

# POLITECNICO DI TORINO

Corso di Laurea Magistrale  
in Ingegneria Biomedica

Tesi di Laurea Magistrale

## Sviluppo di un sistema ibrido basato su Machine Learning e Deep Learning per la classificazione di lesioni tumorali in immagini di mammografia sintetica



### **Relatori**

Prof. Massimo Salvi  
Prof. Filippo Molinari

### **Candidato**

Francesca Pastore

Anno Accademico 2021-2022



# Sommario

Secondo i dati AIOM (associazione italiana oncologi medici) il tumore alla mammella è la patologia neoplastica a più alta prevalenza nelle donne, con una stima di 55 mila nuovi casi nel 2020 [1]. In Italia, questo tumore rappresenta più del 30% di tutte le nuove diagnosi di cancro e provoca il 17% di tutti i decessi oncologici nel sesso femminile. Nonostante ciò, negli ultimi anni il tasso di mortalità è in continua diminuzione, principalmente grazie ad una diffusione sempre maggiore dei programmi di screening. Questi programmi permettono di identificare i tumori prima dell'inizio dei sintomi, quando sono ancora in uno stadio iniziale, il che permette di limitare i trattamenti ed aumentare i tassi di sopravvivenza. La probabilità di guarigione, infatti, è tanto più alta quanto più la diagnosi è anticipata. Per questo motivo, la ricerca spinge a trovare soluzioni sempre più innovative per individuare precocemente le neoplasie.

I metodi classici per diagnosticare il cancro si basano sulla mammografia, ovvero un esame radiologico che indaga i tessuti interni della mammella al fine di rilevare lesioni tumorali ancor prima che possano essere palpabili [2]. L'esame mammografico presenta, tuttavia, non poche limitazioni, tra cui una scarsa sensibilità e il problema dei falsi positivi. La sovrapposizione del tessuto mammario, infatti, o la densità parenchimale possono oscurare i tumori o generare artefatti con proprietà simili a vere neoplasie determinando una successiva diagnosi errata. Per ridurre questo problema si è introdotto un secondo metodo di screening basato sull'esame diagnostico di tomosintesi digitale (DBT) cioè uno strumento diagnostico che si basa sull'acquisizione di immagini della mammella a differenti angolazioni. Un maggior numero di proiezioni consente di analizzare l'intero volume su piani a diverse profondità, limitando l'effetto di sovrapposizione dei tessuti [3]. Tutto questo si traduce, da un lato, in una minore possibilità di generare una diagnosi errata, ma, dall'altro, in un maggior volume di dati da analizzare, portando a costi aggiuntivi e un maggiore stress lavorativo ai medici. In questo contesto, si sono inseriti nuovi sistemi automatici basati su algoritmi di intelligenza artificiale che sono in grado non solo di sostituire e automatizzare le operazioni più standardizzate e ripetitive, ma anche di effettuare delle previsioni utili ai radiologi come supporto decisionale durante l'analisi dell'immagine. Sebbene non possano sostituire l'esperienza degli esperti, questi algoritmi, sembrano aver contribuito in modo significativo a soddisfare la domanda crescente di mammografie di screening, contribuendo efficacemente a ridurre la mortalità per cancro alla mammella [4].

Visto lo sviluppo sempre maggiore di queste tecnologie, il lavoro di tesi propone

la realizzazione un software automatico per il riconoscimento e la classificazione di lesioni tumorali per supportare il medico radiologo nell'analisi delle immagini.

Il primo capitolo, con un breve riferimento all'anatomia, illustra le principali caratteristiche del tumore alla mammella e le tecniche di imaging adottate per la diagnosi, il capitolo seguente descrive, con un approfondimento delle tecniche di intelligenza artificiale, gli algoritmi di deep learning impiegati nella clinica. Successivamente, nel capitolo "Materiali e Metodi", vengono descritti il dataset, i vari passi per l'estrazione di caratteristiche dalle immagini e gli algoritmi utilizzati per la classificazione. Infine, negli ultimi due capitoli, verranno illustrate tutte le prove effettuate e le analisi dei risultati ottenuti e relative "conclusioni e prospettive future" di ricerca nel campo.

# Indice

<b>Elenco delle tabelle</b>	7
<b>Elenco delle figure</b>	8
<b>1 Il tumore alla mammella e la sua diagnosi</b>	11
1.1 Richiami di anatomia	11
1.2 Cos'è un tumore alla mammella	12
1.3 Origine, cause e fattori di rischio	14
1.4 Numeri del cancro	15
1.5 Prevenzione e diagnosi	15
1.6 Mammografia digitale	16
1.7 Tomosintesi	18
1.8 Mammografia sintetica	20
<b>2 Apprendimento automatico</b>	23
2.1 Stato dell'arte	23
2.2 Machine learning e Deep learning	24
2.3 Reti neurali	25
2.3.1 Modello matematico del neurone	26
2.3.2 Architettura della rete	28
2.3.3 Addestramento della rete	30
2.4 Reti convoluzionali	31
2.4.1 Architettura di una CNN	31
<b>3 Materiali e metodi</b>	37
3.1 Dataset	37
3.2 Pipeline	38
3.3 Pre-processing	39
3.4 Estrazione di feature locali	41
3.4.1 Ridimensionamento delle immagini	41
3.4.2 Costruzione e divisione del dataset	42
3.4.3 Transfer learning	43
3.4.4 Architettura delle reti	44
3.4.5 Tuning dei parametri	47
3.4.6 Addestramento della rete e overfitting	48

3.5	Estrazione di feature globali . . . . .	51
3.5.1	Feature estratte . . . . .	51
3.5.2	Descrittori del primo ordine . . . . .	51
3.5.3	Descrittori di Haralick e Galloway . . . . .	52
3.5.4	Local binary patterns . . . . .	53
3.6	Feature selection . . . . .	53
3.7	Metodi di classificazione . . . . .	55
3.7.1	ANN . . . . .	55
3.7.2	SVM . . . . .	55
3.7.3	KNN . . . . .	56
3.7.4	LDA . . . . .	57
3.8	Metodi di valutazione . . . . .	58
3.8.1	Confusion matrix . . . . .	58
3.8.2	Curva ROC e Area Under Curve (AUC) . . . . .	60
3.8.3	k-fold . . . . .	61
<b>4</b>	<b>Risultati</b> . . . . .	<b>63</b>
4.1	Classificazione sulla base delle feature locali . . . . .	64
4.1.1	Batch size . . . . .	64
4.1.2	Data augmentation . . . . .	66
4.1.3	Learning Rate e Regolarizzazione . . . . .	67
4.1.4	Cross-validazione . . . . .	69
4.2	Classificazione sulla base delle feature concatenate . . . . .	71
4.2.1	Feature selection . . . . .	72
4.3	Confronto tra classificazioni . . . . .	73
4.4	Analisi delle misclassificazioni . . . . .	74
<b>5</b>	<b>Conclusioni e Sviluppi futuri</b> . . . . .	<b>77</b>

# Elenco delle tabelle

3.1	Descrittori di I° ordine . . . . .	52
3.2	Descrittori di II° ordine . . . . .	53
4.1	Performance del classificatore migliore . . . . .	74

# Elenco delle figure

1.1	Anatomia della struttura interna della mammella [5] . . . . .	12
1.2	Confronto tra un tumore benigno a sinistra e uno maligno a destra . . . . .	13
1.3	Struttura del dispositivo mammografico e posizionamento della paziente [6] . . . . .	17
1.4	Dispositivo di tomosintesi digitale mammaria. . . . .	19
1.5	Migliore individuazione delle lesioni grazie alla limitata sovrapposizione di esse nell'immagine di tomosintesi [7] . . . . .	20
1.6	Confronto tra immagini ottenute mediante mammografia, tomosintesi e ricostruzione C-View [8] . . . . .	21
2.1	Struttura del neurone biologico [9] . . . . .	26
2.2	Modello del neurone artificiale [10] . . . . .	27
2.3	Modello del neurone con Delta Rule [10] . . . . .	28
2.4	Funzioni di attivazioni: a sinistra la funzione sigmoideale, in centro quella a tangente iperbolica, a destra la ReLu [10] . . . . .	28
2.5	Architettura a strati di una semplice rete neurale [11] . . . . .	29
2.6	Struttura della retro-propagazione degli errori [12] . . . . .	30
2.7	Struttura di una rete neurale convoluzionale [13] . . . . .	32
2.8	Esempio di filtro convoluzionale e della sua applicazione su un'immagine [14] . . . . .	32
2.9	Esempio della convoluzione tra la matrice di input e un kernel 3x3 . . . . .	33
2.10	Relu function . . . . .	34
2.11	Esempio di riduzione della dimensione spaziale di una matrice tramite max pooling. . . . .	34
2.12	Esempio di rete neurale fully connected con 10 classi di output . . . . .	35
3.1	Esempio di due proiezioni (CC a sinistra e MLO a destra) in cui si evidenzia la lesione attraverso il contorno rosso posizionato dalle senologhe . . . . .	38
3.2	Flow-chart del progetto . . . . .	39
3.3	Istogramma dei valori di PixelScaling e relativo valore mediano . . . . .	40
3.4	Pre-processing applicato all'immagine per individuare la lesione . . . . .	40
3.5	Pre-processing applicato all'immagine per individuare la mammella . . . . .	41
3.6	Istogramma delle dimensioni delle bounding-box delle lesioni. Si raffigura anche la dimensione media e quella selezionata. . . . .	41
3.7	Flow-chart riguardante l'utilizzo dei dataset durante l'addestramento e la validazione dei modelli . . . . .	43

3.8	Percentuali in cui si è diviso il dataset a sinistra e numero di lesioni per ogni set a destra. . . . .	43
3.9	Esempio di Transfer Learning applicato al dataset di questo progetto . . . . .	44
3.10	Schema dell'architettura AlexNet . . . . .	44
3.11	Schema dell'architettura VGG . . . . .	45
3.12	Schema dell'architettura ResNet . . . . .	46
3.13	Schema di un blocco dell'architettura DenseNet . . . . .	46
3.14	Esempio di overfitting della rete . . . . .	48
3.15	Esempio di data augmentation offline e online . . . . .	49
3.16	Esempio di rete senza applicazione di dropout (a) e con applicazione di dropout (b) . . . . .	51
3.17	Matrice GLCM . . . . .	52
3.18	Descrittori LBP . . . . .	53
3.19	Struttura dell'algoritmo Random Forest . . . . .	54
3.20	Esempio di divisione in due classi attraverso il classificatore SVM . . . . .	56
3.21	Esempio di classificazione di un nuovo elemento attraverso il metodo di classificazione k-NN . . . . .	56
3.22	Esempio di divisione in tre classi attraverso il classificatore LDA . . . . .	57
3.23	Confusion matrix LBP . . . . .	58
3.24	Esempi di curva ROC: nel primo grafico è rappresentata la retta a scelta casuale, nel secondo grafico è rappresentata una generica curva ROC, nel terzo grafico è rappresentata una curva ROC associata ad un classificatore perfetto . . . . .	60
3.25	Schema riguardante la suddivisione del dataset durante l'applicazione del metodo k-fold . . . . .	61
4.1	Monitoraggio di accuratezza e loss function durante l'allenamento . . . . .	64
4.2	Accuratezza e AUC mediati su tutte le architetture per i tre set di dati e per i tre batch size scelti . . . . .	65
4.3	Confronto in termini di accuratezza, precisione e AUC tra le diverse architetture quando si utilizza un batch size pari a 32 . . . . .	65
4.4	Accuratezza e AUC mediati su tutte le architetture per i tre set di dati e per i tre data augmentation utilizzati . . . . .	66
4.5	Confronto in termini di accuratezza, precisione e AUC tra le diverse architetture quando si effettua data augmentation offline . . . . .	67
4.6	Accuratezza e AUC mediati su tutte le architetture per i tre valori di learning rate utilizzati . . . . .	68
4.7	Accuratezza e AUC mediati su tutte le architetture per i tre valori di lambda utilizzati . . . . .	68
4.8	Confronto in termini di accuratezza, precisione e AUC tra le diverse architetture quando si utilizza un learning rate pari a 0,0005 e un lambda pari a $10^{-5}$ . . . . .	69
4.9	Confronto in termini di accuratezza, recall, precisione e AUC tra le diverse architetture mediando le performance ottenute su ogni fold . . . . .	70
4.10	Accuratezza sui diversi set di dati per ogni combinazione di vettore di feature concatenato e classificatore utilizzata . . . . .	72

4.11	Accuratezza sui diversi set di dati per ogni combinazione di vettore di feature concatenato e classificatore utilizzata . . . . .	73
4.12	Accuratezza sul Test set mediata sui cinque fold nel caso in cui si classifichino solo le feature locali o nel caso in cui si classifichi la concatenazione di feature locali e globali . . . . .	73

# Capitolo 1

## Il tumore alla mammella e la sua diagnosi

La conoscenza generale dell'anatomia del seno e quella delle principali caratteristiche associate al tumore sono essenziali per poter trattare nel dettaglio gli argomenti successivi. Va notato che la discussione non vuole essere in alcun modo esaustiva, poiché una buona descrizione richiederebbe più spazio dell'intera tesi.

### 1.1 Richiami di anatomia

La mammella è un organo ghiandolare, pari e simmetrico, collocato nella parte anteriore del torace davanti al principale muscolo toracico, il gran pettorale [15]. L'organo è presente sia nel corpo femminile che in quello maschile anche se con volume e funzione diversa. Macroscopicamente ciascuna mammella presenta tre componenti principali, quali: la ghiandola mammaria (circondata da uno stroma di tessuto connettivo e da tessuto adiposo), la cute ed il complesso areola capezzolo [16].

La ghiandola è suddivisa in 15-20 strutture ghiandolari indipendenti fra loro, dette lobi, che a loro volta sono costituite da 20-40 unità elementari dette lobuli. Ogni lobulo contiene gruppi di piccole ghiandole capaci di secernere il latte che viene successivamente trasportato dal lobulo stesso al capezzolo attraverso una rete di canali chiamati dotti galattofori. In sintesi, si può pensare alla ghiandola mammaria come un grappolo d'uva formato dai lobuli che rappresentano gli acini all'interno dei quali avviene la produzione di latte materno e dai lobi che rappresentano il raspo, cioè il collegamento tra acini e ramificazioni. È importante conoscere la distinzione tra lobuli e dotti perché entrambe le strutture possono essere colpite da neoplasie che avranno prognosi e comportamento differente [3].

La ghiandola è circondata da una fitta rete di tessuto connettivo comprendente arterie, vene, nervi e vasi linfatici. Quest'ultimi sono dei piccoli canali che conducono dalla ghiandola ai linfonodi (piccole strutture nodulari situate lungo i vasi) un liquido, detto linfa, che contiene sostanze di rifiuto [16]. I linfonodi permettono di

filtrare la linfa bloccando batteri, cellule cancerose e altre sostanze nocive. Per questo motivo il sistema di drenaggio linfatico descrive il principale mezzo di diffusione metastatica del tumore, rivestendo particolare importanza nell'evoluzione della malattia. La stazione di filtraggio della mammella si posiziona nel cavo ascellare ed è in questo luogo che, come conseguenza della comparsa di un tumore al seno, compaiono normalmente le prime metastasi linfonodali. Lo studio e il controllo dei linfonodi del cavo ascellare risultano, pertanto, di grande interesse per la valutazione dello stadio del tumore e in particolare per capire se le cellule tumorali sono già in fase metastatica e stanno diffondendo nel corpo [17].

La mammella è composta anche da tessuto adiposo che è posto intorno e tra i lobuli e colma il vuoto tra il tessuto connettivo, il tessuto ghiandolare e i dotti. A seconda della sua collocazione e quantità, determinata dalla costituzione fisica individuale, dipendono dimensione e forma esterna della mammella [15].

Ogni mammella presenta alla sua estremità il capezzolo, cioè un rilievo di aspetto conico, più o meno pronunciato. Esso è formato dalla parte terminale dei dotti galattofori e da tessuto connettivo, ed è costituito da tessuto muscolare liscio che disposto circolarmente e radialmente è responsabile della sua erezione. Attorno al capezzolo è presente l'areola mammaria, una superficie circolare di cute caratterizzata da prominenze dovute alle sottostanti ghiandole del Montgomery che facilitano la suzione del neonato e donano morbidezza ed elasticità al capezzolo durante l'allattamento.

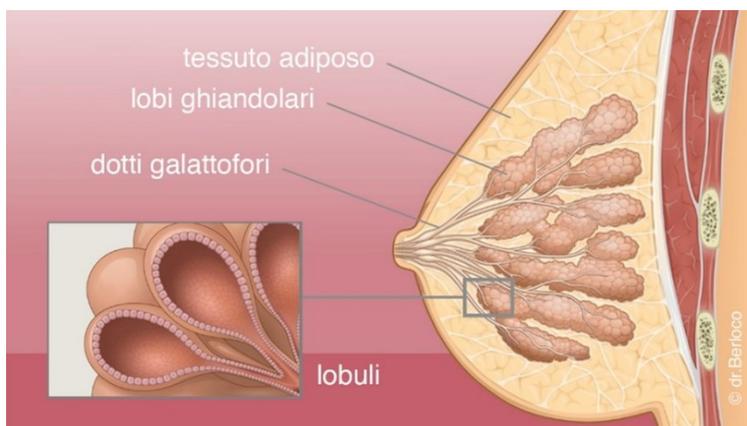


Figura 1.1. Anatomia della struttura interna della mammella [5]

## 1.2 Cos'è un tumore alla mammella

Il tumore alla mammella è una malattia dovuta alla moltiplicazione incontrollata delle cellule del seno. A seconda del comportamento biologico di tali cellule, esistono due forme di tumore: benigno e maligno. I tumori benigni sono formati da cellule che conservano parzialmente la loro morfologia e funzione, sebbene continuino a moltiplicarsi. Sono caratterizzati da un allargamento chiamato espansivo, poiché

cregono fino a formare una massa che comprime i tessuti adiacenti senza annientarli. Si presentano come masse ben definite, che mantengono le caratteristiche del tessuto da cui hanno avuto origine e non invadono né diffondono in altre zone del corpo. I tumori maligni, detti anche cancro, invece sono caratterizzati da morfologia e funzione differente rispetto a quella associata ai tessuti sani. Durante la loro riproduzione, formano prolungamenti che infiltrano i tessuti vicini, avvinghiando le cellule normali e distruggendole, fenomeno noto come "invasività neoplastica". Inoltre, i tumori maligni a differenza di quelli benigni, anche se rimossi potrebbero ricomparire.

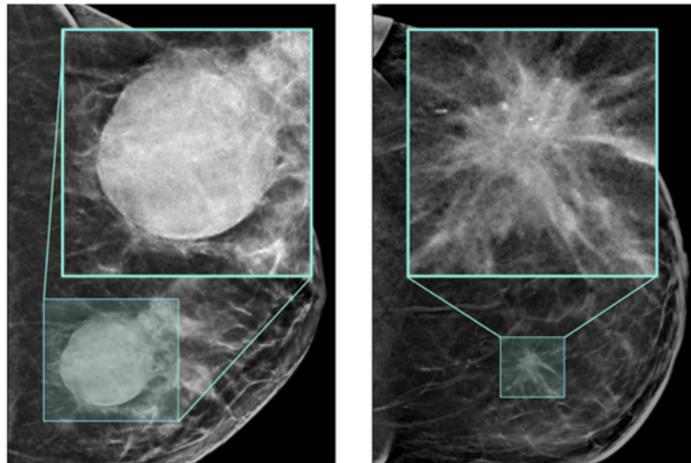


Figura 1.2. Confronto tra un tumore benigno a sinistra e uno maligno a destra

I tumori mammari possono essere classificati anche in base alla loro morfologia cioè a seconda del tipo di cellule che danno origine al tumore. Secondo questa suddivisione, la maggior parte dei casi viene identificata come carcinoma, ovvero un tumore che origina da cellule epiteliali [18].

