione il che indica difficoltà di caratterizzazione delle lesioni: vengono considerati sia parametri di tessitura più complessi che parametri statistici del prim'ordine, ma la valenza statistica associata è nettamente inferiore (Figura 3.25).

Specialmente in questo caso, ma più in generale in tutte le prove, è frequente l'inserimento nel gruppo di features più significative di parametri calcolati come rapporto tra regione esterna e regione esterna: è un risultato positivo in quanto questi non dipendono direttamente dall'intensità della regione ma danno un'informazione relativa e non dipendono dal contorno manuale, che introduce variabilità nell'analisi.

È interessante notare come quest'ultimo test sia in assoluto il più paragonabile all'approccio utilizzato normalmente dal radiologo per la diagnosi della lesione, in quanto utilizza entrambe le tipologie di immagini sfruttandone le diverse caratteristiche. Nonostante questo, le prestazioni ottenute con l'algoritmo non sono migliori di quelle ottenute sulle singole modalità, il che sottolinea il divario tra il funzionamento della mente umana e l'apprendimento automatico. Analizzando tutti i risultati è evidente come le prestazioni sulle immagini C-View siano quasi nella totalità dei casi migliori rispetto a quelle sulle key-slice delle BTO (Figura 3.26) e, soprattutto, si ottengono con metodi di classificazione diversa.

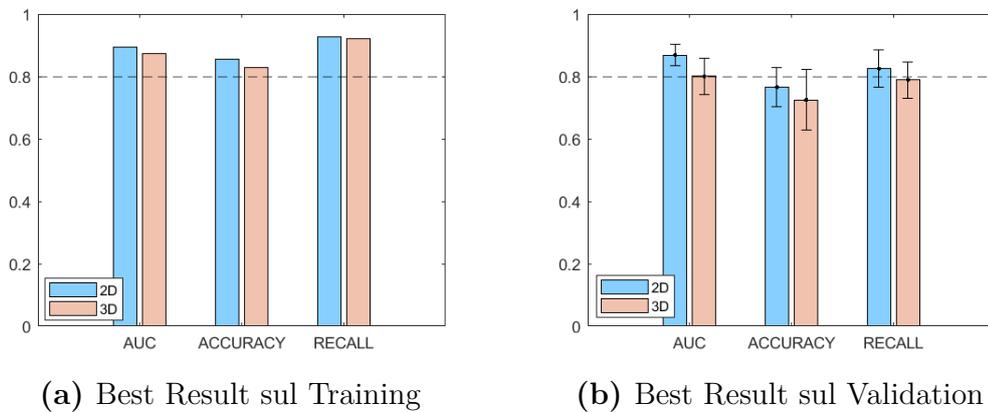


Figura 3.26: Risultati ottenuti in termini di AUC, Accuracy e Recall con il miglior classificatore, SVM, sul training e sul validation mettendo a confronto le performance con immagini 2D sintetiche e con le tomosintesi.

Considerando tutte le prove effettuate al variare delle strategie attuate, della tipologia di immagini e dei parametri utilizzati, sulla base

delle tre metriche utilizzate per il confronto ossia AUC, Accuracy e Recall, il metodo di classificazione migliore si rivela essere l'SVM con kernel lineare in quanto:

- Mantiene un buon livello di generalizzazione in tutte le prove effettuate, con una variabilità relativamente bassa;
- Presenta i migliori valori di AUC in tutti test e a prescindere dal tipo di immagini;
- Sia in fase di training che in fase di validation è il metodo che permette di ottenere il maggior valore di Recall, spesso maggiore dell'80%;

Nonostante la maggiore complessità delle immagini di mammografia 2D sintetica, esse si dimostrano più utili al fine della discriminazione tra lesioni maligne e benigne e permettono di ottenere maggiori livelli di sensibilità. Essendo la maschera manuale utilizzata la stessa per le due tipologie di immagini, ad incidere sono i parametri di tessitura.