Se invece si classificano i tumori in base alla loro capacità di infiltrarsi nei tessuti circostanti, si hanno tumori invasivi e non invasivi. I secondi sono caratterizzati da cellule tumorali che sono confinate alla zona di origine, e vengono definiti di Stadio 0. In questa classe sono inclusi il carcinoma duttale e il carcinoma lobulare, entrambi in situ, nati rispettivamente dalle cellule dei dotti e quelle dei lobuli. Viceversa, se il tumore si introduce anche nei tessuti adiacenti viene definito invasivo e viene catalogato negli Stadi I, II, III, IV a seconda dell'aggressività. In questa classe è incluso il tipo di cancro più diffuso, ovvero il carcinoma duttale che rappresenta il 70-80% dei casi. Proveniente dai dotti galattofori questa neoplasia può diffondersi anche nel tessuto connettivo o adiposo. Il restante 20% dei casi si suddivide tra il carcinoma lobulare, originario delle cellule dei lobuli, e altre forme meno comuni come i carcinomi tubulari, papillari, mucinosi o cribriformi [19].

### 1.3 Origine, cause e fattori di rischio

Le cellule tumorali, come già citato precedentemente, hanno caratteristiche che le contraddistinguono dalle cellule normali e che consentono loro di crescere incontrollate. Ci si chiede, però, perché il corpo permetta a queste cellule di proliferare senza freni. Normalmente le cellule si moltiplicano seguendo delle regole definite dalle sequenze di geni presenti nel DNA. Le istruzioni permettono di generare diversi tipi di proteine, alcune in grado di controllare la proliferazione, promuovendola o frenandola, altre in grado indurre la morte cellulare. Nel caso di cellule tumorali però questi processi non avvengono in quanto una parte dei geni è alterata e di conseguenza portano all'interruzione dei processi ordinati, e ad una riproduzione cellulare incontrollata. Dal momento che però il nostro organismo ha innumerevoli sistemi di controllo che fanno sì che ogni processo fisiologico vada a buon fine, la formazione di un tumore avviene solo se avvengono in contemporanea diverse variazioni genetiche, processo che spesso richiede un tempo piuttosto lungo [2], [20].

Le cause alla base dei tumori sono sconosciute, ma oggi è possibile identificare dei fattori che favoriscono il loro sviluppo, tra i quali vi sono [19]:

- **l'età:** in generale la probabilità di sviluppare la malattia aumenta con l'invecchiamento e questo è dimostrato dai dati che riportano più del 75% dei casi di tumore diagnosticato a donne con più di 50 anni.
- **fattori ormonali:** il rischio aumenta anche a causa dell'esposizione per un lungo periodo di tempo agli estrogeni (ormoni prodotti dall'organismo che permettono di mantenere le caratteristiche sessuali femminili). La produzione di estrogeni è direttamente correlata ad un elevato numero di cicli mestruali nell'intera vita riproduttiva della donna che sono influenzati da diverse situazioni, tra cui una prematura comparsa del primo ciclo mestruale, una prima gravidanza dopo i 30 anni, un mancato allattamento al seno, una menopausa tardiva e infine dall'assunzione della pillola anticoncezionale.
- **stile di vita:** il consumo eccessivo di alcool, il fumare, la sedentarietà e l'obesità, sono scelte di vita che incrementano la probabilità di ammalarsi.
- **pregressa esposizione a radioterapia toracica:** nelle donne sottoposte a radioterapia al torace è presente un pericolo maggiore di incorrere in neoplasie che aumenta tanto più è giovane la paziente al momento dell'esposizione alle radiazioni.
- **densità del seno:** gli studi mostrano un rischio maggiore di sviluppare la malattia nelle donne che presentano un tessuto prevalentemente denso.
- **pregressa esperienza di cancro al seno:** le donne che hanno già sviluppato la malattia in passato hanno maggiori possibilità presentarla anche nell'altro seno.
- **familiarità:** anche se la maggior parte dei tumori alle mammelle sono forme sporadiche, circa il 5-7% dei casi è associato ad una predisposizione genetica.

La crescita di questo tumore, infatti, è resa favorevole da diverse mutazioni genetiche, tra cui quelle che interessano i geni BRCA-1 e BRCA-2, da cui dipende il 50% circa dei tumori ereditari. Va notato che in questi casi viene ereditata solo la suscettibilità alla malattia e non la malattia stessa.

Altri possibili fattori di rischio sono in fase di approvazione, tra questi vi sono l'impatto della dieta e il legame tra la presenza di alcune sostanze nell'ambiente e la probabilità di ammalarsi. È comunque di notevole importanza adottare uno stile di vita sano in quanto alcuni dei fattori appena elencati, dipendono dalle abitudini individuali che possono essere evitate o corrette.

## 1.4 Numeri del cancro

In base ai dati riportati nel report “I numeri del cancro in Italia 2021” a cura dell’Associazione italiana registri tumori (AIRTUM) e l’Associazione italiana di oncologia medica (AIOM), il carcinoma della mammella è il tumore più frequentemente diagnosticato nelle donne per tutte le fasce d’età [2]. Con circa 55 mila nuove diagnosi in un anno, questo tumore delinea più del 30% di tutte le neoplasie che si manifestano nelle donne italiane. La curva di incidenza in base all’età delle pazienti cresce fino ai 50-55 anni, poi la curva si ferma nel periodo contrassegnato dalla menopausa, per poi riprendere ad aumentare dopo i 60 anni. Ogni donna ha un rischio dell’11% circa di sviluppare questa malattia prima degli 85 anni e del 3% di morire per questa causa. In base alle stime dell’AIOM, in Italia, le neoplasie al seno sono state la prima causa di morte per tumore nel sesso femminile con 12500 decessi nel 2021. Nonostante l’incidenza sia in leggera crescita (soprattutto nelle donne più giovani), dalla fine degli anni novanta si riscontra una costante diminuzione della mortalità (-0,8% all’anno), attribuibile ad una maggior diffusione dei programmi di diagnosi precoce e ai progressi terapeutici. Questo tumore è, infatti, uno di quelli con il maggior tasso di guarigione: in Italia vivono quasi 800 mila donne che in passato sono state colpite da questa malattia. I tassi di sopravvivenza netta delle donne, però, variano molto a seconda della zona geografica studiata. Mentre in Italia la sopravvivenza a cinque anni dalla diagnosi è dell’88% e in generale nei paesi sviluppati si aggira intorno all’80%, nei paesi a basso reddito è inferiore al 40%. Questa enorme differenza è attribuibile soprattutto alla mancanza di programmi di screening e alla carenza di strutture diagnostiche e sanitarie dei paesi poco sviluppati [21], [22].

## 1.5 Prevenzione e diagnosi

Come accennato precedentemente la prevenzione basata sui programmi di screening risulta essere uno tra gli strumenti più efficaci per diminuire la mortalità associata alle neoplasie. Negli ultimi anni, in Italia, la diagnosi precoce è diventata sempre più frequente grazie ai “programmi di screening” garantiti dal servizio sanitario nazionale, che consiglia l’esecuzione di una mammografia gratuita ogni due anni a tutte le donne con età compresa tra i 50 e i 69 anni [18]. In questa fascia d’età, infatti,

è più probabile sviluppare la malattia e, secondo gli esperti dell'Agenzia internazionale per la ricerca sul cancro (IARC), invitando le donne a partecipare attivamente allo screening con queste modalità e frequenza, si può ridurre la mortalità legata a questa malattia del 40%. In diverse regioni italiane, inoltre, si sta sperimentando l'efficacia dei programmi di screening a donne in una fascia di età più ampia, in particolare tra i 45 e i 74 anni [23].

Il gold standard nei programmi di screening è rappresentato dalla mammografia, cioè un esame diagnostico preciso ed affidabile, in grado di rilevare lesioni nodulari, anche di piccole dimensioni, non ancora individuabili al tatto. Le masse tumorali possono essere circoscritte, cioè presentare bordi e forme ben definite, o spicolate, ossia possono presentare un nucleo centrale dal quale si diffondono diverse ramificazioni prive di regolarità. La mammografia è in grado di rilevare anche la presenza di microcalcificazioni, cioè piccoli depositi di calcio dovuti alle secrezioni delle cellule mutate che possono essere indice di lesioni precancerose [19]. Se l'esame mammografico ha un esito positivo, ovvero se si riscontrano delle anomalie strutturali all'interno del seno, vengono eseguiti degli accertamenti che prevedono una seconda mammografia, un'ecografia e una visita clinica che confermano o smentiscono l'effettiva presenza del tumore. A questi esami si può affiancare una biopsia, cioè un prelievo di materiale tissutale mammario che verrà sottoposto all'esame istologico, attraverso cui si valutano alcune caratteristiche del cancro, tra cui la velocità di crescita e il grado della malattia. Conclusi questi controlli si ottiene la conferma della diagnosi e, in caso di positività, si procede al trattamento che varia a seconda del grado di tumore e che prevede sempre l'eliminazione chirurgica del tessuto malato [24]. La scelta del trattamento dipende anche da alcune caratteristiche delle cellule, dall'età, dal peso, dalla presenza del ciclo mestruale e dalle condizioni generali di salute.

## 1.6 Mammografia digitale

Lo screening con mammografia è una tecnica di imaging che analizza forma e struttura della ghiandola mammaria permettendo di individuare formazioni anomale che possono avere dimensioni anche dell'ordine delle decine di micron. È una tecnica non invasiva che utilizza radiazioni ionizzanti a bassa dose per generare immagini proiettive della mammella e che offre una risoluzione spaziale non raggiungibile da altre tecniche diagnostiche come l'ecografia.

La modalità di acquisizione consiste nella generazione di un fascio di raggi X mediante un tubo radiogeno, orientato in modo da essere tangente allo sterno del paziente e da investire interamente la mammella. I raggi X sono radiazioni in grado di penetrare all'interno dei tessuti biologici e di arrivare al rivelatore con un segnale più o meno intenso a seconda dell'attenuazione che impongono le strutture attraversate dalla radiazione luminosa. Quando un fascio di raggi X viene inviato al corpo, infatti, parte di esso viene assorbita e l'altra parte viene trasmessa. La quantità di radiazione assorbita dipende dal tipo di oggetto in esame e in particolare è tanto maggiore quanto più è denso il corpo attraversato. La radiazione trasmessa contiene, quindi, un'informazione relativa al tipo di tessuto presente nel seno. Ogni struttura

della mammella è caratterizzata da un proprio coefficiente di assorbimento: i tessuti caratterizzati da un maggiore assorbimento della radiazione, come il tessuto ghiandolare, lasceranno passare una quantità minore di raggi X emergenti, producendo una immagine scura; viceversa, i tessuti con un basso coefficiente di assorbimento, come il tessuto adiposo, lasceranno passare una quantità maggiore di raggi X, producendo una immagine chiara. A seconda, quindi, del tessuto presente all'interno della mammella, l'immagine presenterà un tono di grigio differente, proprietà che permette di indagare l'anatomia del seno.

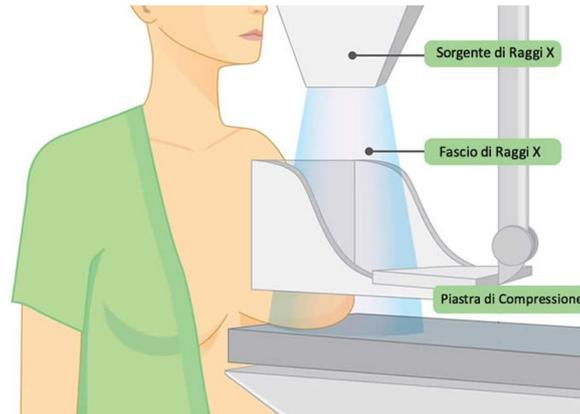


Figura 1.3. Struttura del dispositivo mammografico e posizionamento della paziente [6]

Questo tipo di esame presenta, però, delle problematiche legate alla natura tridimensionale dell'oggetto in esame. La mammella è costituita da più tessuti anatomici che, durante il passaggio dei raggi X, si sovrappongono generando un'immagine che può non rappresentare la realtà. La sovrapposizione delle strutture può, infatti, portare ad un mascheramento delle lesioni o generare strutture molto simili ad esso che possono essere interpretate in modo errato dai radiologi. Si possono ottenere due tipi di diagnosi errate: la classificazione del paziente sano come malato, detto falso positivo, e la classificazione del paziente malato come sano, detto falso negativo [25]. Gli errori falsi negativi sono quelli maggiormente preoccupanti dal momento che determinano un ritardo nella diagnosi che può danneggiare irrimediabilmente la salute della paziente. D'altra parte, un risultato falso positivo, può comportare danni importanti alla salute psicologica della donna sotto forma di stress, insonnia e preoccupazione. Anche se un certo numero di falsi allarmi è inevitabile in ogni processo di screening, è importante che tali errori siano ridotti il più possibile.

Una strategia che permette di ridurre in parte l'effetto di sovrapposizione è la compressione della mammella. Vista la geometria volumetrica delle ghiandole, la compressione permette di aumentare la qualità dell'immagine rendendo più uniforme lo spessore del tessuto attraversato dai raggi X e al contempo permettendo a strutture anatomiche che sarebbero state sovrapposte di allontanarsi. Durante l'esame, la mammella viene, perciò, schiacciata fra due piastre mediante uno strumento, detto appunto, di compressione. Un'altra strategia utile per limitare il problema è acquisire più proiezioni della mammella permettendo di massimizzare la visualizzazione

di eventuali anomalie. L'esame di screening classico prevede, infatti, l'acquisizione di una proiezione dalla testa ai piedi, detta cranio caudale e indicata come CC e una proiezione angolata in vista laterale, detta medio laterale obliqua e indicata come MLO. Le acquisizioni vengono eseguite su entrambi i seni, ottenendo un totale di quattro immagini. Nel caso in cui vengano riscontrate delle lesioni sospette e sia necessario effettuare una seconda mammografia di accertamento (detta di secondo livello), è possibile decidere di analizzare solo il seno che presenta l'anomalia in modo da non somministrare una dose inutile alla paziente.

Un altro limite che rende questo esame non affidabile al 100% risulta essere la densità mammografica, cioè un'alta percentuale di tessuto denso all'interno del seno [26]. Il tessuto adiposo è radiologicamente traslucido, quindi i raggi X possono attraversarlo senza ostacoli, mentre il tessuto ghiandolare, come le lesioni tumorali, sono radiologicamente densi e quindi bloccano i raggi X più del tessuto adiposo. La maggior parte delle donne nella fascia d'età tra i 50 e i 69 anni, presenta mammelle prevalentemente adipose, e quindi essendo il tessuto «trasparente» ai raggi, si possono facilmente identificare radiopacità anche di piccole dimensioni. Nel caso in cui, però, la mammella presenti un'alta percentuale di tessuto ghiandolare, a causa della densità del tessuto, le lesioni potrebbero non essere facilmente individuabili in quanto coperte dal tessuto radiopaco. Questa situazione è molto frequente nelle giovani donne, soprattutto nella fascia d'età tra i 20 e i 40 anni, in cui una elevata componente ghiandolare è una condizione più che normale. Per questo motivo, la mammografia viene generalmente prescritta dopo i 40 anni, in quanto prima di questa età, l'esame, salvo casi particolari, non è indicato.

## 1.7 Tomosintesi

Per limitare il problema della sovrapposizione tissutale che, come è stato detto, può portare ad errori nella diagnosi, si è introdotto un nuovo sistema diagnostico chiamato tomosintesi digitale mammaria (DBT). Tale tecnica fornisce una riproduzione quasi tridimensionale delle strutture interne alla mammella, in grado di facilitare il riconoscimento delle lesioni tumorali differenziandole dal tessuto sano. Sviluppata negli ultimi anni, viene utilizzata sempre più di frequente sia per lo screening secondario, cioè in caso di dubbio clinico, sia per la diagnosi di tumore, in quanto presenta un'accuratezza diagnostica migliore rispetto alla mammografia tradizionale [27]. Il principio di funzionamento si basa sull'acquisizione di un numero finito di proiezioni bidimensionali del seno a diversi angoli di inclinazione. In questo modo, l'organo può essere visualizzato come una successione di piani consecutivi che, uniti attraverso degli algoritmi di ricostruzione, permettono di generare il volume tridimensionale.

I sistemi DBT si compongono di un tubo a raggi X mobile che può ruotare in un intervallo angolare limitato per ottenere dati tridimensionali del seno delle pazienti. Il tubo compie una rotazione su di un arco da  $15^\circ$  a  $50^\circ$  acquisendo mediamente tra 9 e 25 immagini digitali. In generale, un numero maggiore di proiezioni, a discapito di un aumento del tempo di acquisizione, porta ad una migliore qualità dell'immagine

grazie all'aumento della risoluzione, alla riduzione della sfocatura fuori dal piano per le strutture a basso contrasto e alla riduzione degli artefatti. La tomosintesi è, tuttavia, vincolata dalla dose di radiazioni, che deve essere mantenuta nello stesso intervallo della mammografia. Per questo motivo, la dose totale viene distribuita in frazioni su tutte le proiezioni bidimensionali. L'aumento del numero di proiezioni si traduce, di conseguenza, in una dose più bassa per proiezione, che porta ad un aumento del rumore dell'immagine e ad un scarso rapporto contrasto-rumore. Inoltre, le proiezioni, come nel caso della mammografia tradizionale, vengono acquisite sulle due viste cranio-caudale e medio laterale obliqua per entrambi i seni [28].



Figura 1.4. Dispositivo di tomosintesi digitale mammaria.

Il principale vantaggio della tomosintesi è l'elevata risoluzione nel piano perpendicolare al verso di acquisizione che permette di analizzare in modo più approfondito la lesione, valutando meglio forma e dimensione al variare della profondità. È inoltre possibile determinare la posizione e la profondità del tumore esaminando ogni slice individualmente. Anche nel caso di seni densi, la DBT ha dimostrato una sensibilità più elevata rispetto alle tradizionali mammografie digitali, permettendo di ridurre al minimo il livello di “richiamo”, pure nei casi dubbi o difficili da analizzare [29]. La tomosintesi mammaria è stata approvata dalla FDA come tecnica di imaging supplementare alla mammografia, seppur non sia considerata il gold standard a causa di alcuni limiti. Nonostante produca immagini migliori, infatti, i radiologi trovano maggiore complessità nell'esaminarla dal momento che non sono abituati a refertare immagini di questo tipo in tre dimensioni. Inoltre, a causa del numero più elevato di proiezioni, i tempi di lettura risultano essere nettamente superiori a quelli necessari in mammografia comportando maggiori tempi e costi da parte dei medici. [30]. Oltre a ciò, le immagini di tomosintesi possono presentare una serie di artefatti legati all'algoritmo di ricostruzione ed alla tecnica di acquisizione. Tra quest'ultimi, un esempio è l'artefatto causato dal movimento involontario del paziente durante l'acquisizione delle varie proiezioni in rapida successione [28]. Diversi studi comparativi

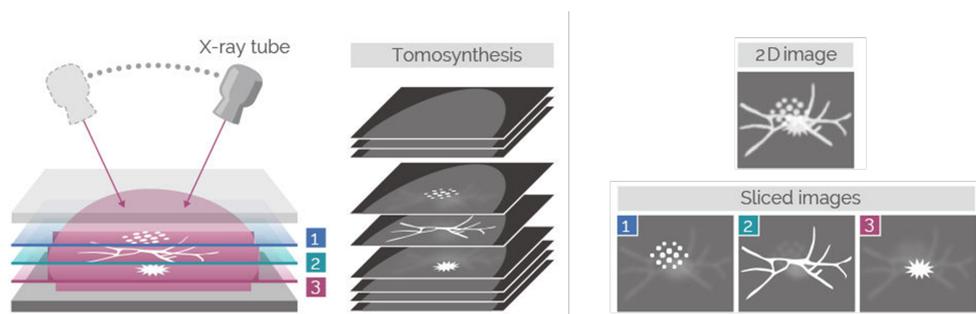


Figura 1.5. Migliore individuazione delle lesioni grazie alla limitata sovrapposizione di esse nell'immagine di tomosintesi [7]

hanno dimostrato che la tomosintesi (DBT) combinata alla mammografia tradizionale permette di aumentare i tassi di rilevamento del cancro del 9,5% rispetto alla sola mammografia digitale [27] [31]. Tuttavia, l'acquisizione combinata di queste due tecniche ha lo svantaggio di raddoppiare approssimativamente la dose di radiazioni e aumentare il tempo di acquisizione delle immagini. Pertanto, le preoccupazioni sulle radiazioni rendono la combinazione delle due meno adatta per i programmi di screening della popolazione. Per risolvere il problema, è stato sviluppato un nuovo strumento diagnostico che verrà dettagliato di seguito.

## 1.8 Mammografia sintetica

La mammografia sintetica bidimensionale (SM) fornisce immagini di ricostruzione simili a quelle della mammografia tradizionale, ma utilizzando i dati delle acquisizioni di tomosintesi. L'obiettivo della SM è quello di generare un'immagine il più simile possibile alla mammografia tradizionale, ma che allo stesso tempo permetta agli specialisti una visualizzazione più accurata dell'intero seno. La mammografia sintetica si basa sulla sovrapposizione delle singole slices della tomosintesi con un enhancing mirato sulle strutture di interesse, tra cui simil-lesioni e microcalcificazioni che possono essere individuate tramite filtraggio passa alto e, allo stesso tempo, con un'attenuazione del tessuto fibro-ghiandolare e adiposo. In questo modo si ottiene una sola immagine da refertare che permette di ridurre i tempi di analisi da parte dei medici, principale svantaggio della tomosintesi [32]. Questa tecnica non è però priva di limiti: alcuni derivano dalle criticità già presenti nella DBT (come gli artefatti da movimento o quelli dovuti alle microcalcificazioni), altri dipendono dall'algoritmo di ricostruzione utilizzato per generare l'immagine (che può causare un'errata accentuazione di strutture fisiologiche o definire in modo impreciso i margini di alcune lesioni) [32].

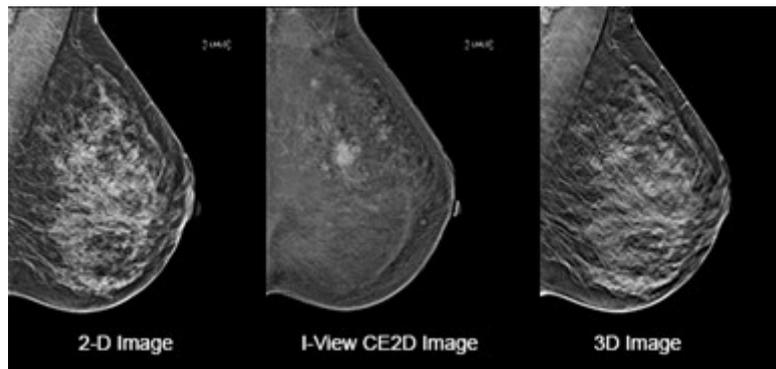


Figura 1.6. Confronto tra immagini ottenute mediante mammografia, tomosintesi e ricostruzione C-View [8]

La mammografia sintetica mostra un'ottima accuratezza diagnostica, in particolare quando si ricercano e analizzano microcalcificazioni e bordi spiculati, mentre le masse e le simmetrie vengono individuate con la stessa accuratezza che si ottiene combinando mammografia tradizionale e tomosintesi. Per questo motivo la SM è stata approvata dalla FDA come alternativa alla mammografia digitale nello screening DBT, anche se alcuni ricercatori mostrano preoccupazioni sulla qualità dell'immagine per il rilevamento e la caratterizzazione delle calcificazioni [33].



## Capitolo 2

# Apprendimento automatico

### 2.1 Stato dell'arte

L'individuazione dei tumori al seno, a partire da immagini di tomosintesi, avviene per ispezione visiva da parte di più radiologi. Questi specialisti, una volta identificata la lesione, oltre a fornirne la localizzazione, ne prevedono una classificazione in benigna o maligna, in quanto a seconda di questa diagnosi il trattamento sarà diverso. Nell'interpretare le immagini, i radiologi fanno affidamento sulla loro capacità di analisi delle lesioni, capacità che dipende principalmente dalla loro conoscenza ed esperienza. La diagnosi, quindi, è dipendente dallo specialista che la esegue e può essere accertata solo con una successiva biopsia. Il prelievo tissutale comporta, però, una procedura invasiva e traumatica per il seno che, se possibile, è preferibile evitare. Per questo motivo, una peculiarità dei programmi di screening è la presenza della doppia lettura, cioè una valutazione separata da parte di due medici che permette di diminuire gli errori [34]. Nella situazione in cui il responso degli specialisti non è il medesimo, essi si confrontano e nel caso di dubbio, la paziente viene sottoposta ad ulteriori accertamenti.

La combinazione tra un maggior numero di immagini da analizzare dovute alla DBT e la doppia refertazione, comporta un grande dispendio di tempo e costi per gli ospedali. In uno studio del 2015 è emerso che, per soddisfare la richiesta diagnostica, un radiologo dovrebbe impiegare mediamente non più di 3-4 secondi per analizzare ogni esame [35]. È in questo contesto che si sono sviluppati algoritmi automatici basati su intelligenza artificiale (AI) con l'obiettivo di migliorare i tempi di refertazione e di ridurre i falsi sospetti. La diagnosi assistita basata su questi algoritmi è in fase di sviluppo dalla fine degli anni '50, ma solo grazie al recente interesse nell'apprendimento profondo, detto anche deep learning, si sono ottenuti ottimi risultati che fanno ben sperare in questa tecnologia. I sistemi automatici possono essere utilizzati come secondo strumento diagnostico, aiutando i medici nel riconoscimento delle regioni sospette dell'immagine e affiancandoli nella classificazione delle lesioni. Gli algoritmi di deep learning hanno diversi vantaggi, tra cui la coerenza del risultato in presenza degli stessi dati di input, la non dipendenza dalla stanchezza dovuta a grandi carichi di lavoro e la possibilità di essere addestrati con un'ampia quantità di

immagini. Tra le principali criticità da affrontare, invece, vi è la ridotta disponibilità di dataset open source bilanciati tra le due classi di tumori (benigne e maligne) e il difficile riconoscimento sia di masse che di calcificazioni, a causa delle caratteristiche tipiche molto diverse.

Oggigiorno esistono sistemi di intelligenza artificiale che, nonostante le criticità, presentano performance paragonabili a quelle di uno specialista [30]. La capacità di discriminazione dei medici è, però, costruita anche su una serie di informazioni al di là dell'immagine. Questo vuol dire che i radiologi, a seconda dell'immagine che analizzano e del suo contesto clinico, possono richiedere accertamenti per comprendere meglio la situazione ed evitare trattamenti inutili. I sistemi automatici, viceversa, nella maggior parte dei casi, non hanno a disposizione altre informazioni se non quelle derivanti dall'imaging e producendo una risposta "astratta" e spesso incompleta.

Bisogna, infine, aver presente che questi sistemi sono basati su algoritmi che non sono in grado di pensare o di avere un giudizio proprio, ma che semplicemente propongono un risultato ottenuto su calcoli matematici. Il ruolo del radiologo non è dunque messo in discussione e rimane di fondamentale importanza nella diagnosi della malattia.

Nel prossimo paragrafo, verrà trattato un approfondimento sulle tecniche di intelligenza artificiale, partendo dai primi sistemi introdotti fino al più moderno Deep Learning.

## 2.2 Machine learning e Deep learning

Oggigiorno si sente parlare sempre più frequentemente di intelligenza artificiale e per questo è lecito pensare che sia una scienza recente. In realtà la sua nascita risale agli anni '50 quando gli scienziati dell'epoca svilupparono le prime macchine "intelligenti", cioè in grado di eseguire dei task emulando i comportamenti umani. Ci si può chiedere come mai una disciplina così innovativa e promettente, si sia sviluppata solo negli ultimi anni. Uno dei motivi è legato al fatto che per molto tempo gli scienziati hanno cercato di modellizzare il processo mentale umano basato sulla conoscenza a priori degli esperti, attraverso una serie di regole codificate [36]. Questo approccio fallì a causa della difficoltà nel replicare i meccanismi dell'intelligenza umana attraverso dei calcoli matematici. Tale problematica trovò soluzione con la nascita del Machine Learning (ML) che cambiò totalmente i paradigmi informatici. Invece di descrivere ogni processo mentale umano attraverso delle regole dettate dalla conoscenza degli esperti, si iniziò a fornire alle macchine un gran numero di esempi (composti sia dai dati di input che dalle risposte attese) attraverso cui le stesse macchine riuscivano a generare le regole. Quest'ultime non erano più specifiche come quelle inserite manualmente nei sistemi precedenti, ma potevano essere applicate ad altri dati per produrre nuove risposte. Venne in questo modo introdotta l'idea di "apprendimento automatico", in cui un sistema non viene programmato esplicitamente, ma viene "addestrato" sulla base dell'esperienza. Anche se rivoluzionario, questo concetto non venne applicato nella pratica per ben 30 anni a causa

dalla bassa potenza di calcolo disponibile all'epoca. Solo dagli anni 2000, grazie al miglioramento della tecnologia ed in particolare grazie alla commercializzazione delle GPU (graphics processing unit), si è potuto sviluppare il Machine Learning. I sistemi di ML, utilizzando funzioni matematiche e statistiche, sono in grado di apprendere un compito specifico e di fare previsioni senza essere stati programmati appositamente per quello. Grazie all'incremento degli studi in questo settore, negli ultimi anni i sistemi si sono evoluti sempre di più, dando origine ad un nuovo sottoinsieme del ML chiamato Deep Learning (DL) [37]. Questi algoritmi vengono impiegati con lo stesso obiettivo dei precedenti modelli di Machine Learning, ma hanno la caratteristica di permettere un apprendimento più dettagliato. Per far ciò, impiegano le reti neurali artificiali a più strati cioè modelli che simulano il funzionamento dei neuroni del cervello umano. Il "profondo" nel Deep Learning si riferisce proprio alla struttura stratificata delle reti che permette di estrarre dai dati in input informazioni molto complesse, grazie a cui il sistema guadagna efficienza. A differenza del Machine Learning che aveva comunque bisogno dell'intervento umano per identificare le caratteristiche distintive dei dati di input, il DL automatizza gran parte delle operazioni manuali, poiché la rete stessa riesce a comprendere quali informazioni sono utili e quali no al completamento del task. L'apprendimento profondo trova oggi diverse applicazioni in tanti ambiti, tra cui nella computer vision, nell'analisi dei testi, nel riconoscimento del contenuto di immagini, raggiungendo spesso prestazioni analoghe a quelle dell'osservatore umano [38].

Nei prossimi paragrafi verranno approfondite le basi biologiche e matematiche su cui si fondano le Reti Neurali e un particolare sottogruppo di reti, cioè le Reti Neurali Convolutionali, ovvero gli algoritmi fulcro di questo lavoro di tesi.

## 2.3 Reti neurali

Le reti neurali artificiali sono una famiglia di modelli matematici che imitano il comportamento dalle reti neurali biologiche durante la fase di apprendimento. Si ispirano al funzionamento biologico del cervello caratterizzato dalla trasmissione di informazioni all'interno di una rete di neuroni. Ogni neurone biologico è composto da una parte centrale, detta corpo cellulare, in cui risiede il nucleo che regola tutte le attività del neurone stesso. Dal corpo cellulare hanno origine due tipi di prolungamenti, detti dendriti e assone. I dendriti hanno il compito di ricevere i segnali da migliaia di altri neuroni, per poi propagarli verso il nucleo, l'assone, invece, conduce il segnale dal nucleo verso i dendriti di altri neuroni permettendo la propagazione dell'impulso nervoso tra cellule diverse [39]. Ciascun neurone esegue una elaborazione molto semplice (acquisisce i segnali, elabora la risposta, trasmette il nuovo segnale) e grazie alla connessione rapida ed efficiente tra tutte le cellule si sviluppa quella che definiamo intelligenza.

Analogamente alla struttura appena descritta, una rete neurale artificiale è costituita dall'interconnessione di tante semplici unità di elaborazione, dette neuroni artificiali, che, come nel caso dei neuroni anatomici, prendono tanti input dall'esterno, li codificano nel corpo cellulare e li propagano verso le altre unità.

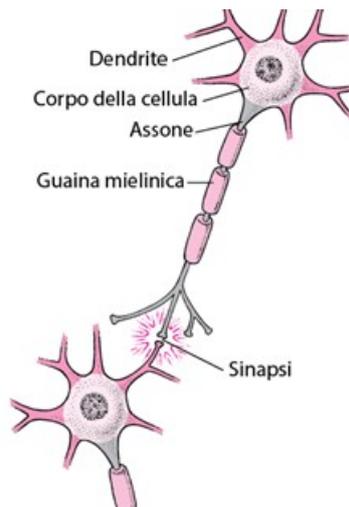


Figura 2.1. Struttura del neurone biologico [9]

Per tener conto dell'importanza del segnale di ingresso, si inserisce nel modello artificiale un parametro che influenza il risultato della rete, detto peso sinaptico. Questo parametro può assumere un qualsiasi valore reale e caratterizza la connessione di ogni unità neurale. La rete neurale artificiale è un sistema adattivo che durante l'allenamento cambia la propria struttura con l'obiettivo di avere un output della rete più simile possibile a quello desiderato. Per far ciò, vengono variati i valori dei pesi sinaptici, processo che nei prossimi paragrafi verrà trattato in modo più approfondito.

Tutti i modelli di Deep learning, come la maggior parte dei sistemi di intelligenza artificiale, sono formulati come problemi di ottimizzazione. A partire da una serie di dati di input e dagli output associati agli ingressi, l'obiettivo della rete è di minimizzare la differenza tra output reale e quello predetto dall'algoritmo. Durante l'addestramento della rete, si cercherà dunque una combinazione di pesi che permetta di minimizzare questa differenza. In questo modo, la rete, a partire dagli esempi con cui si è allenata, costruisce una mappa input-output del sistema.

### 2.3.1 Modello matematico del neurone

Per proseguire con la trattazione è necessario descrivere la funzione matematica che modella ogni neurone artificiale, fondamentale per capire il ruolo dei pesi all'interno della rete. L'unità di elaborazione riceve in ingresso una serie di input e dà come uscita una combinazione lineare dei suoi ingressi nel seguente modo:

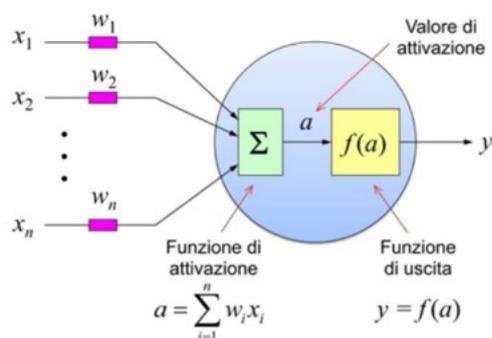


Figura 2.2. Modello del neurone artificiale [10]

dove:

- $x_n$  = ingresso relativo all'assone del n-esimo neurone a monte della connessione;
- $w_n$  = peso, o weight, della connessione n-esima;
- $a$  = funzione di attivazione del neurone, ovvero la legge che regola la sua uscita.

Secondo il modello, il neurone artificiale riceve dei valori in ingresso ( $x_1, x_2, \dots, x_n$ ) (tanti quanti sono i dati di input scelti dall'operatore) che vengono moltiplicati per il peso della connessione corrispondente ( $w_1, w_2, \dots, w_n$ ). Successivamente i risultati di tutte le moltiplicazioni vengono sommati tra loro, ottenendo un valore di attivazione  $a$ . Se tale valore supera una soglia, il neurone genera un segnale di uscita  $y$ , altrimenti rimane silente.

$$\begin{cases} y = 1, & \text{se } \sum_{i=1}^N w_i x_i > 0 \\ y = 0, & \text{se } \sum_{i=1}^N w_i x_i < 0 \end{cases}$$

Il neurone così descritto può essere assimilato ad un modello matematico di classificazione che, a seconda del risultato della moltiplicazione tra input e peso, restituisce un output.

Dal momento che i dati di input e il valore soglia sono definiti prima dell'inizio dell'addestramento, l'unico parametro che può essere cambiato al fine di ottenere un output differente è il valore dei pesi  $w_n$ . Durante l'allenamento l'algoritmo cercherà di modificare i pesi in modo tale che l'output sia più vicino possibile a quello desiderato. Questa variazione, conosciuta come Delta Rule, permette di variare i pesi in modo proporzionale ai dati di input e all'errore ( $\delta$ ) operato dal neurone; l'errore è valutato come la differenza tra l'uscita reale ( $y$ ) e quella target ( $y_d$ ), mentre la costante di proporzionalità è detta tasso di apprendimento (parametro che verrà approfondito in seguito).

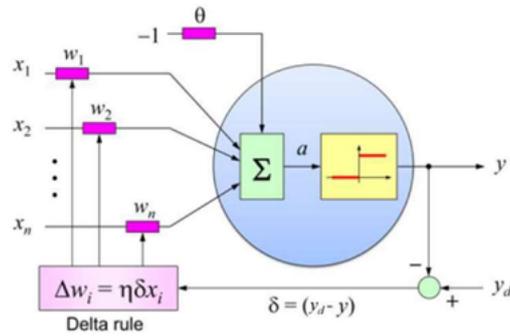


Figura 2.3. Modello del neurone con Delta Rule [10]

Questo tipo di modello non permette però di gestire i problemi complessi, dal momento che realizza una semplice combinazione lineare dei valori di ingresso. Per questo motivo viene introdotta una funzione di attivazione diversa dalla semplice soglia che permette di creare un modello fortemente non lineare. Esistono diverse funzioni di attivazioni, tra cui le principali sono la funzione sigmoideale (che determina valori di uscita compresi tra 0 e 1) e la tangente iperbolica (che determina valori compresi tra -1 e 1). Per alleggerire l'onere computazionale, spesso vengono utilizzate funzioni di attivazione ancora più semplici, come la Relu che permettono di mantenere comunque delle capacità di apprendimento molto elevate.

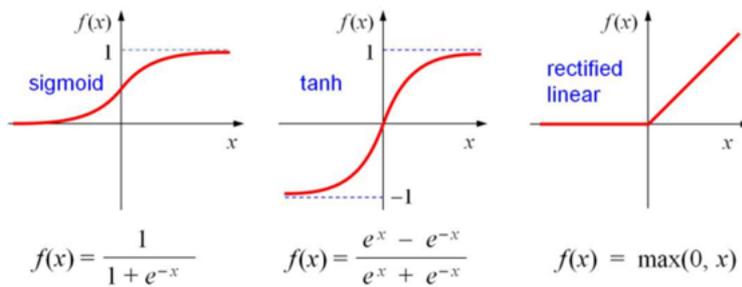


Figura 2.4. Funzioni di attivazioni: a sinistra la funzione sigmoideale, in centro quella a tangente iperbolica, a destra la ReLu [10]

### 2.3.2 Architettura della rete

Un singolo neurone non è abbastanza per costruire un algoritmo capace di prendere decisioni complesse. Per questo motivo si utilizzano tante unità neurali connesse tra loro che formano appunto una rete neurale. Le reti assumono una struttura stratificata cioè sono composte da diversi livelli che contengono quantità distinte di neuroni. L'unità neurale di ogni strato riceve come input gli output dei neuroni dello strato precedente e fornisce il proprio output ad ognuno dei neuroni dello strato

successivo, che lo accetta come input pesato.

L'architettura più semplice di una rete neurale artificiale è composta da tre layer in cui ogni neurone elabora i segnali ricevuti dal livello precedente e trasmette il risultato alle unità successive.