In conclusione, la migliore curva ROC dell'SVM, ottenuta come media delle prestazioni raggiunte nella task di classificazione di immagini 2D sintetiche in fase di validazione, è stata confrontata con quella ottenuta come media delle performance di 101 radiologi su immagini di mammografia DM nello studio di A. Rodriguez-Ruiz, K. Lang, A. Gubern-Merida et al. pubblicato nel 2019 dalla Oxford University [61]. Non essendo le immagini utilizzate per questo studio risultati di esami di screening ma di approfondimenti in clinica, non è stato possibile ricavare una curva di performance delle senologhe sul Dataset in esame e, per questo motivo, nonostante non sia stata ottenuta sulla stessa tipologia di immagini, è stata presa come riferimento in quanto un buon risultato in letteratura.

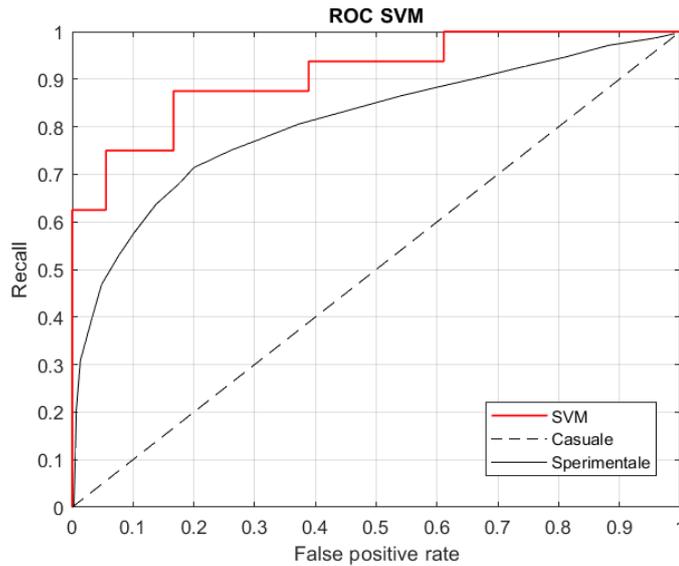


Figura 3.27: Confronto tra la curva AUC ottenuta sulle performance di 101 radiologi [61] su immagini DM e la curva ROC dell'SVM sul Validation set.

Il numero di immagini considerate nello studio è decisamente maggiore, infatti conta circa 2600 esami in un Dataset complessivo multi-centro e multi-macchinario. La curva è associata ad un valore di AUC pari a 0.814, mentre il valore ottenuto in questo studio è pari a $0,87 \pm 0,03$ quando le singole proiezioni sono state classificate separatamente, e una media di 0,89 quando unite. Infine, i risultati ottenuti integrando le due proiezioni sono promettenti ma ancora affetti da un'importante variabilità che, sulla base del trend osservato con il raddoppiamento del Dataset, può sicuramente essere migliorata aumentando il numero di casi.

Capitolo 4

Conclusioni e Sviluppi futuri

Lo studio condotto fa emergere importanti aspetti legati alla caratterizzazione automatica delle lesioni. I risultati migliori, in termini di variabilità e valore medio, si ottengono con l'SVM, che si rivela essere un metodo molto robusto, nonché quello con sensibilità più alta in quasi tutti i test. Il più alto valore medio di AUC ottenuto è pari all'87% ed è superiore a quello ottenuto da radiologi esperti su immagini di mammografia: nell'ipotesi di utilizzare un sistema del genere come aiuto alla decisione, avere prestazioni paragonabili o maggiori a quelle di un esperto che valuta l'immagine diagnostica è importante e può comportare una riduzione anche solo di una piccola percentuale del numero di biopsie. Le prestazioni dell'SVM non vengono raggiunte dalla rete CNN, nonostante le applicazioni basate sul deep learning siano in generale più performanti in quanto permettono di raggiungere un livello di astrazione decisamente maggiore. Questo risultato risente sicuramente delle due limitazioni principali dello studio: la numerosità dei dati e la casistica selezionata. Nonostante il Dataset utilizzato sia costituito da casi monocentrici e mono-macchinario, caratteristica che da una parte favorisce la CNN, include un'unica tipologia di lesioni il che probabilmente limita il livello di generalizzazione e la possibilità di rendere l'analisi più completa e incide concretamente sulla difficoltà