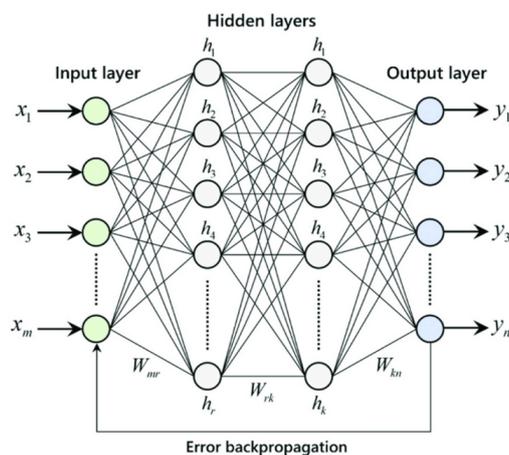


Figura 2.5. Architettura a strati di una semplice rete neurale [11]

Di seguito vengono descritti i singoli strati:

- Il primo strato viene detto input layer ed è formato da neuroni che ricevono come ingresso i dati del problema definiti dall'operatore.
- Lo strato intermedio, detto hidden layer, esegue trasformazioni non lineari degli input immessi nella rete. Nell'architettura più semplice è presente solo un livello di questo tipo, ma nei modelli più complessi ci sono diversi strati nascosti. Maggiore è il numero di livelli nascosti, maggiore sarà la profondità della rete, il che permette di determinare astrazioni sempre maggiori. Una rete che più di due hidden layer viene chiamata Deep e il meccanismo di apprendimento legato ad essa viene chiamato Deep Learning.
- L'ultimo strato, chiamato output layer, sarà formato da tanti neuroni quanti sono gli output del problema che si vuole risolvere. Se ad esempio si vuole determinare il contenuto di una immagine che può rappresentare due oggetti diversi, l'ultimo strato della rete sarà formato da due neuroni associati agli oggetti da riconoscere. Nel caso di problemi di classificazione, il compito delle unità neuronali consiste nel definire a quale categoria appartiene l'ingresso. Per far ciò, l'output di ogni neurone di questo livello viene convertito attraverso una funzione. Anche in questo caso si possono usare diverse funzioni, ma la più comune è la funzione softmax che permette di trasformare i valori output in una distribuzione di probabilità normalizzata. Ciò permette una facile individuazione del neurone a cui è associata una probabilità maggiore. Questo neurone identifica la classe di appartenenza dell'oggetto in ingresso alla rete.

### 2.3.3 Addestramento della rete

Come già è stato accennato, l'addestramento della rete si basa su un problema di ottimizzazione. A partire da una serie di dati di input e dagli output associati agli ingressi (che devono essere noti durante l'allenamento), l'obiettivo della rete è di minimizzare la differenza tra output reale e quello predetto dall'algoritmo. Uno dei metodi più efficaci per addestrare una rete neurale è quello di utilizzare l'algoritmo di retro-propagazione dell'errore, in cui i valori dei pesi della rete vengono modificati iterativamente al fine di avvicinare sempre di più l'uscita prodotta dalla rete al valore desiderato (cioè l'output noto).

Prima di addentrarci nel meccanismo di addestramento, è bene dire che l'errore commesso dall'algoritmo, ovvero la distanza che intercorre tra gli output desiderati e i corrispondenti output della rete, può essere descritto da una funzione detta funzione di costo. Durante l'addestramento, l'algoritmo aggiorna iterativamente i pesi dei neuroni fino al raggiungimento del minimo assoluto della funzione di costo.

L'algoritmo di retro-propagazione dell'errore ha inizio iniziando in maniera casuale i pesi e si compone di due fasi: la forward propagation e la backward propagation. Durante la prima fase si procede dal layer di input e si arriva al layer di output, calcolando le attivazioni dei neuroni in base al valore corrente dei pesi. Arrivati al risultato dell'ultimo layer, si confronta questo con il valore di output desiderato e si calcola l'errore attraverso la funzione di costo. A questo punto inizia la fase di backward propagation in cui si correggono i pesi a partire dal layer di output fino al layer di input.

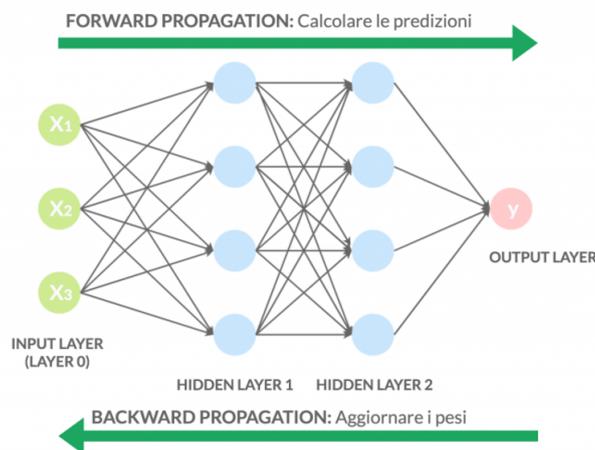


Figura 2.6. Struttura della retro-propagazione degli errori [12]

L'aggiustamento dei pesi sinaptici è funzione dell'errore e di un parametro del sistema molto importante detto tasso di apprendimento (o learning rate). Questo parametro assume valori tra zero e uno e condiziona la velocità di apprendimento dell'algoritmo: l'allenamento è tanto più rapido quanto tale parametro si avvicina a uno dal momento che questo causa una modifica maggiore del peso sinaptico verso il valore ottimo. Nonostante ciò, un learning rate molto alto accentua il rischio di

variazioni esagerate del risultato e, di conseguenza, un'instabilità del sistema. È bene quindi attribuire a tale parametro un valore adeguato, in modo tale da portare a convergenza l'algoritmo nel minor tempo possibile.

## 2.4 Reti convoluzionali

Le reti neurali convoluzionali (CNN) sono una tipologia di reti stratificate ispirate alla corteccia visiva dei mammiferi, ovvero una regione del cervello che elabora le immagini provenienti dal sistema visivo. Per far ciò, le cellule appartenenti alla corteccia si suddividono in gruppi, ognuno dei quali estrae informazioni in una specifica zona dell'immagine, agendo da filtro locale. Questo tipo di architettura è ottimale per cercare forti correlazioni spaziali, caratteristica tipica delle scene naturali. Visto le grandi performance di queste strutture, gli studiosi hanno cercato di riprodurle con l'obiettivo di identificare e classificare il contenuto all'interno delle immagini in modo molto accurato.

Quando si analizza un'immagine, essa viene rappresentata come una matrice di pixel, ovvero una matrice di valori che indicano l'intensità di colore dei vari punti dell'immagine. Se si utilizza una rete neurale per analizzare l'immagine, ogni valore della matrice rappresenterà un dato di input. Questa tipologia di rete, però, solitamente non viene utilizzata per risolvere problemi di questa natura, in quanto presenta due limitazioni. La prima riguarda il numero di neuroni di input nel primo strato della rete. Questo risulterebbe essere troppo elevato dal momento che è associato al numero totale dei pixel dell'immagine (si noti che anche nel caso si immagini con dimensioni contenute, il numero di pixel risulta essere molto alto). Questo comporta un numero enorme di pesi di collegamento tra l'input layer e l'hidden layer che aumenta molto la complessità e il costo dell'algoritmo. Un secondo problema è legato alla perdita della spazialità dei dati. Considerando ogni pixel come un dato separato dagli altri, la rete non risulta più essere influenzata dalle traslazioni e distorsioni dell'immagine, caratteristica fondamentale durante l'analisi visiva. Questi limiti vengono superati grazie allo sviluppo delle reti convoluzionali, chiamate così perché al loro interno si utilizza un'operazione matematica detta convoluzione che permette di estrarre delle caratteristiche all'interno dell'immagine.

### 2.4.1 Architettura di una CNN

Una rete neurale convoluzionale è strutturata come le reti classiche (è costituita da un blocco di input, uno o più blocchi nascosti e un blocco di output che effettua la classificazione), ma a differenza delle reti precedentemente descritte presenta degli ulteriori blocchi composti da tre livelli: un layer di convoluzione, uno di attivazione e uno di pooling che consentono l'estrazione di caratteristiche dalle immagini. Di seguito verranno descritti più in dettaglio i singoli layer.

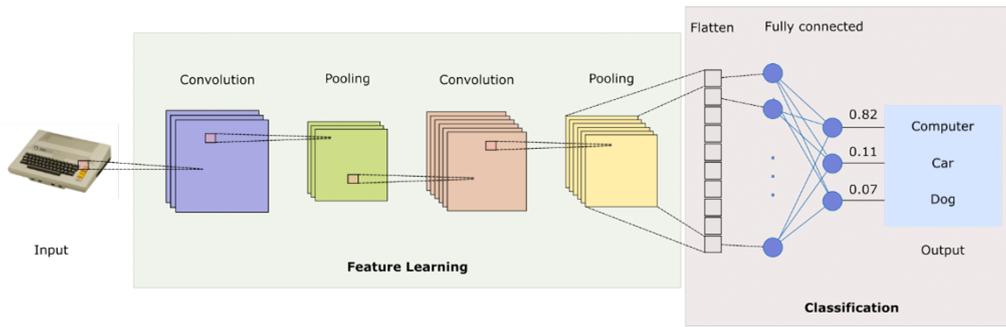


Figura 2.7. Struttura di una rete neurale convoluzionale [13]

### Input layer o strato di input

Il primo strato della rete prende in input le immagini del Dataset, introdotte come matrici di pixel. Ogni pixel è un dato singolare a cui è associato un neurone della rete formano così per ogni immagine una matrice di neuroni.

### Convolutional layer o strato di convoluzione

Il layer convoluzionale ha il compito di estrarre delle caratteristiche salienti presenti all'interno delle immagini. L'estrazione avviene attraverso l'utilizzo di filtri, ovvero delle matrici numeriche (solitamente di dimensione ridotta rispetto all'immagine da analizzare) capaci di evidenziare peculiarità ed elementi distintivi nelle immagini. Come si può notare dalla Figura, a seconda del filtro applicato, si estrae una proprietà diversa dell'immagine originale, ottenendo comunque in ogni caso una immagine (detta feature map) di dimensione uguale a quella originale.

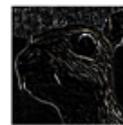
Original	Gaussian Blur	Sharpen	Edge Detection
$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$
			

Figura 2.8. Esempio di filtro convoluzionale e della sua applicazione su un'immagine [14]

Le CNN applicano diversi filtri in sequenza e il risultato del filtraggio viene “impilato” formando un blocco che ha spessore pari al numero di filtri utilizzati. L'operazione di filtraggio è un'operazione composta da un prodotto scalare tra i valori dei filtri e quelli dell'immagine e una successiva somma dei risultati ottenuti. Essendo il filtro di dimensione più piccole rispetto all'immagine, la moltiplicazione avviene



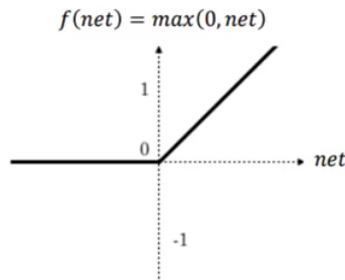


Figura 2.10. Relu function

### Pooling layers o strati di raggruppamento

I layer di pooling vengono utilizzati con lo scopo principale di ridurre la dimensione dell'input al layer successivo e, quindi, di limitare i parametri che deve apprendere la rete. In questo modo si producono delle nuove feature map che mantengono solo le caratteristiche essenziali, eliminando quelle che sono fonte di disturbo ed errore. Esistono diverse tecniche per effettuare pooling, le più comuni sono: l'average pooling (in cui si estrae la radice quadrata della somma dei quadrati dei pixel osservati), il max pooling (in cui si estrae il valore massimo fra i pixel nella finestra di osservazione). Anche in questo livello, se è presente il problema dei bordi, è possibile introdurre un padding per preservare le dimensioni della matrice di output.

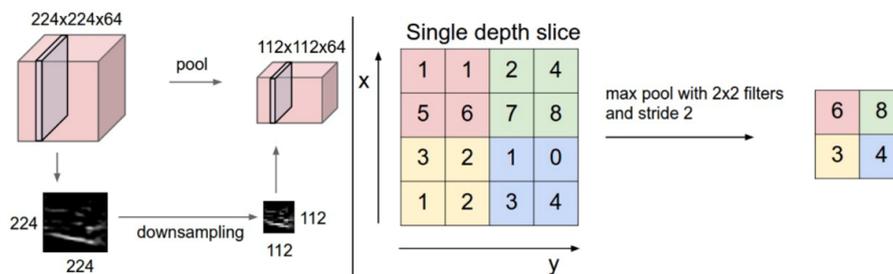


Figura 2.11. Esempio di riduzione della dimensione spaziale di una matrice tramite max pooling.

### Fully connected layer o strati completamente connessi

Il Fully connected layer è l'ultimo livello di una rete neurale convoluzionale. Prima di questo livello è sempre necessario applicare un Flatten layer, cioè un livello che permette di trasformare le matrici tridimensionali in uscita dal layer precedente in un vettore lineare.

A questo punto si introduce il livello Fully connected, cioè una rete neurale classica in cui ciascun neurone è collegato a tutti i neuroni dello strato precedente, ma due neuroni appartenenti allo stesso layer non possono essere connessi tra loro. L'output

di questo layer avrà un numero di neuroni pari al numero di oggetti che si vogliono classificare. Infine, i valori di output verranno trasformati in un dato di probabilità normalizzata che definiscono la probabilità che l'oggetto appartenga alla classe  $i$ -esima.

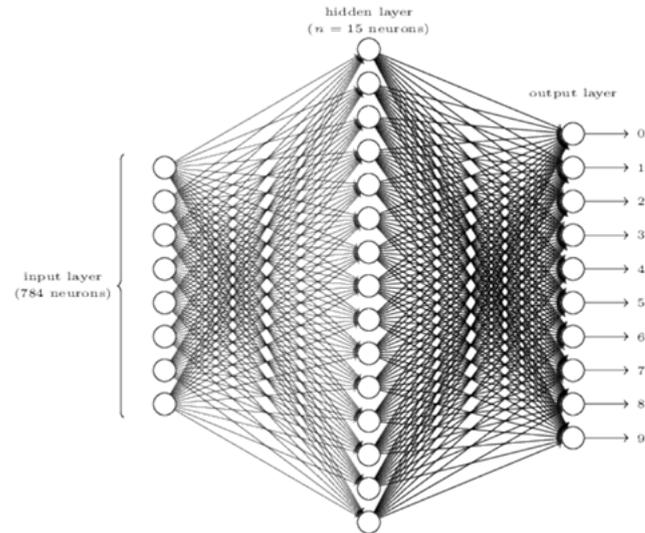


Figura 2.12. Esempio di rete neurale fully connected con 10 classi di output



# Capitolo 3

## Materiali e metodi

Il presente capitolo contiene nel dettaglio le diverse fasi di lavoro. Nello specifico, si descrive il dataset utilizzato, le modalità con cui si caratterizzano le immagini, i metodi utilizzati per la classificazione delle lesioni e i test effettuati.

### 3.1 Dataset

Il dataset si compone di immagini di mammografie sintetiche C-View di 153 pazienti che si sono sottoposte all'esame diagnostico. Le immagini sono state acquisite mediante il sistema Selenia Dimensions Mammography di Hologic e concesse dal reparto di Senologia delle Molinette tramite il supporto tecnico di Tecnologie Avanzate. Il dataset comprende le C-View di entrambe le mammelle lungo le proiezioni cranio-caudale e medio laterale obliqua. Tutte le immagini sono state rese anonime e fanno riferimento ad esami effettuati tra settembre 2014 e ottobre 2021. Le pazienti che si sono sottoposte all'esame hanno un'età compresa tra i 30 e gli 87 anni, con una media di 60 anni.

Il dataset utilizzato nel progetto è formato da un numero limitato di immagini rispetto al dataset concesso originariamente dal reparto di Senologia, in quanto alcuni casi sono stati esclusi perché non in linea con i seguenti criteri:

- Presenza di entrambe le proiezioni CC e MLO.
- Presenza all'interno dell'immagine della mammella di almeno una lesione tumorale riconducibile ad una sola tipologia (massa e opacità) ben visibile in entrambe le proiezioni.
- Presenza all'interno dell'immagine della mammella di lesioni di diametro non superiore ai 600 pixel.

Dal momento che ogni seno può contenere più di un tumore, le immagini sono state etichettate e nominate a seconda della lesione presente al loro interno. In questo modo si è ottenuto un dataset composto da 414 immagini comprendenti ognuna un'unica lesione tumorale, metà delle quali in proiezione CC e l'altra metà in proiezione MLO. Per ogni lesione è stata fornita la classe (benigna o maligna) comprovata

dall'esame istologico o dalla diagnosi della radiologia. I dati a disposizione includono 293 lesioni benigne e 121 maligne. Oltre alla classe, le senologhe che hanno effettuato la diagnosi di tumore, hanno introdotto nel dataset anche i contorni delle lesioni. I contorni sono stati colorati di rosso per essere facilmente riconoscibili come si può vedere in Figura 3.1.

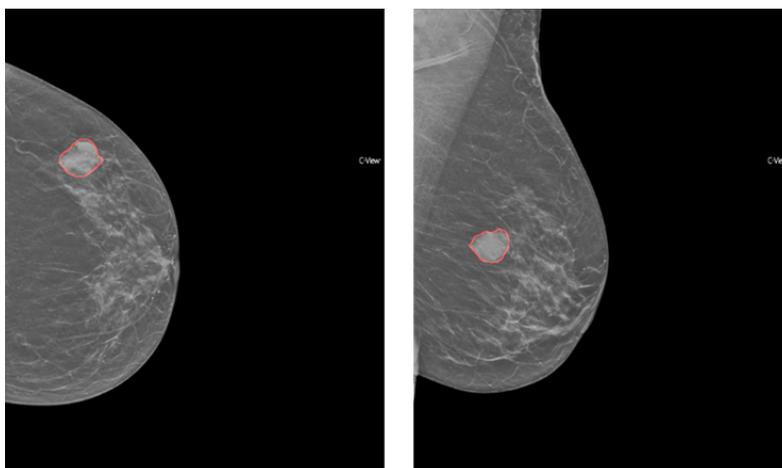


Figura 3.1. Esempio di due proiezioni (CC a sinistra e MLO a destra) in cui si evidenzia la lesione attraverso il contorno rosso posizionato dalle senologhe

## 3.2 Pipeline

L'obiettivo del progetto è la realizzazione un software automatico che permetta di predire la classe di appartenenza (benigna o maligna) di una lesione a partire dall'immagine di mammografia sintetica in cui è presente la lesione stessa. Tutti gli algoritmi di apprendimento automatico che hanno il fine di classificare un oggetto si basano su un processo detto estrazione di caratteristiche (o features extraction in inglese). Tale processo permette di trasformare il dato grezzo (in questo progetto rappresentato dalle immagini C-View) in caratteristiche numeriche che permettono di distinguere i due oggetti da classificare (lesioni benigne e maligne).

Esistono diversi metodi di estrazione di caratteristiche, in questo lavoro vengono utilizzati due approcci diversi: uno automatico e uno manuale.

- In una prima fase, si attua un'estrazione automatica delle funzionalità attraverso l'utilizzo di algoritmi di deep learning che non prevedono l'intervento umano. In particolare, vengono implementate quattro architetture CNN quali: ResNet, AlexNet, DenseNet e VGG [40] [41] [42] [43]. Questa estrazione viene effettuata unicamente nella zona dell'immagine in cui è presente il tumore, si ricavano così gli elementi peculiari che permettono la caratterizzazione delle due tipologie di lesioni al fine della classificazione.

- Nella seconda fase del progetto, vengono estratte delle funzionalità manuali che, sulla base di ricerche scientifiche, appartengono ai descrittori di primo e secondo ordine e LBP (spiegati più nel dettaglio nei paragrafi successivi). Queste feature vengono estratte sull'intera mammella per descrivere la natura del tessuto.

Una volta estratte tutte le feature, viene creato un unico vettore che unisce i due gruppi di funzionalità, utile per la caratterizzazione sia della lesione che del tessuto mammario [44] [45] [46] [47] [48]. Il vettore di feature così creato viene dato in input ad un classificatore basato su un algoritmo di machine learning. Vengono utilizzati tre tipi di classificatori, quali LDA, SVM, e KNN, la cui classificazione finale viene comparata in modo da permettere la selezione dell'algoritmo migliore a livello di performance [49].

Durante la fase di estrazione automatica, oltre ad estrarre le feature, si è anche effettuata classificazione delle immagini di input, le cui performance sono state confrontate con quelle degli algoritmi di ML allenati con l'intero set di caratteristiche. Questo processo ha permesso di determinare l'impatto della natura del tessuto mammario (descritto dalle feature manuali) sulla classificazione delle lesioni.

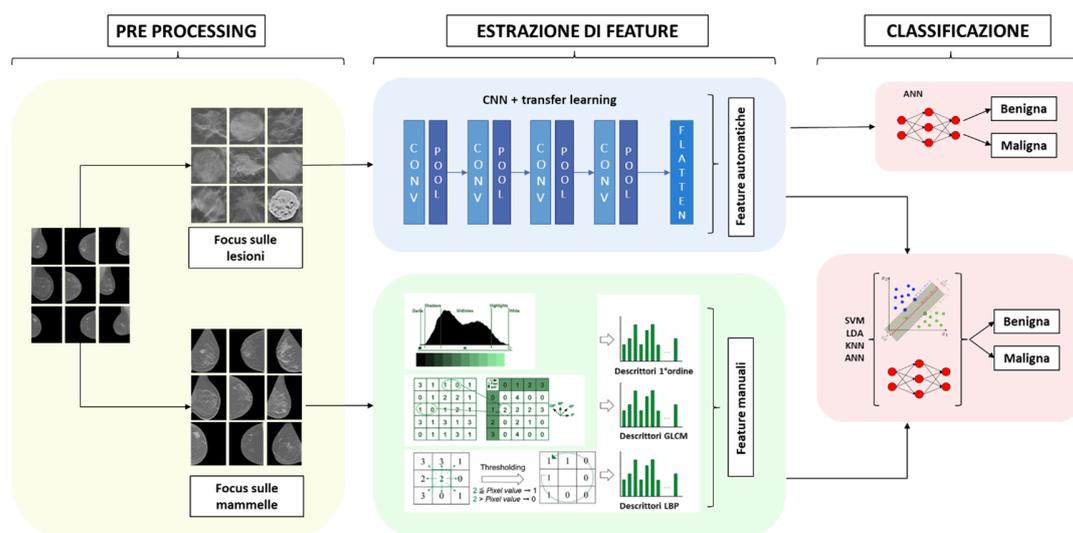


Figura 3.2. Flow-chart del progetto

### 3.3 Pre-processing

Prima della fase di features extraction, tutte le immagini vengono ricampionate, ovvero vengono variate le dimensioni dei pixel al fine di avere una qualità delle immagini equiparabile. Il ricampionamento avviene tramite il valore di Pixescaling (valore che indica la distanza fisica tra i centri di due pixel dell'immagine), proprietà presente in ogni immagine DICOM. Durante il processo di ricampionamento, le

immagini a maggiore risoluzione perdono qualità. Per limitare tale condizione, si è deciso di effettuare un ricampionamento con un valore di PixelScaling pari alla mediana dei PixelScaling di tutte le immagini. Come si può vedere dalla Figura 3.3, tale valore è pari a 0,1 mm/pixel.

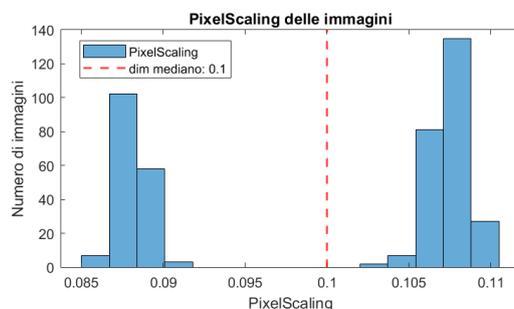


Figura 3.3. Istogramma dei valori di PixelScaling e relativo valore mediano

Successivamente, a seconda delle feature da estrarre (automatiche sulle lesioni o manuali sulla mammella), si effettuano due diversi passaggi di pre-processing.

- Nel primo caso, si ritaglia l'immagine in modo da individuare unicamente la zona in cui è presente il tumore. Per far ciò, si crea una bounding box rettangolare intorno al contorno manuale (colorato di rosso) proposto dalle senologhe.

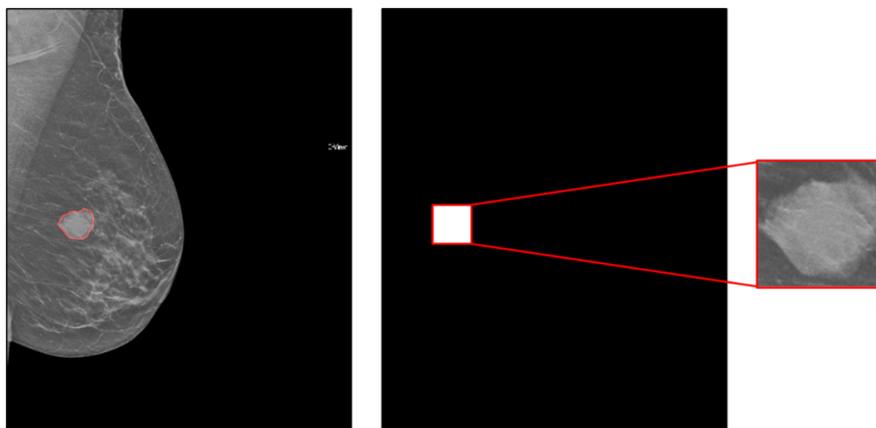


Figura 3.4. Pre-processing applicato all'immagine per individuare la lesione

- Nel secondo caso, si eliminano tutte le parti dell'immagine che non contengono la mammella, ovvero lo sfondo e nel caso di proiezioni MLO il pettorale e la parte superiore al cavo ascellare. Per far ciò, si crea una maschera in cui i pixel bianchi rappresentano la zona del seno su cui si estraggono le feature. Questa zona viene individuata attraverso specifici punti che sono stati posizionati e salvati in modo automatico dal sistema di acquisizione.

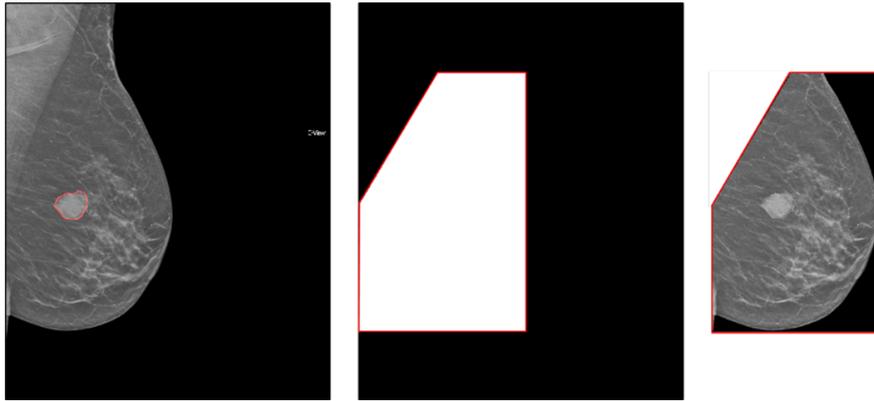


Figura 3.5. Pre-processing applicato all'immagine per individuare la mammella

## 3.4 Estrazione di feature locali

Nei paragrafi seguenti vengono riportate le modalità di estrazioni di feature sulle lesioni attraverso l'approccio automatico.

### 3.4.1 Ridimensionamento delle immagini

L'estrazione automatica di caratteristiche mediante l'utilizzo di reti convoluzionali richiede un'omogeneità delle dimensioni dei dati di input. Per questo motivo, dal momento che i tumori presentano grandezze differenti, è necessario ridimensionare ad una dimensione comune le bounding box che racchiudono le lesioni. Nella Figura 3.6 si riportata la distribuzione dei valori del lato delle bounding box, il valore della dimensione media e il valore scelto per il ridimensionamento, ovvero la potenza di due più vicina al valor medio. Le immagini ritagliate contenenti i tumori vengono quindi ridimensionate a ROI quadrate di dimensione 128x128 pixel.

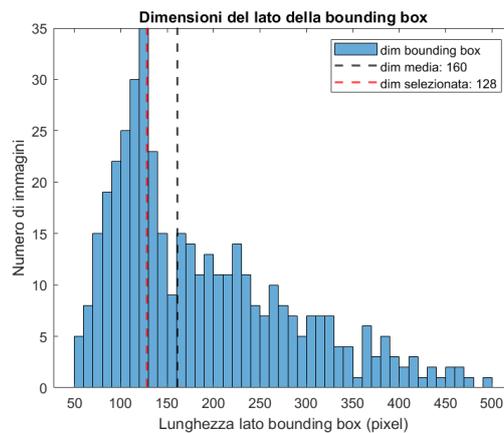


Figura 3.6. Istogramma delle dimensioni delle bounding-box delle lesioni. Si raffigura anche la dimensione media e quella selezionata.

### 3.4.2 Costruzione e divisione del dataset

Gli algoritmi di apprendimento automatico si basano sulla divisione dei dati di input in tre set di dati: Training, Validation e Test set. La suddivisione è necessaria in quanto non si può valutare il modello sugli stessi dati utilizzati durante l'addestramento. Questo perché si vuole creare un sistema automatico che possa classificare nel modo corretto non solo le immagini presenti nel dataset di partenza, ma anche dati diversi da quelli su cui è stato allenato. Così facendo l'algoritmo non si specializza sui dati di input, ma riesce ad estrarre feature ed informazioni generalizzabili a diversi nuovi casi. Il dataset viene quindi diviso in tre gruppi con le seguenti funzioni:

- Il modello è inizialmente adattato su un set di dati di addestramento detto Training set. Questo set è formato da una serie di esempi che vengono utilizzati per adattare i parametri del modello (nel caso di reti neurali artificiali permette di adattare le connessioni sinaptiche tra i neuroni). Durante la fase di allenamento, il sistema costruisce delle relazioni tra le immagini di input e la classe di appartenenza della lesione presente nell'immagine. Eseguendo poi un confronto tra il risultato predetto e quello atteso sulla base del quale verranno adattati i parametri.
- Successivamente, il modello addestrato viene utilizzato per predire gli output di un secondo set di dati chiamato set di dati di convalida o Validation set. Questo set di dati permette di valutare la qualità del sistema allenato e la sua capacità di generalizzare, ovvero di predire nel modo corretto immagini diverse da quelle utilizzate nella fase di addestramento. Se le performances sul Validation set non risultano essere elevate, si modificano gli iperparametri del modello e si allena nuovamente l'algoritmo sul Training set, finché il risultato sul set di convalida non è soddisfacente.
- Infine, l'ultimo set di dati è detto Test set ed è utilizzato per fornire una valutazione imparziale su un set di dati che il sistema non ha mai visto. Ciò permette di ottenere delle performance simili a quelle che si otterrebbero con un qualsiasi nuovo set di dati.

Il dataset di partenza composto da 414 lesioni (293 benigne e 121 maligne) è stato quindi diviso nei tre set appena descritti. La percentuale di immagini in ogni set è la seguente: il 70% del dataset fa parte del Training set, il 20% del Validation set e il rimanente 10% del Test set. Come si può notare, sia dalle percentuali appena elencate, sia dal grafico sottostante, il set di addestramento è il più numeroso in quanto è necessario un grande numero di immagini nella fase di allenamento per ottenere un algoritmo che abbia la capacità di generalizzare. Inoltre, per evitare che il rapporto tra lesioni benigne e maligne sia diverso nei vari set, le percentuali vengono applicate ai dati di entrambe le classi in modo separato (ovvero si dividono le lesioni benigne dalle maligne e successivamente per i due gruppi si effettua la suddivisione in set). Il numero di lesioni per ogni classe e per ogni set è riportato nella tabella seguente (Figura 3.8).

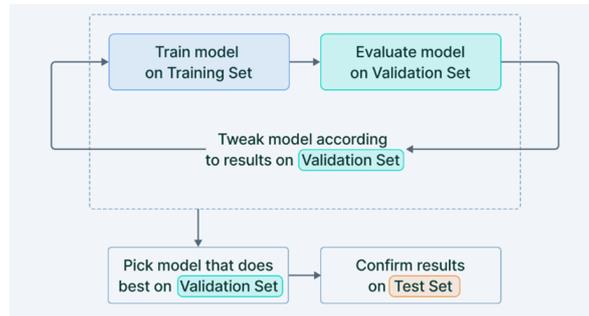


Figura 3.7. Flow-chart riguardante l'utilizzo dei dataset durante l'addestramento e la validazione dei modelli

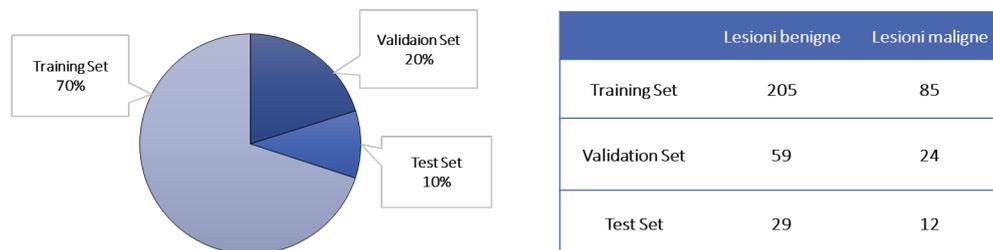


Figura 3.8. Percentuali in cui si è diviso il dataset a sinistra e numero di lesioni per ogni set a destra.

### 3.4.3 Transfer learning

Suddiviso il dataset, si può passare alla fase di costruzione del modello CNN. Le reti neurali profonde richiedono un grande costo computazionale per l'addestramento e costruire una rete da zero può portare, oltre ad un grande dispendio di tempo, anche ad un'architettura con scarse prestazioni. Per risolvere questo problema spesso si utilizza una strategia detta transfer learning che consiste nello sfruttare una rete già addestrata per un nuovo task [50]. Esistono alcune architetture che sono note per le loro elevate performance in quanto vincitori di concorsi che puntano appunto a premiare la rete migliore. Attraverso il transfer learning, i modelli performanti già addestrati su dataset di grandi dimensioni (il più comune è ImageNet composto da 14 milioni di immagini divise in 1000 classi) vengono utilizzati come punto di partenza per risolvere un nuovo task che in questo progetto è la classificazione delle lesioni. Per rendere il modello capace di eseguire il nuovo compito, si riutilizzano i pesi dell'architettura pre-addestrata e si rimuove il layer di classificazione e la softmax adattandoli al nuovo obiettivo.

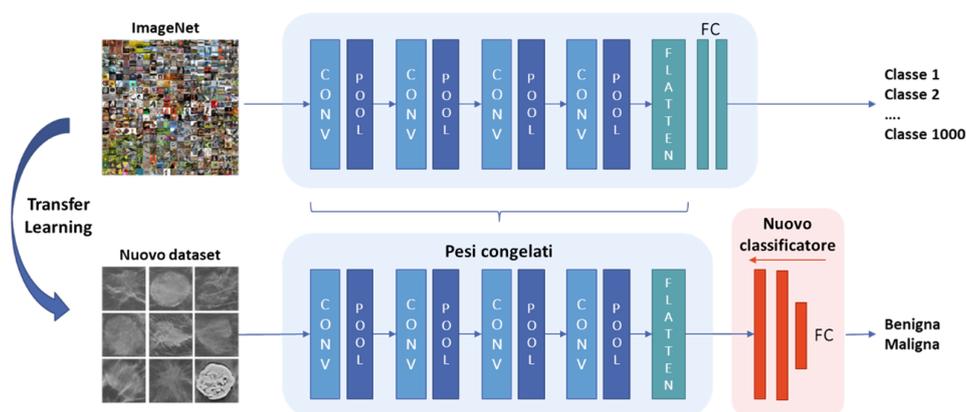


Figura 3.9. Esempio di Transfer Learning applicato al dataset di questo progetto

### 3.4.4 Architettura delle reti

In questo lavoro vengono implementate quattro architetture di reti neurali convoluzionali, tra le più comuni in letteratura [40] [41] [42], [43], le cui performance sono confrontate al fine di determinare il modello più adatto allo scopo.

#### AlexNet

Alexnet è una rete neurale convoluzionale profonda per la classificazione di immagini proposta nel 2012 da Alex Krizhevsky. È stata la prima rete convoluzionale ad utilizzare la GPU per aumentare le prestazioni, attraverso cui ha vinto il concorso ImageNet ILSVRC-2010 con un tasso di errore del 17%. Questo modello presenta otto strati, di cui i primi cinque sono convoluzionali, mentre gli ultimi tre sono livelli completamente connessi. Gli strati completamente connessi hanno 4096 neuroni ciascuno e il secondo strato completamente connesso viene collegato ad un ultimo strato formato da classificatore softmax con 1000 classi. Caratteristica principale di questa rete è la presenza di un livello di max pooling sovrapposto nei primi due livelli convoluzionali. Questi livelli applicano il classico max pooling su finestre che però non sono adiacenti ma si sovrappongono una con l'altra.

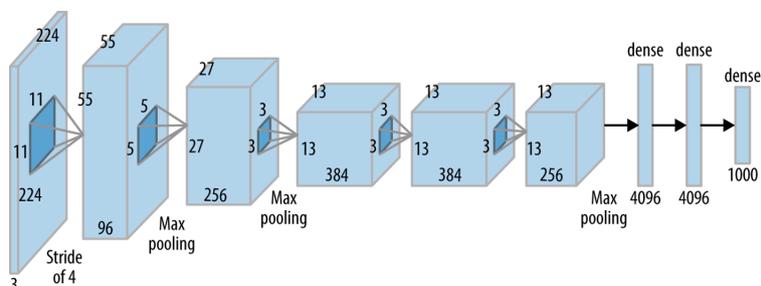


Figura 3.10. Schema dell'architettura AlexNet

## VGG

Il modello VGG o VGGNet è un modello di rete neurale convoluzionale che ha raggiunto il 92,7% di accuratezza dei primi 5 test su ImageNet. Il modello VGG16, così chiamato perché presenta 16 strati di cui 13 convoluzionali, ha una struttura uniforme che, nonostante il grande numero di strati, la rende una rete semplice. La principale differenza rispetto alle classiche CNN è che utilizza tanti filtri di dimensione ridotta, principalmente  $3 \times 3$ . Il numero di filtri raddoppia ad ogni livello di convoluzione, arrivando a 512 nell'ultimo livello. L'architettura si conclude con tre livelli completamente connessi, dei quali i primi due hanno 4096 canali ciascuno e il terzo è un classificatore softmax a 1000 classi. Il limite più grosso di questa architettura è il grande numero di strati (e di conseguenza il grande numero di parametri da determinare che arriva a 138 milioni) che la rendono molto pesante dal punto di vista computazionale e difficile da implementare su sistemi con bassa potenza di calcolo.