di raggiungere una numerosità adeguata. L'analisi di tessitura implementata mira a quantificare gli aspetti normalmente considerati a livello clinico, traducendo ad esempio l'intensità dell'opacità in media e media normalizzata della regione o identificando microcalcificazioni e spicule, oltre a cogliere pattern di tessitura più complessi. L'immagine 2D sintetica, che di fatto appiattisce l'informazione, viene classificata principalmente sulla base dei descrittori di forma che si rivelano essere molto significativi, mentre sulle singole slices della BTO, dove grazie all'assenza di sovrapposizione le caratteristiche del tessuto sono preservate, la correlazione con i parametri di tessitura è più rilevante. Il fatto che i risultati migliori si ottengano utilizzando solamente le immagini 2D sintetiche, rappresenta un aspetto molto positivo nell'ottica di ridurre i lunghi tempi di lettura che caratterizzano le immagini di tomosintesi: nella normale attività clinica, infatti, quando presenti, vengono consultate entrambe le immagini al fine della localizzazione e definizione della lesione, ma a livello di algoritmo l'analisi di una sola tipologia è sufficiente. Possibili sviluppi futuri vedono non solo l'aumento della numerosità ma, soprattutto, l'introduzione di casi diversi per garantire una rappresentatività maggiore. Tuttavia è complicato prevedere un miglioramento del sistema solo in base all'ampliamento del Dataset, in quanto, nonostante sia stato verificato l'effettivo incremento delle prestazioni all'aumentare delle immagini, è possibile che esista una soglia di performance superabile solo integrando informazioni che non appartengono all'immagine ma che siano dati anamnestici della paziente o il risultato di esami complementari. Per questo, dal punto di vista clinico, potrebbe essere interessante integrare queste informazioni e correlare le caratteristiche dell'immagine con esami di anatomia patologica. Dal punto di vista degli algoritmi, l'analisi può essere sicuramente migliorata con l'implementazione di algoritmi di deep learning e CNN che, con un Dataset adeguato, permettono di smarcarsi dalla segmentazione della lesione e di analizzare direttamente patch di tessuto restituendo un grado di malignità con accuratezza decisamente maggiore di quello che si ottiene con la sola analisi di tessitura.

Ringraziamenti

Vorrei ringraziare innanzi tutto il Professor Filippo Molinari per avermi concesso l'opportunità di lavorare a questo progetto di Tesi durante il quale ho ampliato le mie conoscenze, sono entrata in contatto con nuove realtà e ho lavorato su tematiche che mi appassionano molto.

Un grande ringraziamento va all'ingegner Massimo Salvi per l'infinita disponibilità, per i preziosi consigli e gli insegnamenti, e per essere stato un fondamentale punto di riferimento durante tutto il periodo di tesi. Ringrazio inoltre Matilde ed Enrico di TA per il supporto, la collaborazione e soprattutto la fiducia nel mio lavoro, e le dottoresse del reparto di Senologia delle Molinette per l'incredibile disponibilità e gentilezza il cui lavoro è stato fondamentale per portare avanti il progetto e mi ha permesso di imparare molto.