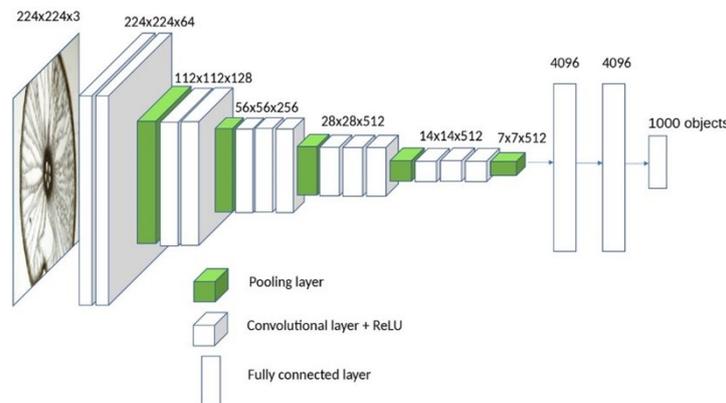


Figura 3.11. Schema dell'architettura VGG

## ResNet

La ResNet è una CNN profonda molto utilizzata nell'ambito della classificazione di immagini che nel 2015 ha vinto il primo posto nella competizione ILSVRC per la classificazione del dataset ImageNet. È un'architettura nata per limitare un problema caratteristico delle reti profonde, ovvero la scomparsa del gradiente. Nelle reti che presentano molti strati, a causa di una retro-propagazione ridotta, i primi livelli non hanno una grande variazione dei pesi e di conseguenza non danno un contributo nell'estrazione di feature. La ResNet risolve questo problema introducendo la skip connections (o blocchi residui) ovvero l'aggiunta di più connessioni tra blocchi distanti che permettono di propagare l'input saltando alcuni livelli. L'input di alcuni strati non sarà, quindi, soltanto l'output dello strato precedente ma comprenderà anche l'output di alcuni livelli precedenti.

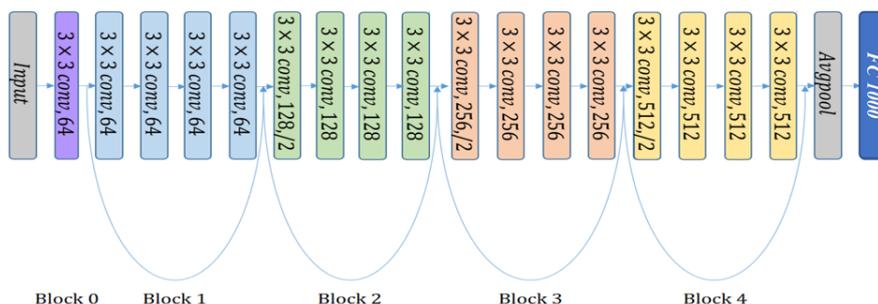


Figura 3.12. Schema dell'architettura ResNet

Ci sono diverse architetture ResNet che si differenziano per il numero di strati e per le skip connections. In questo progetto viene implementata la ResNet18 che, come dice il nome, contiene 18 layer e quattro blocchi residui. L'ultimo blocco residuo è connesso ad uno strato di average pooling e ad un livello completamente connesso seguiti che permette di classificare l'immagine in ingresso in una delle 1000 classi di oggetti attraverso l'applicazione della funzione softmax.

## DenseNet

La rete di DenseNet è una delle innovazioni nel campo del machine learning. Risulta molto simile alla struttura della ResNet, ma invece di sommare l'output del layer successivo con l'input del layer precedente, concatena gli output di più layer. Come la ResNet, la DenseNet è stata sviluppata con il fine di risolvere il problema della scomparsa del gradiente in reti molto profonde. Il principio su cui si fonda questa tipologia di rete è la connessione di ogni layer con tutti gli altri. Nello specifico la rete DenseNet è composta da una serie di blocchi Dense in cui ogni layer ha come input l'output di tutti i layer precedenti. Ogni blocco si compone di un livello di batch Normalization, uno di attivazione mediante ReLU e uno di convoluzione con kernel 3x3.

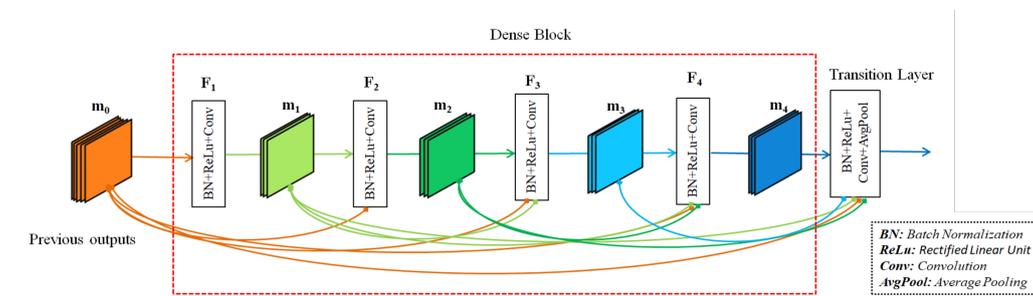


Figura 3.13. Schema di un blocco dell'architettura DenseNet

### 3.4.5 Tuning dei parametri

A partire da un modello è poi possibile apportare diverse modifiche sia a livello strutturale sia a livello degli iperparametri (ovvero tutte le variabili della rete che possono essere inizializzati e ottimizzati prima della fase di allenamento). In questo lavoro si è deciso di non apportare modifiche alle architetture (se non per i layer di classificazione che sono stati adattati al nostro scopo), mentre si sono effettuate numerose sperimentazioni per trovare la combinazione di iperparametri migliore.

#### Funzione di costo

Come già detto, le reti neurali imparano attraverso il confronto tra l'output che predicono e quello reale. Questo confronto viene espresso in termini matematici attraverso una funzione che è chiamata funzione di costo (o loss function in inglese). Affinché la risposta sia la più esatta possibile, è necessario minimizzare questa funzione. Esistono diverse tipologie di funzioni da poter adoperare in base al task predefinito, in questo lavoro si è utilizzata la cross-entropy, definita in termini matematici dalla seguente espressione:

$$C = -\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^c [y_i \log(a_i) + (1 - y_i) \log(1 - a_i)] \quad (3.1)$$

dove  $y$  è il vettore degli output target,  $a$  è il vettore degli output predetti dalla rete,  $n$  rappresenta il numero degli input e  $c$  in numero di classi (nel nostro caso pari a 2).

#### Tasso di apprendimento

Un altro parametro fondamentale per l'addestramento della rete è il tasso di apprendimento detto anche learning rate. Esso determina lo spostamento verso il minimo della funzione di costo che viene effettuato ad ogni iterazione. Mentre la direzione di discesa è solitamente determinata dal gradiente della loss function, il tasso di apprendimento determina quanto grande è il passo in quella direzione. Poiché influenza la misura in cui le informazioni appena acquisite sostituiscono le vecchie informazioni, rappresenta metaforicamente la velocità con cui un modello di apprendimento automatico "apprende". Se il tasso di apprendimento è molto piccolo, il modello convergerà troppo lentamente, mentre, se è troppo grande, il modello divergerà. Al fine di ottenere una convergenza più rapida, prevenire oscillazioni e rimanere bloccati in minimi locali indesiderati, il learning rate viene spesso variato durante l'allenamento attraverso un suo adattamento che consiste solitamente nella diminuzione del suo valore.

In questo lavoro vengono effettuate numerosi allenamenti con valori di learning rate differenti, in particolare si è partiti da un tasso di apprendimento pari a 0,001 e lo si è diminuito fino a 0,0001.

## Dimensioni dei batch

Nel caso di dataset numerosi, spesso non si allena la rete con tutte le immagini, ma si preferisce dividere il set in gruppi di dimensioni più piccole, detti batch. La dimensione del batch definisce quindi il numero di campioni che vengono propagati attraverso la rete. Questo permette di allenare la rete utilizzando meno memoria e più velocemente. Il limite di questo approccio consiste nel fatto che più è piccolo il lotto, meno accurata sarà la stima del gradiente.

Per le quattro architetture proposte prima si è scelto di effettuare tre prove dividendo il dataset in batch contenenti rispettivamente di 8, 16 e 32 immagini.

### 3.4.6 Addestramento della rete e overfitting

A seguito dell'inizializzazione degli iperparametri, dei pesi e del settaggio della GPU (quando disponibile), inizia il vero e proprio allenamento della rete sul Training set. In questa fase risulta essere di fondamentale importanza capire quali possono essere i possibili errori di predizione. Questi errori provocano una scarsa capacità di generalizzazione della rete detto problema dell'overfitting. In questa situazione, il modello apprende eccessivamente gli esempi del Training set, estraendo non solo le informazioni fondamentali per la classificazione, ma anche dettagli e rumori casuali che causano performance non soddisfacenti su un nuovo set di dati (Validation set). La Figura 3.14 descrive tale situazione in termini di loss, che decresce per entrambi i set fino ad un certo punto in cui la funzione del Training continua a decrescere mentre quella del Validation set inizia a crescere.

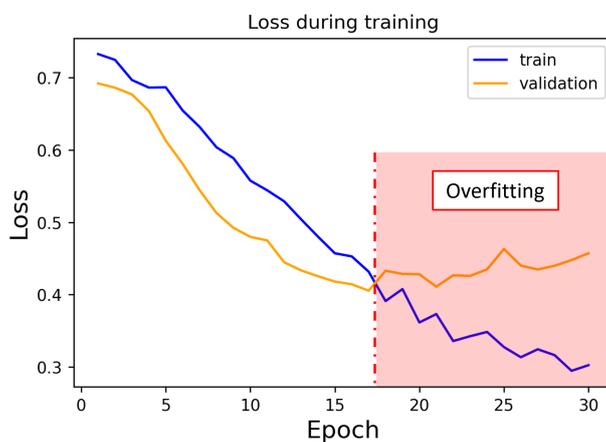


Figura 3.14. Esempio di overfitting della rete

Vi sono diversi approcci per limitare l'overfitting, i principali verranno trattati nei prossimi paragrafi.

## Data augmentation

L'aumento dei dati, detto anche data augmentation, è un primo metodo per limitare l'overfitting e aumentare la capacità di generalizzazione dell'algoritmo [51] [52]. Consiste nell'aumentare “artificialmente” la numerosità del dataset, applicando delle trasformazioni alle immagini (tra cui rotazione, flip, variazione del contrasto o della luminosità, distorsioni, zoom, etc.) che permettono di avere nuovi esempi durante la fase di allenamento. Questa tecnica è molto utilizzata in letteratura, soprattutto per ottenere prestazioni migliori quando si hanno set di dati limitati.

In questo lavoro, si implementano due tipi di data augmentation: uno online e uno offline.

- Nella tecnica online le immagini di ogni batch vengono trasformate in modo casuale, prima di ogni iterazione durante il Training del modello (non si creano quindi delle copie di immagini ma si modificano quelle originali). In questo caso, le immagini trasformate non vengono salvate sul disco ed è impossibile sapere quale immagine è stata modificata. Le trasformazioni che vengono applicate con questa tecnica sono le seguenti: flip orizzontale, flip verticale, variazione del contrasto e variazione della luminosità tutti applicati con una probabilità del 50% di verificarsi.
- Nella tecnica offline, l'aumento dei dati avviene su tutto il dataset prima dell'inizio dell'addestramento. Il dataset così creato viene salvato su disco e conterrà tante immagini “copie” quanti aumenti vengono effettuati. Il modello viene quindi allenato attraverso un dataset composto da molte più immagini rispetto a quello originale che permette di aumentarne la robustezza e limitare l'overfitting. Le trasformazioni che vengono applicate con questa tecnica sono le seguenti: rotazione di 90°, 180°, 270°, flip orizzontale e flip verticale. In figura sono riportati i due tipi di data augmentation applicati ad un'immagine del Training set.

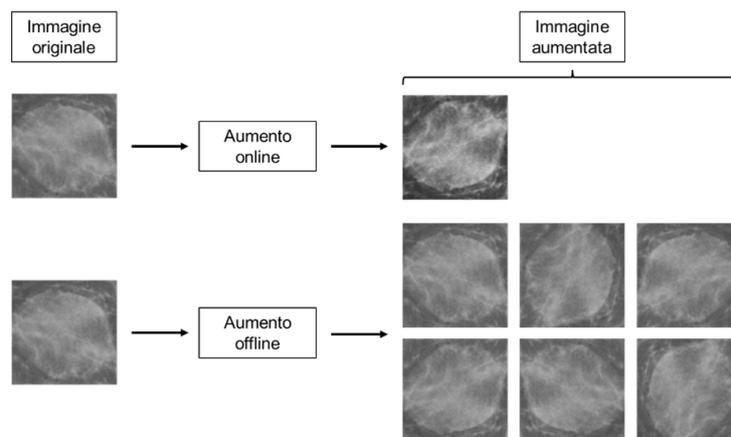


Figura 3.15. Esempio di data augmentation offline e online

## Numero di epoche ed arresto anticipato

Gli algoritmi di apprendimento automatico aumentano le proprie performance all'aumentare del tempo di addestramento, ovvero all'aumentare del numero di epoche di allenamento. Nel caso di overfitting non è, però, efficiente allenare l'algoritmo su un numero di epoche molto grande in quanto ad un certo punto le performance aumenteranno solo sul Training set e non più sul set di validazione. In questo progetto si allenano le reti per 50 epoche, applicando una particolare tecnica chiamata arresto anticipato o early stopping. Tale tecnica consente di interrompere l'allenamento quando le prestazioni su un set di convalida smettono di aumentare su un certo numero di epoche, prevenendo così l'overfitting.

## Regolarizzazione

Un'ulteriore tecnica che limita l'overfitting prevede l'introduzione di un termine di regolarizzazione nella descrizione della funzione di costo. Questo termine è definito come la somma dei quadrati dei pesi moltiplicati per un termine costante lambda. Il valore di lambda determina l'importanza del termine di regolarizzazione: più è elevato, più la funzione di costo è influenzata da questo termine.

$$\begin{aligned} \text{Cost function} &= \text{Loss function} + \text{Regularization Term} \\ &= \text{Cross Entropy} + \lambda \sum_{j=1}^n w_j^2 \end{aligned} \quad (3.2)$$

Grazie a tale termine, la funzione di loss (e quindi l'errore commesso dalla rete) aumenta nel caso in cui i valori dei pesi siano molto alti. Riducendo il valore dei pesi, si limita l'importanza dei dati di input del Training set che può causare una bassa generalizzazione su altri set di dati. Così facendo, si riduce l'overfitting e dal momento che i pesi tendono allo zero, si modera la complessità del modello.

## Dropout

Il dropout è una tecnica che "ignora" alcuni neuroni dalla rete durante l'allenamento. Durante ogni iterazione del Training, alcune unità neurali vengono silenziate casualmente, il che causa una diminuzione di neuroni che lavorano nel processo di apprendimento. Tale processo porta ad avere una struttura della rete più semplice e, allo stesso tempo, un modello più robusto, dal momento che l'apprendimento dei pesi non può fare affidamento solo su alcuni neuroni specifici. La rete sarà quindi costretta a dare importanza a tutti i neuroni presenti, senza sensibilizzarsi su alcuni di essi che potrebbero essere silenziate durante l'allenamento.

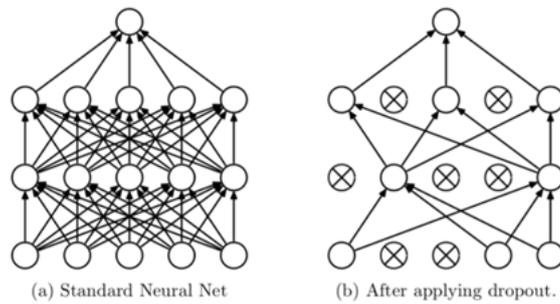


Figura 3.16. Esempio di rete senza applicazione di dropout (a) e con applicazione di dropout (b)

## 3.5 Estrazione di feature globali

Nei paragrafi seguenti vengono riportate le tipologie di feature estratte sulle immagini delle mammelle attraverso l'approccio manuale.

### 3.5.1 Feature estratte

L'estrazione manuale di caratteristiche nella zona delle immagini in cui è presente la mammella avviene attraverso l'analisi di tessitura (detta *texture analysis* in inglese). Questa tecnica permette di descrivere la natura del tessuto attraverso le informazioni relative al valore assoluto dei pixel e alla loro distribuzione all'interno dell'immagine. Per estrarre le caratteristiche si utilizzano delle precise definizioni matematiche, chiamate "descrittori", che sono indipendenti dalla modalità di imaging. I descrittori che estraggono feature a partire dai valori nativi delle immagini, detti del primo ordine, si basano sullo studio dell'istogramma delle luminosità, mentre i descrittori che analizzano i rapporti fra due o più pixel dell'immagine, detti di secondo ordine o superiore, si basano su particolari matrici su cui vengono successivamente calcolati i descrittori. Nei paragrafi seguenti vengono descritti nel dettaglio i descrittori utilizzati in questo progetto.

### 3.5.2 Descrittori del primo ordine

I descrittori del primo ordine si ricavano dall'istogramma delle luminosità, ovvero un grafico che sull'asse x riporta i valori di grigio e sull'asse y il numero di volte che compare tale valore nei pixel dell'immagine e descrive, perciò, la distribuzione dei valori di grigio all'interno della mammella. Questi descrittori sono operatori statistici che non forniscono informazioni a livello morfologico, ma permettono di ottenere una stima della luminosità e dell'omogeneità dell'immagine. I descrittori ricavabili sono il valore medio e la deviazione standard delle intensità, la skewness, la kurtosis e l'entropia dell'istogramma, la cui formulazione matematica è riportata nella tabella seguente (Tabella 3.1).

Feature I° ordine	Formula
Media	$\frac{1}{N} \sum_{i=1}^N A_i$ (3.3)
Varianza	$\frac{1}{N-1} \sum_{i=1}^N A_i - v^2$ (3.4)
Skewness	$\frac{E(x-v)^3}{\sigma^3}$ (3.5)
Kurtosis	$\frac{E(x-v)^4}{\sigma^4}$ (3.6)
Entropia	$-\sum hist \log_2 hist$ (3.7)

Tabella 3.1. Descrittori di I° ordine

### 3.5.3 Descrittori di Haralick e Galloway

I descrittori di Haralick e Galloway permettono di valutare la dipendenza spaziale tra pixel. Per calcolarli è necessario costruire la matrice delle co-occorrenze o GLCM, ovvero una matrice quadrata e simmetrica, le cui righe e colonne rappresentano tutti i valori di grigio che possono assumere i pixel dell'immagine. Ogni cella della matrice contiene il numero di volte in cui un pixel con un determinato livello di grigio è vicino ad un altro pixel con un altro colore. Successivamente si conta quante volte occorre nell'immagine l'accoppiamento di due pixel. Attraverso alla GLCM si può valutare la presenza di pattern e l'omogeneità dell'immagine. La matrice delle co-occorrenze è stata calcolata per quattro angoli di inclinazione, ovvero per 0°, 45°, 90° e 135° e poi successivamente i parametri estratti da ogni matrice sono stati mediati.

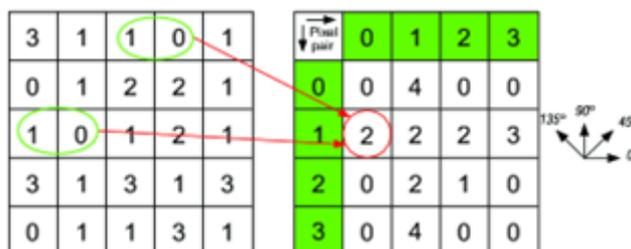


Figura 3.17. Matrice GLCM

Una volta creata la matrice, vengono calcolati cinque descrittori, quali: il contrasto, la correlazione, l'energia, l'omogeneità e il valore massimo, la cui formulazione matematica è riportata nella tabella seguente (Tabella 3.2).

Feature II° ordine	Formula
Contrasto	$\sum_{i=1}^{N-1} \sum_{j=1}^{N-1}  i - j ^2 P(i, j)$ (3.8)
Correlazione	$\frac{1}{\sigma_x \sigma_y} \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} (i - v_x)(j - v_y) P(i, j)$ (3.9)
Energia	$\sum_{i=1}^{N-1} \sum_{j=1}^{N-1} P(i, j)^2$ (3.10)
Omogeneità	$\sum_{i=1}^{N-1} \sum_{j=1}^{N-1} 11 + (i - j)^2 P(i, j)$ (3.11)
Massima probabilità	$\max(P(i, j))$ (3.12)

Tabella 3.2. Descrittori di II° ordine

### 3.5.4 Local binary patterns

I Local Binary Patterns (LBP) sono dei descrittori che indagano il contrasto locale all'interno dell'immagine. Per far ciò si analizza l'intorno 8-connesso di ogni pixel e si assegna il valore 1 se un pixel dell'intorno ha un valore superiore a quello del pixel centrale mentre si assegna il valore 0 se ha valore minore di quello centrale. Successivamente, si leggono, in senso orario o antiorario, i valori associati all'insieme 8-connesso (zero e uno) trasformandoli prima in un numero binario e poi in codifica decimale. Questo si esegue per tutti i pixel dell'immagine e i valori che si ottengono vengono mediati per ricavare un solo parametro.