Alle mie compagne di magistrale, amiche, nonché Dream Team, Giulia, Laura e Michi per aver condiviso con me i momenti migliori e i peggiori, per avermi fatto ridere anche con l'ansia, perché insieme a voi lavorare è sempre stato più leggero. E aggiungo le mie compagne di sempre Ale, Giu, Sara, Marghe e Giusy. Grazie a tutte voi, per aver reso tutto più semplice, per la vostra amicizia e la vostra presenza, per lo studio insieme, per i pranzi sulle panchine, per i caffè dell'ultimo minuto, per la complicità e il sostegno, ma soprattutto per la forza che spesso mi avete dato. Ogni momento è impresso nella mia mente e sarà per sempre un ricordo meraviglioso e indelebile.

Agli amici di sempre, ma soprattutto, a Martina, che chiamerò così solo in questa occasione, perché con te ho condiviso tutto e come solo una sorella può fare, mi hai sempre sostenuta e hai creduto in me più di quanto a volte riuscissi a fare io.

A Nicola, grazie per essere stato la mia ancora e per avermi tenuta a galla, per avermi sostenuta e spronata a dare sempre il meglio e a cogliere le occasioni, per essere stato accanto a me ogni singolo istante di ogni singolo giorno senza lasciarmi mai, so per certo che non avrei fatto molte cose senza di te.

Infine, ma prima di tutti, ringrazio il mio papà al quale devo tutto. Perché non c'è nulla al mondo, per me, che significhi casa più di te. Se sono arrivata fino a qui è grazie a te e alla mamma, il mio angelo, la mia forza.

Bibliografia

- [1] L.Olivetti. *Professione TSRM*. Springer, 2013. Cap. 13.
- [2] R. Bellantone F. Basile. *Manuale di Semeiotica e Metodologia Chirurgica*. Edra, 2014. Cap. 12.
- [3] *The breasts*. Amboss. Apr. 2021. URL: <https://next.amboss.com/us>.
- [4] Adam W. Mitchell Richard L. Drake A. Wayne Vogl. *Anatomia del Gray*. Edra, 2015. Cap. 3.
- [5] Fondazione AIOM. «I numeri del cancro in Italia 2020». In: (2019). URL: <https://www.registri-tumori.it/cms/pubblicazioni/i-numeri-del-cancro-italia-2020>.
- [6] Osservatorio nazionale screening M.Zappa P.Mantellini. *Rapporto 2019*. 2019.
- [7] A. Goldhirsch et al. U. Veronesi P. Boyle. «Breast Cancer». In: *Seminar* 365.9472 (mag. 2005).
- [8] Pinku Mukherjee Shayan Shaghayeq Nazari. «An overview of mammographic density and its association with breast cancer». In: (2018).
- [9] *I geni Brca aumentano il rischio di tumore al seno. Non la mortalità*. Gen. 2018. URL: <https://www.fondazioneveronesi.it/magazine/articoli/oncologia>.
- [10] et al. Yi-Sheng Sun Zhao Zhao Zhang-Nv Yang. «Risk Factors and Preventions of Breast Cancer». In: *International Journal of Biological Sciences* 13.11 (2017), pp. 1387–1397.

-
- [11] Aiom. *Linee guida NEOPLASIE DELLA MAMMELLA*. Ott. 2019.
- [12] et al. Akram M. «Awareness and current knowledge of breast cancer». In: *Biological Research* 50.33 (2017).
- [13] G. Cserni. «Histological type and typing of breast carcinomas and the WHO classification changes over time». In: *Pathologica - Journal of the Italian Society of Anatomic Pathology and Diagnostic Cytopathology* 112.1 (mar. 2020).
- [14] «Benign Breast Diseases in Women». In: *Deutsches Arzteblatt international* 116.33 (2019), pp. 565–574.
- [15] *Mammografia*. Senologia diagnostica. URL: <https://senologiadagnostica.it/mammografia/>.
- [16] F. Squartini. *Anatomia patologica*. edizioni CEA, 1993.
- [17] S. Pacifici et al. A. Haouimi. *Digital breast tomosynthesis*. Radiopaedia. URL: <https://radiopaedia.org/articles/digital-breast-tomosynthesis>.
- [18] I. Sechopoulos. «A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications». In: *Medical Physics* 40.1 (gen. 2013).
- [19] S. Vedantham et al. P.S.Sujlana M. Mahesh. «Digital breast tomosynthesis: Image acquisition principles and artifacts». In: *Clinical Imaging* 55 (apr. 2018).
- [20] D.Dreizin et al. N.Tirada G.Li. «Digital Breast Tomosynthesis: Physics, Artifacts, and Quality Control Considerations». In: *Radiographics* 39.2 (mar. 2019).
- [21] L. Moy K.J.Geras R.M. Mann. «Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives». In: *Radiology* (2019).
- [22] A. Smith. «Synthesized 2D Mammographic Imaging Theory and Clinical Performance». In: (2016).

- [23] P.Fonio et al. G.Mariscotti M.Durando. «Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories». In: (ago. 2017).
- [24] A.S.Holen S.Hofvind T. Hovda. «Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program». In: *Radiology* 287.3 (giu. 2018), pp. 787–793.
- [25] M.Pellegrini D.Bernardi P.Macaskill. «Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study». In: *Lancet Oncol* (giu. 2016).
- [26] E.B.Eben et al. P. Skaane A. I. Bandos. «Two-View Digital Breast Tomosynthesis Screening with Synthetically Reconstructed Projection Images: Comparison with Digital Breast Tomosynthesis with Full-Field Digital Mammographic Images». In: *Radiology* 271.3 (giu. 2014).
- [27] C.S.Giess L. Ratanaprasatporn S.A.Chikarmane. «Strengths and Weaknesses of Synthetic Mammography in Screening». In: *RadioGraphics* 37.7 (nov. 2017), pp. 1913–1927.
- [28] Medixant. *RadiAnt DICOM Viewer*. Ver. 2020.2. 19 Lug. 2020. URL: <https://www.radiantviewer.com>.
- [29] A. Karellas et al. S.Vedantham L. Shi. «Semi-automated segmentation and classification of digital breast tomosynthesis reconstructed images». In: (gen. 2013).
- [30] B. Sahiner H. Chan Y. Wu. «Characterization of masses in digital breast tomosynthesis: Comparison of machine learning in projection views and reconstructed slices». In: *Medical Physics* 37 (lug. 2010).
- [31] R. Ammar et al. D. Abdelhafiz C. Yang. «Deep convolutional neural networks for mammography: advances, challenges and applications». In: *BMC Bioinformatics* 20 (ott. 2017).

- [32] A. Jain D.Lévy. «Breast Mass classification from mammograms using deep convolutional Neural Networks». In: (dic. 2016).
- [33] S.Ertekin et al. E.Sert. «Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms». In: (2017).
- [34] S. F. Ershad L. Armi. «Texture image analysis and texture classification methods - a review». In: *International Online Journal of Image Processing and Pattern Recognition* 2.1 (2019).
- [35] Wout Oude Elferink. *Gray Level Run Length Image Statistics*. Ver. 2020.2. URL: <https://www.mathworks.com/matlabcentral/fileexchange/52640-gray-level-run-length-image-statistics>.
- [36] Michael A. Wirth. *Texture Analysis*. University Lecture. 2004.
- [37] A. Rosebrock. *Local Binary Patterns with Python OpenCV*. pyimagesearch. Dic. 2015. URL: <https://www.pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/>.
- [38] Mäenpää T. Ojala T. Pietikäinen M. «Gray scale and rotation invariant texture classification with local binary patterns». In: *Proceedings of the European Conference on Computer Vision* (lug. 2000).
- [39] Zwiggelaar R. Oliver A. Freixenet J. «Automatic classification of breast density». In: *Proceedings of the IEEE International Conference on Image Processing 2* (set. 2005).
- [40] Y. Purnama A.Setiawan J. Wesley. «Mammogram Classification using Law's Texture Energy Measure and Neural Networks». In: *Procedia Computer Science* 59 (2015).
- [41] C.Y. Suen M.Yousefi A. Krzyzak. «Mass detection in digital breast tomosynthesis data using convolutional networks and multiple instance learning». In: *Computers in Biology and Medicine* 96 (apr. 2018).