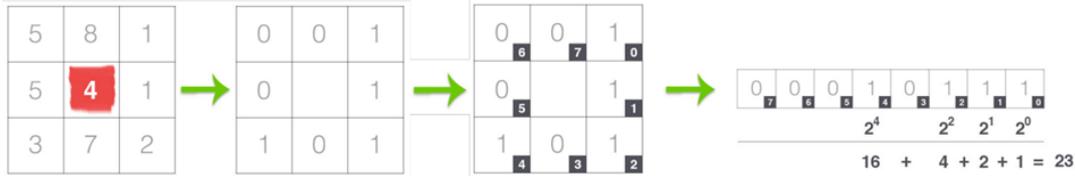


Figura 3.18. Descrittori LBP

## 3.6 Feature selection

A seconda dell'architettura scelta si hanno un numero diverso di feature estratte, partendo da 512 feature estratte da ResNet fino ad arrivare a 4096 feature estratte dalle reti AlexNet o Vgg. Visto il grande numero di funzionalità estratte dall'apprendimento profondo in confronto a quelle estratte manualmente, si è deciso di provare a riequilibrare i due gruppi in modo da dare più peso alle feature globali. Per questo motivo si è scelto di eseguire feature selection attraverso la tecnica del Random Forest applicata esclusivamente alle feature locali.

Il Random Forest è un algoritmo di apprendimento automatico basato sull'insieme di tanti alberi decisionali (o Decision Tree) i cui output vengono combinati per raggiungere una singola previsione. A loro volta, gli alberi decisionali sono degli algoritmi di apprendimento supervisionato che sono, però, soggetti ai problemi di

bias e overfitting. La potenza del Random Forest consiste nella combinazione di tanti alberi che riduce i problemi appena citati. Unendo diverse previsioni, si ottengono risultati più accurati e meno soggetti a bias, in quanto ogni albero effettua una diversa misclassificazione che una volta mediata avrà un peso minore.

Il problema dell'overfitting viene, invece, superato allenando i Decision Tree con input diversi. Gli alberi non utilizzano l'intero dataset per l'addestramento, ma ogni algoritmo viene allenato su un campione causale di dati del set di partenza. Maggiore è il numero di alberi e minore sarà l'overfitting, a discapito però delle performance sul set di Testing. In questo progetto si è selezionato un numero di decision tree pari a 5 che ha permesso di limitare l'overfitting, ma allo stesso tempo non ha peggiorato le prestazioni sul Test set.

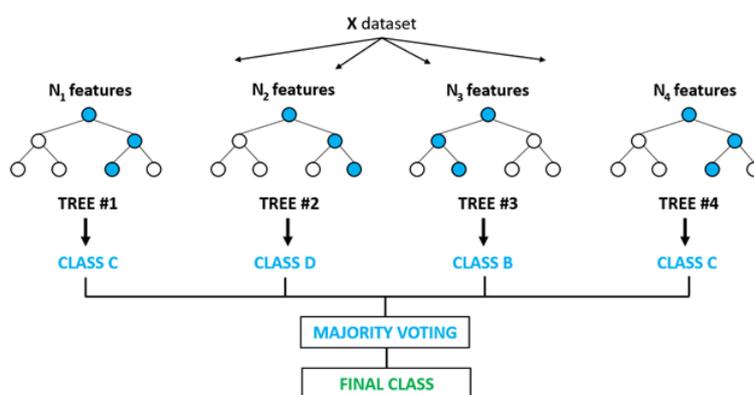


Figura 3.19. Struttura dell'algoritmo Random Forest

Il Random Forest oltre a predire la classe, determina anche un punteggio di importanza associato ad ogni variabile. Questi punteggi permettono di individuare le feature che secondo l'algoritmo sono di maggiore importanza, eliminando le funzionalità che hanno poco impatto nella classificazione. In questo lavoro si è impostata una soglia di importanza che varia tra 0,1 e 0,2 in modo da estrarre mediamente 50 funzionalità per ogni insieme di feature.

Dopo aver selezionato le caratteristiche più importanti, sono state concatenate tali feature con quelle manuali. Infine, tutte le caratteristiche sono state normalizzate attraverso la tecnica del min-max scaling che permette di migliorare il confronto diretto tra valori con grandezze molto diverse.

## 3.7 Metodi di classificazione

L'ultima parte del progetto consiste nella classificazione delle feature che sono state estratte in precedenza. La classificazione viene eseguita sia per le sole feature profonde estratte dalle architetture CNN, sia per l'unione delle feature profonde e quelle estratte manualmente. Nel primo caso, la classificazione avviene mediante una rete neurale semplice (ANN), mentre nel secondo caso si concatenano i due gruppi di funzionalità che vengono classificati attraverso tre algoritmi di machine learning, spesso utilizzati in letteratura: un SVM, un KNN e un LDA [49] [53] [54]. I diversi metodi di classificazione sono stati scelti tra quelli più comunemente utilizzati in letteratura e verranno descritti più nel dettaglio nei prossimi paragrafi. Tutti gli algoritmi hanno come obiettivo comune la classificazione delle immagini nelle due classi benigna e maligna.

### 3.7.1 ANN

Le Artificial Neural Network o ANN sono stati discussi in precedenza nel capitolo 2.4.1. In questo progetto è stata utilizzata una rete neurale formata da tre layer completamente connessi. Il primo strato contiene un numero di neuroni pari al numero di feature in input, mentre il secondo e il terzo layer hanno rispettivamente 512 e 128 neuroni. Dopo ogni livello lineare è presente un livello di attivazione con funzione Relu e una dropout con probabilità di 0,7. Infine, il classificatore si conclude con una funzione di attivazione softmax che permette di trasformare i valori output in una distribuzione di probabilità normalizzata.

### 3.7.2 SVM

L'SVM, dall'inglese support-vector machines, è un metodo di apprendimento supervisionato utilizzato per la classificazione di feature estratte sia manualmente che attraverso deep learning. Questa tecnica si basa sull'individuazione di una funzione che permetta la separazione degli elementi di classi diverse. Il confine che separa le classi è detto confine decisionale (decision boundary). Il modello rappresenta i dati di input come punti nello spazio, mappati in modo tale che gli elementi appartenenti a due classi differenti siano separabili da uno spazio il più possibile ampio. Mentre gli elementi che si posizionano lontano dall'iperpiano vengono classificati con molta probabilità nella classe corretta, gli elementi vicini ad esso rischiano di essere misclassificati in base all'aggiunta di nuovi punti. L'obiettivo dell'algoritmo è quello di avere una elevata distanza tra i punti più vicini (support vector) e il confine decisionale dal momento che maggiore è la distanza, migliore è la generalizzazione dell'algoritmo nel classificare nuovi elementi.

L'equazione dell'iperpiano è del tipo  $wx - b = 0$  e viene trovata risolvendo un problema di ottimizzazione con il fine di massimizzare la distanza tra i support vector.

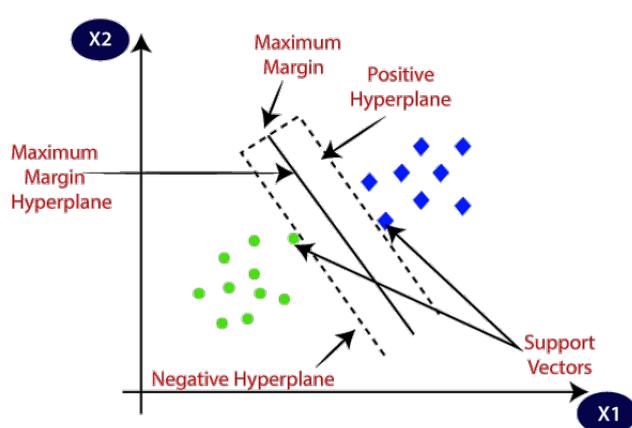


Figura 3.20. Esempio di divisione in due classi attraverso il classificatore SVM

### 3.7.3 KNN

Il k-nearest neighbors (k-NN) è uno dei più semplici metodi di classificazione tra quelli utilizzati nell'apprendimento automatico. È un algoritmo che classifica i dati in base alle caratteristiche degli oggetti vicini a quello in esame. Per stabilire la vicinanza tra gli elementi, quest'ultimi che vengono rappresentati come dei punti nello spazio multidimensionale. La classificazione di un nuovo elemento avviene sulla base della classe più frequente tra i k punti più vicini. La vicinanza si basa sul calcolo della distanza tra punti (solitamente si utilizza la distanza euclidea, ma anche altri tipi di distanza sono ugualmente utilizzabili).

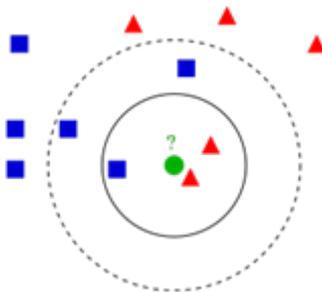


Figura 3.21. Esempio di classificazione di un nuovo elemento attraverso il metodo di classificazione k-NN

Di fondamentale importanza è, quindi, il parametro che viene definito in fase di progettazione.  $k$  è un intero positivo tipicamente non molto grande. Se  $k$  fosse pari a uno, allora l'oggetto verrebbe assegnato alla classe dell'elemento più vicino ad esso. In un contesto binario in cui sono presenti esclusivamente due classi, è opportuno scegliere  $k$  dispari per evitare di ritrovarsi in situazioni di parità. Generalmente, all'aumentare di  $k$  si riduce il rumore e l'overfitting, ma il criterio di scelta per la classe diventa più labile. Considerando solo i voti dei  $k$  oggetti vicini c'è l'inconveniente dovuto alla predominanza delle classi con più oggetti. In questo caso può

risultare utile pesare i contributi dei vicini in modo da dare, nel calcolo della media, maggior importanza in base alla distanza dall'oggetto considerato. La scelta di  $k$  dipende dalle caratteristiche dei dati. Infine, dal momento che questo algoritmo si basa sulla distanza, è necessario normalizzare i dati perché altrimenti grandezze fisiche molto diverse possono impattare sulla classificazione dell'oggetto. In questo progetto è stato implementato l'algoritmo con  $k$  pari a 7 sulla base della distanza euclidea.

### 3.7.4 LDA

L'analisi discriminante lineare (LDA), è un metodo che ha come fine quello di trovare una combinazione lineare di caratteristiche che separi il più possibile le classi in esame. Viene spesso utilizzato quando il dataset non è ampio grazie alla capacità dell'algoritmo di generalizzare. Per classificare gli elementi, questo metodo proietta le caratteristiche di input da uno spazio di dimensioni superiori ad uno di dimensioni inferiori. Per far ciò è necessario creare un nuovo asse su cui verranno proiettate le variabili. Per ridurre la dimensionalità il nuovo asse deve poter possedere i seguenti criteri: massimizzare la distanza tra le medie delle due classi e minimizzare la varianza all'interno della singola classe. Il valore medio di ciascun input per ciascuna delle classi può essere calcolato dividendo la somma dei valori per il numero totale di valori, mentre la varianza viene calcolata in tutte le classi come media del quadrato della differenza di ciascun valore dalla media. I modelli LDA si basano sul calcolo della probabilità che un nuovo set di dati di input appartenga ad una delle classi di output. La funzione utilizzata per il calcolo di questa probabilità è denominata funzione discriminante, implementata come:

$$d(x)_i = \log(f(x)_i) + \log(p_i) \quad (3.13)$$

Nella fase di classificazione, per ogni nuovo dato si calcola il valore della funzione di appartenenza ad ogni classe e il valore maggiore rappresenta la classe a cui verrà assegnato.

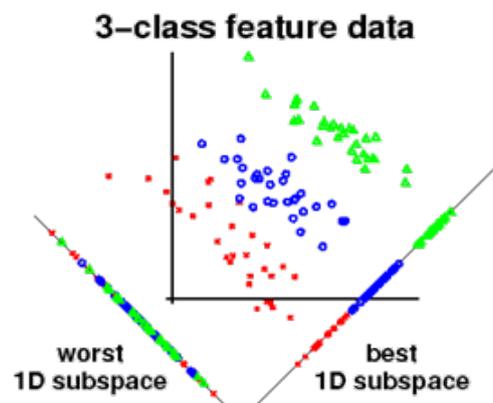


Figura 3.22. Esempio di divisione in tre classi attraverso il classificatore LDA

## 3.8 Metodi di valutazione

Per quantificare numericamente la bontà dei risultati di classificazioni verranno utilizzate le metriche di validazione discusse nei seguenti paragrafi. In generale, nei problemi di classificazione, si confronta la classe attesa con quella predetta dall'algoritmo. I risultati verranno comparati nel capitolo 4 con le annotazioni manuali, chiamate ground truth, realizzate da due senologhe esperte.

### 3.8.1 Confusion matrix

La confusion matrix o tavola di contingenza è la metrica di validazione più comune per valutare le performance di un algoritmo nell'ambito dell'intelligenza artificiale. Come si può vedere in Figura 3.23, è formata da una tabella che contiene sulle righe la classe reale (actual condition), mentre sulle colonne la classe predetta dal modello (prediction condition). La matrice ha dimensioni  $n \times n$  dove  $n$  è il numero di classi totali del dataset che in questo lavoro è pari a due: lesione benigna e lesione maligna. Se la classificazione è binaria, le due classi sono indicate come positiva (positive) e negativa (negative). Alla classe positiva, che in questo progetto è associata alla classe benigna, viene anche indicata come 0, mentre la classe negativa, che nel progetto è associata alla classe maligna, viene indicata come 1.

		Predicted condition	
		Positive (PP)	Negative (PN)
Actual condition	Total population = P + N		
	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Figura 3.23. Confusion matrix LBP

All'interno della tabella sono riportate:

- il numero di elementi corretti classificati lungo la diagonale principale. In particolare, TN e TP che sono, rispettivamente, i corretti classificati per la classe positiva e i corretti classificati per la classe negativa;
- il numero di elementi misclassificati (ovvero gli errori di classificazioni commessi dall'algoritmo) lungo l'altra diagonale. In particolare, FN e FP che sono, rispettivamente, i casi che sono stati predetti negativi anche se nella realtà sono positivi e viceversa.

A partire dai valori presenti nella confusion matrix, si possono calcolare diverse metriche di valutazione. Le principali sono le seguenti:

- **Accuratezza:** è la metrica più utilizzata per un problema di classificazione e rappresenta la proporzione delle previsioni corrette sul totale.

$$\text{Accuratezza} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.14)$$

- **Recall/ Sensibilità/ True positive rate (TPR):** rappresenta la proporzione tra le previsioni positive correttamente classificate (TP) sul totale delle istanze positive reali.

$$\text{Recall}(TPR) = \frac{TP}{TP + FN} \quad (3.15)$$

- **Precisione/ Valore predittivo positivo (PPV):** rappresenta la proporzione delle previsioni positive correttamente classificate (TP) sul totale delle previsioni positive (sia reali che predette).

$$\text{Precisione}(PPV) = \frac{TP}{TP + FP} \quad (3.16)$$

- **Specificità/ True negative rate (TNR):** rappresenta la proporzione delle previsioni negative correttamente classificate (TN) sul totale delle istanze negative reali.

$$\text{Specificità}(TNR) = \frac{TN}{TN + FP} \quad (3.17)$$

- **Valore predittivo negativo (NPV):** rappresenta la proporzione delle previsioni negative correttamente classificate (TN) sul totale delle previsioni negative (sia reali che predette).

$$NPV = \frac{TN}{TN + FN} \quad (3.18)$$

- **F1 score:** rappresenta la media armonica ponderata delle metriche Precisione e Recall.

$$F1score = \frac{2 \dots \text{RecallPrecision}}{\text{Recall} + \text{Precisione}} \quad (3.19)$$

### 3.8.2 Curva ROC e Area Under Curve (AUC)

La curva ROC è una curva di probabilità che permette di valutare le performance di un classificatore binario. È raffigurata su un piano cartesiano in cui l'asse x rappresenta la sensibilità (True Positive Rate) e l'asse y rappresenta 1-specificità (False Positive Rate). ROC è l'acronimo di Receiver Operating Characteristic in quanto il grafico mette in relazione le due caratteristiche operative (veri positivi e falsi positivi) in funzione della soglia di discriminazione impiegata durante la classificazione. Nella Figura 3.24 si possono vedere degli esempi di curve ROC.

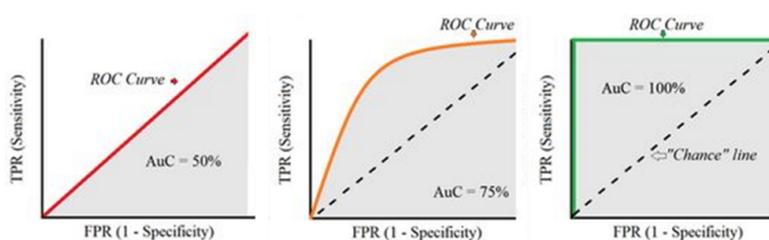


Figura 3.24. Esempi di curva ROC: nel primo grafico è rappresentata la retta a scelta casuale, nel secondo grafico è rappresentata una generica curva ROC, nel terzo grafico è rappresentata una curva ROC associata ad un classificatore perfetto

Il primo grafico riporta quella che è definita la retta a scelta casuale (retta rossa bisettrice del quadrante) che descrive il comportamento di un classificatore che classifica gli elementi in maniera totalmente casuale. Questa retta è importante perché determina il limite di performance accettabili. Un algoritmo, infatti, per avere delle buone prestazioni, deve avere una curva ROC che si trova al di sopra di tale retta. Più la curva sarà lontana dalla retta a scelta casuale e più vicina all'angolo in alto a sinistra del piano e maggiore sarà la qualità del modello. Questo perché la curva verde (terzo grafico) che raffigura il classificatore ideale presenta un valore di TPR pari a uno, ovvero una sensibilità del 100% e un valore di FPR pari a zero, ovvero una specificità del 100%.

A partire da questo grafico si può calcolare un ulteriore parametro chiamato AUC, acronimo di Area Under Curve. Come dice il nome, determina l'area totale di piano sottesa alla curva ROC dando una informazione quantitativa sulla capacità del classificatore di separare le due classi. È un parametro molto comune in quanto indipendente dalla soglia di classificazione. Il massimo valore di AUC è 100%, condizione ideale che viene rappresentata dalla curva verde riportata nel terzo grafico. Minore sarà il valore di AUC e peggiori saranno le performance del modello analizzato. In generale, un buon classificatore è 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.

### 3.8.3 k-fold

Come riportato nel paragrafo 3.1, per applicare nel modo corretto gli algoritmi di apprendimento automatico, è necessario dividere il set di dati in Training, Validation e Test set. Il problema di questa strategia è che il modello potrebbe essere influenzato dalla divisione dei dati. Alcune immagini particolarmente complesse o con caratteristiche insolite, a seconda del set di dati a cui appartengono, possono far variare le performance dell'algoritmo. Per questo motivo, si introduce un nuovo metodo di convalida chiamato k-fold cross validation (convalida incrociata K-fold) che valuta le performance del modello utilizzando diversi set di dati. Questa tecnica prevede di dividere il dataset in K blocchi e a rotazione utilizzare un blocco come Test set. In questo lavoro si è deciso di implementare un k-Fold con k pari a 5 che comporta una divisione del dataset in 5 gruppi e l'esecuzione di 5 allenamenti. In particolare, per ogni allenamento si utilizza un blocco come Test set e gli altri blocchi come Training e Validation set.