- [42] M. Matsui et al. A. Sakai Y. Onishi. «A method for the automated classification of benign and malignant masses on digital breast tomosynthesis images using machine learning and radiomic features». In: *Radiological Physics and Technology* 13 (mag. 2019).
- [43] Shawhin Talebi. *The Wavelet Transform*. Towards Data Science. URL: <https://towardsdatascience.com/the-wavelet-transform-e9cfa85d7b34>.
- [44] C. K. Chua et al. K. Ganesan U. R. Acharya. «Automated Diagnosis of Mammogram Images of Breast Cancer Using Discrete Wavelet Transform and Spherical Wavelet Transform Features: A Comparative Study». In: *Technology in Cancer Research and Treatment* 13.6 (dic. 2014).
- [45] J.H. Lopez W.G. Flores. «Assessment of the invariance and discriminant power of morphological features under geometric transformations for breast tumor classification». In: *Computer Methods and Programs in Biomedicine* 185 (2020).
- [46] J. Tennenbaum. *Analysis of variance and its variations*. Towards data science. Ott. 2019. URL: <https://towardsdatascience.com/analysis-of-variance-and-its-variations-6ef3f8fbeb05>.
- [47] C.D.Costa. *A tour of Machine Learning Algorithms*. Towards Data Science. Mag. 2020.
- [48] Z. Akkus et al. B.Erickson P.Korfiatis. «Machine learning for medical imaging». In: *Radiographics* 37 (mar. 2017).
- [49] MATLAB. *Creating Discriminant Analysis Model*. URL: <https://www.mathworks.com/help/stats/creating-discriminant-analysis-model.html>.
- [50] Y. Xiaozhou. *Linear Discriminant Analysis Explained*. Towards data science. Mag. 2020. URL: <https://towardsdatascience.com/linear-discriminant-analysis-explained-f88be6c1e00b>.

- [51] MATLAB. *Support vector machine (SVM) for one-class and binary classification*. URL: <https://www.mathworks.com/help/stats/classificationsvm.html>.
- [52] S.Periaswamy et al. E.F.Conant A.Y.Toledano. «Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis». In: *Radiology:Artificial Intelligence* (2019).
- [53] J. Park et al. N. Wu J. Phang. «Deep neural networks improve radiologists' performance in breast cancer screening». In: (mar. 2019).
- [54] S.Saha. *A comprehensive Guide to Convolutional Neural Networks*. URL: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- [55] Q.He et al. X.Li G.Qin. «Digital breast tomosynthesis versus mammography: integration of image modalities enhances deep learning-based breast mass classification». In: (nov. 2019).
- [56] D. Sheth et al. K. Mendel H. Li. «Transfer Learning From Convolutional Neural Networks for Computer-Aided Diagnosis:A Comparison of Digital Breast Tomosynthesis and Full-Field Digital Mammography». In: (giu. 2018).
- [57] J.Brownlee. *A gentle introduction to k-fold Cross-Validation*. Machine Learning Mastery. URL: <https://machinelearningmastery.com/k-fold-cross-validation/>.
- [58] *Confusion Matrix*. Comprehensive Analytical Chemistry. 2018. URL: <https://www.sciencedirect.com/topics/engineering/confusion-matrix>.
- [59] MedCalc. *ROC curve analysis*. URL: <https://www.medcalc.org/manual/roc-curves.php>.
- [60] D. Steen. *Precision-Recall curves*. URL: <https://medium.com/@douglassteen/precision-recall-curves-d32e5b290248>.

- [61] A. G. Merida et al. A. R. Ruiz K. Lang. «Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists». In: *JNCI J Natl Cancer Inst* (mar. 2019).