Figura 3.25. Schema riguardante la suddivisione del dataset durante l'applicazione del metodo k-fold



# Capitolo 4

## Risultati

Nel presente capitolo vengono illustrati i risultati ottenuti a seguito dell'applicazione dei metodi discussi nel capitolo 3. La trattazione prevede l'analisi delle performance ottenute dai diversi classificatori, in una prima fase allenati sulle sole feature estratte attraverso gli algoritmi di deep learning e in una seconda fase allenati sul vettore di feature concatenate che comprendono sia quelle automatiche sia quelle estratte manualmente. Per ogni allenamento eseguito vengono calcolate (su tutti e tre i set di dati) le metriche che sono state descritte nel paragrafo 3.8. In questo capitolo, per evitare di rendere i risultati di difficile comprensione, verranno però graficamente solamente le tre metriche seguenti:

- ACC o accuratezza: importante per una valutazione generale della qualità del classificatore definendo il numero dei corretti classificati sul numero totale di elementi presi in considerazione;
- PPV o precisione: importante in quanto se tendente a 1 è indice di un numero ristretto di falsi positivi (ovvero pochi tumori maligni vengono identificati come benigni)
- AUC: importante in quanto unisce il valore di altre due metriche ovvero il TPR e il NPR. Maggiore è il valore di tale metrica e minore sarà il numero di falsi classificati.

I risultati delle classificazioni vengono riportati e confrontati al fine di indagare quale strategia e quale modello porti a raggiungere le migliori performance sul dataset a disposizione.

## 4.1 Classificazione sulla base delle feature locali

Le quattro architetture CNN (AlexNet, VGG, ResNet e DenseNet), oltre ad estrarre le feature profonde (successivamente concatenate con quelle manuali), vengono anche allenate al fine della classificazione delle lesioni. In questo paragrafo vengono riportate le prestazioni di tali modelli in funzione dei parametri impostati dall'utente.

Durante la fase di addestramento, le reti vengono monitorate attraverso il calcolo dell'accuratezza e della loss function. Questi parametri vengono graficati in funzione delle epoche in modo da valutare come e quanto apprende la rete durante la fase di allenamento e se è presente il fenomeno di overfitting. Un esempio è riportato nella Figura 4.1.

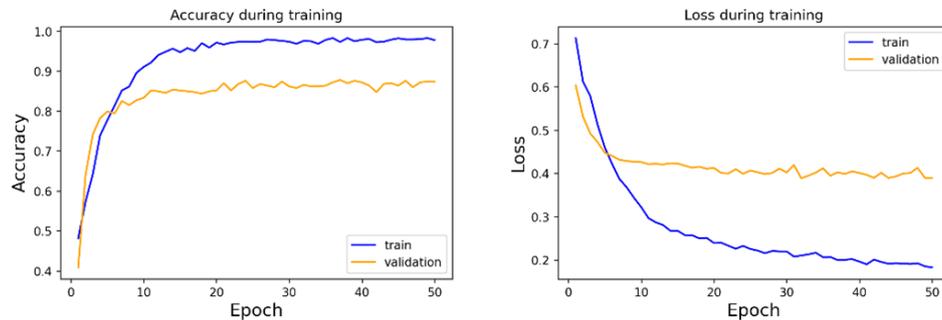


Figura 4.1. Monitoraggio di accuratezza e loss function durante l'allenamento

Le reti sono state allenate per un totale di 50 epoche, utilizzando la cross-entropy come funzione di costo e Adam come algoritmo di ottimizzazione. Inizialmente si applica un batch size pari a 8, un learning rate pari a 0,001, e non si introduce regolarizzazione. Questi parametri verranno variati nel corso degli allenamenti al fine di ottenere la configurazione che porta alle migliori performance.

### 4.1.1 Batch size

Come anticipato nel capitolo precedente, le architetture vengono allenate dividendo il dataset in batch di diversa numerosità. In particolare, si effettuano tre prove con batch rispettivamente di 8, 16 e 32 immagini. In seguito, vengono riportati due grafici che mostrano l'accuratezza e l'AUC medi su tutte le architetture per i tre set di dati e per i tre batch size scelti (Figura 4.2).

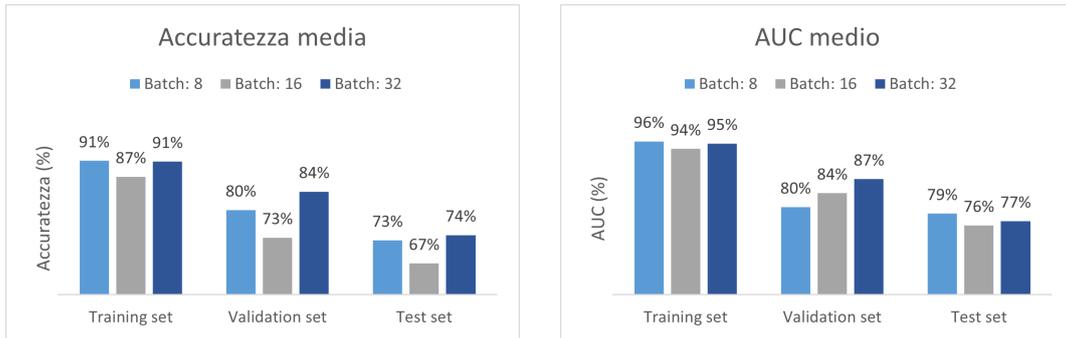


Figura 4.2. Accuratezza e AUC mediati su tutte le architetture per i tre set di dati e per i tre batch size scelti

Si può notare che a parità di architettura, le performance presentano un leggero miglioramento quando il batch size è pari a 32. Nei grafici seguenti (Figura 4.3) vengono, quindi, riportati i risultati ottenuti con tale valore di batch size per i diversi set di dati al fine di confrontare le prestazioni delle quattro architetture allenate.



Figura 4.3. Confronto in termini di accuratezza, precisione e AUC tra le diverse architetture quando si utilizza un batch size pari a 32

Si può osservare che per i set di dati di allenamento e di validazione, tutte e tre le metriche risultano essere migliori nelle architetture ResNet e DenseNet rispetto ad AlexNet e Vgg. Le prestazioni sul Test set invece non hanno un andamento fisso,

ma presentano valori fino al 30% più bassi rispetto a quelli del Training set. Questo dimostra la presenza del fenomeno di overfitting (particolarmente evidente nelle architetture ResNet e DenseNet). In base a questa considerazione vengono proposti approcci che puntano a ridurre tale differenza di performance.

#### 4.1.2 Data augmentation

Una prima variazione al fine di limitare l'overfitting consiste nell'aumentare il dataset a disposizione. Dal momento che non si dispone di ulteriori immagini, si è effettuato un aumento virtuale dei dati. Le prestazioni degli allenamenti effettuati con il set di dati aumentato con tecnica offline sono state calcolate sulla media delle predizioni delle immagini aumentate (ovvero si effettua le predizioni su tutte le immagini e si media il risultato sulle stesse immagini trasformate, a questo punto si calcolano le metriche).

Di seguito vengono confrontati gli addestramenti effettuati senza aumento dei dati, con aumento dei dati online e con aumento dei dati offline (Figura 4.4). I grafici mostrano le performance mediate sulle quattro architetture per tre set di dati. Tutti gli allenamenti sono stati effettuati ponendo un batch size pari a 32.

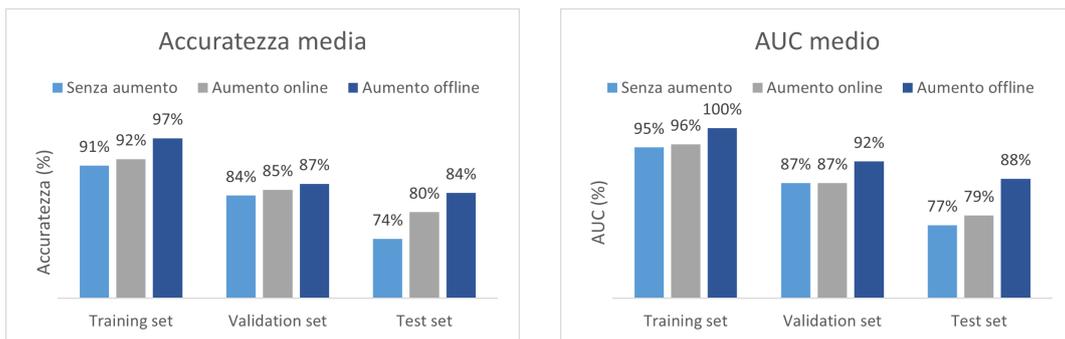


Figura 4.4. Accuratezza e AUC mediate su tutte le architetture per i tre set di dati e per i tre data augmentation utilizzati

Dall'osservazione dei dati, è evidente che l'aumento delle immagini migliori le performance su tutti i set (Training, Validation e Test set). Entrambi i tipi di data augmentation portano accuratezze migliori rispetto al dataset originale, ma la tecnica offline ottiene mediamente, in termini di accuratezza, il 2% in più sul Validation set e il 5% in più sul Test set a rispetto all'aumento dei dati online.

Nella figura 4.5 vengono riportati i valori delle metriche che si ottengono allenando le reti sul dataset aumentato con tecnica offline. I risultati vengono riportati in modo distinto per le diverse architetture per un migliore confronto delle reti.



Figura 4.5. Confronto in termini di accuratezza, precisione e AUC tra le diverse architetture quando si effettua data augmentation offline

Si osserva che l'architettura DenseNet risulta essere la migliore, con valori di AUC che arrivano o superano il 90% su tutti e tre i set di dati. Mediamente il valore di AUC è pari al 100% sul Training set, al 92% sul Validation set e all'88% sul Test set. Si nota ancora, soprattutto in termini di accuratezza, la presenza di overfitting. È per questo motivo che i prossimi passaggi prevedono la variazione del learning rate e l'applicazione del termine di regolarizzazione.

### 4.1.3 Learning Rate e Regolarizzazione

Come anticipato nei paragrafi 3.4.5 e 3.4.6, sia il tasso di apprendimento (o learning rate) sia l'aggiunta di un termine di regolarizzazione nella funzione di costo sono due parametri fondamentali per l'addestramento al fine di limitare l'overfitting e aumentare le performance. Si è deciso quindi di modificare tali parametri nel seguente modo:

- in primo luogo, si è diminuito il tasso di apprendimento da 0,001 (valore utilizzato negli allenamenti precedenti) a 0,0001 con passi di 0,0005
- in secondo luogo, si è introdotto il termine di regolarizzazione L2 con parametro lambda pari a  $10^{-3}$  e  $10^{-5}$ .

Nelle Figure 4.6 e 4.7 vengono confrontate le metriche accuratezza e AUC. Viene riportata la media sulle quattro architetture di tali metriche per i tre set di dati e per gli addestramenti effettuati con i diversi valori di learning rate e valori lambda

del termine di regolarizzazione. Tutti gli allenamenti sono stati effettuati ponendo un batch size pari a 32 ed applicando data augmentation offline.



Figura 4.6. Accuratezza e AUC mediati su tutte le architetture per i tre valori di learning rate utilizzati

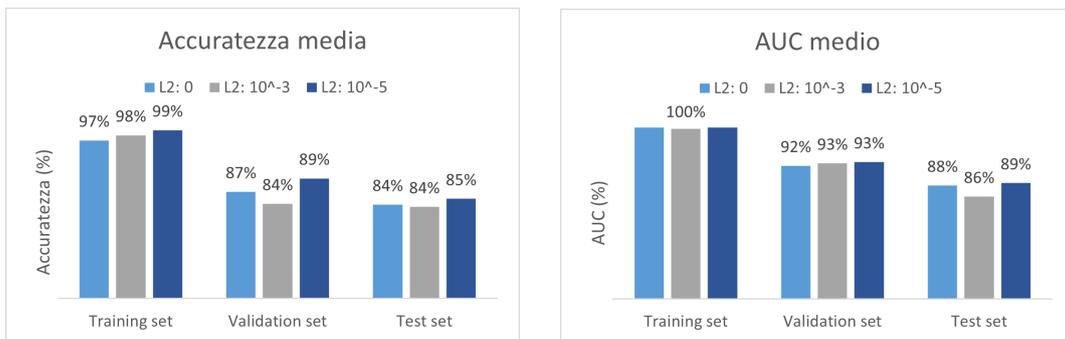


Figura 4.7. Accuratezza e AUC mediati su tutte le architetture per i tre valori di lambda utilizzati

Dall'analisi dei risultati, si nota che entrambi gli approcci hanno poco impatto sulle prestazioni medie degli algoritmi. Sul Training e sul Validation set le variazioni di accuratezza e AUC sono inferiori al 2%, mentre sono leggermente più importanti le modifiche sul Test set. In particolare, si nota un valore maggiore di AUC sul Test set quando si pone il learning rate pari a 0,0005, mentre l'accuratezza cresce quando si aumenta l'impatto del termine di regolarizzazione. Si è scelto come modello migliore, quindi, quello con combinazione di tasso di apprendimento e lambda pari rispettivamente a 0,0005 e  $10^{-5}$ .

Nei grafici seguenti (Figura 4.8) vengono riportati i valori delle metriche ottenuti allenando le reti sul dataset su tutte le architetture con tale combinazione di parametri. Si riportano i valori delle metriche in modo distinto tra le diverse architetture per migliore confronto delle reti.



Figura 4.8. Confronto in termini di accuratezza, precisione e AUC tra le diverse architetture quando si utilizza un learning rate pari a  $0,0005$  e un lambda pari a  $10^{-5}$

Si può notare che tutte le reti a parte AlexNet mostrano accuratezza pari al 100% e 90% rispettivamente sul Training e sul Validation set, mentre sul Test set la rete migliore in termini di accuratezza risulta essere DenseNet con una percentuale pari all'87%. Per quanto riguarda la precisione l'architettura ResNet è quella che presenta prestazioni migliori con valori pari al 100% sul Training set e pari al 90% sugli altri due set di dati (Validation e Test). Infine, la rete AlexNet supera in termini di Area Under Curve le altre architetture con un picco del 91% sul Test Set.

#### 4.1.4 Cross-validazione

Come discusso nel paragrafo 3.8.3 è necessario applicare la convalida incrociata k-fold per valutare se le performance ottenute con un set di dati ristretto (come quelle fornite nei paragrafi precedenti) possono essere estendibili al dataset intero in modo da valutarne la generalizzabilità. Le performance ottenute mediante convalida incrociata (eseguita utilizzando un valore di k pari a cinque) sono state mediate e vengono riportate nella Figura 4.9.



Figura 4.9. Confronto in termini di accuratezza, recall, precisione e AUC tra le diverse architetture mediando le performance ottenute su ogni fold

Si nota che il comportamento delle prestazioni mediate sui cinque fold rispetto a quelle del singolo fold (analizzate nel paragrafo precedente) varia a seconda del dataset considerato. In dettaglio:

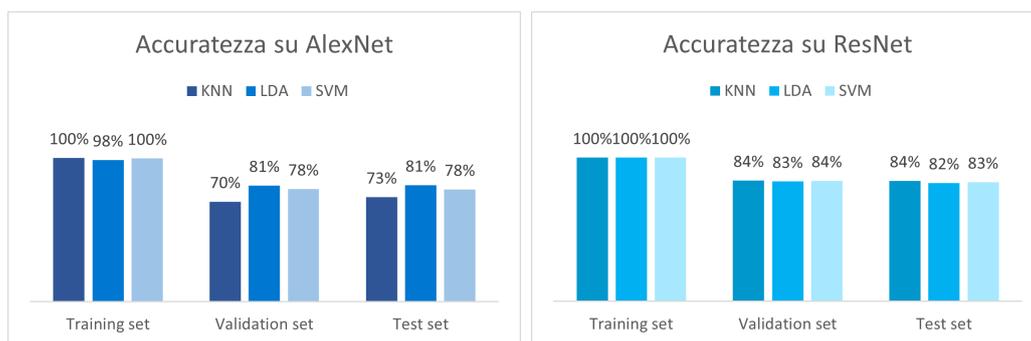
- sul Training set i valori delle metriche non hanno grandi variazioni attestandosi attorno al 100% su tutte le architetture;
- sul Validation set notiamo un importante abbassamento delle performance rispetto al singolo fold di prova pari a circa il 5% in termini di accuratezza e AUC e al 3% in termini di precisione;
- anche sul Test set le performance diminuiscono, anche se meno evidente rispetto al Validation set. La particolarità di questo set risulta essere la similitudine tra le prestazioni delle diverse architetture che si attestano intorno agli stessi valori (pari all'83,5% in termini di accuratezza, all'87% in termini di precisione e di AUC).

## 4.2 Classificazione sulla base delle feature concatenate

Come è stato anticipato nel paragrafo 3.2 le feature concatenate (ovvero l'unione delle feature estratte attraverso le reti neurali convoluzionali e quelle estratte manualmente attraverso i descrittori) vengono date in input a tre classificatori (KNN, LDA e SVM) che prevedono la classificazione benigna o maligna delle lesioni. Le feature di input hanno dimensione differente a seconda delle feature estratte dagli algoritmi di deep learning, nel dettaglio si hanno:

- 4096 feature locali estratte dall'architettura AlexNet che sommate alle 11 feature manuali estratte attraverso l'analisi di tessitura determinano un vettore di 4107 feature;
- 512 feature locali estratte dall'architettura ResNet che sommate alle 11 feature manuali estratte attraverso l'analisi di tessitura determinano un vettore di 523 feature
- 1024 feature locali estratte dall'architettura DenseNet che sommate alle 11 feature manuali estratte attraverso l'analisi di tessitura determinano un vettore di 1035 feature
- 4096 feature locali estratte dall'architettura Vgg che sommate alle 11 feature manuali estratte attraverso l'analisi di tessitura determinano un vettore di 4107 feature.

Ogni vettore di feature concatenato è stato utilizzato per addestrare i tre classificatori descritti nel paragrafo 3.7, le cui performance in termini di accuratezza vengono riportate nei grafici seguenti (Figura 4.10).



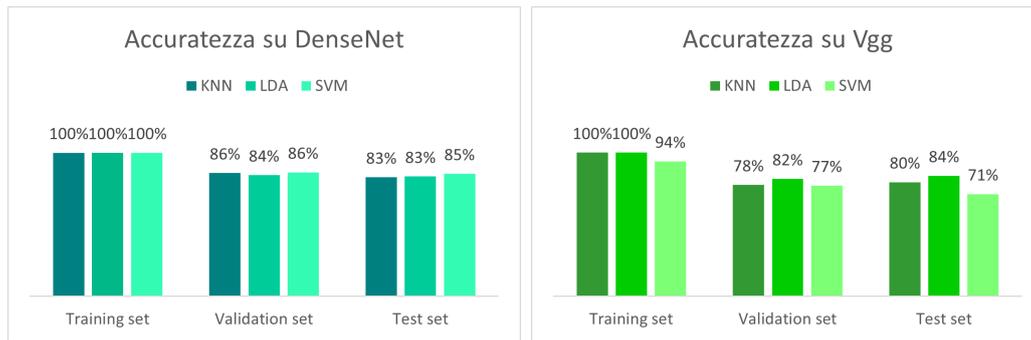


Figura 4.10. Accuratezza sui diversi set di dati per ogni combinazione di vettore di feature concatenato e classificatore utilizzata

Si può notare che le accuratze dei tre classificatori (knn, lda e svm) non si discostano molto quando applicati alle stesse feature. Se si analizzano le performance ottenute quando si utilizzano in fase di addestramento, oltre alle feature globali, quelle estratte dalle reti ResNet o DenseNet, si osserva che i valori non differiscono più del 2% tra i diversi classificatori. Nel caso di concatenazione con le feature estratte da AlexNet o Vgg, invece, si riscontrano variazioni maggiori pari al 5%-9%. Per le successive analisi si è scelto di mantenere i risultati ottenuti con il classificatore più performante applicato ad ogni vettore di feature. Si sono quindi scelti i seguenti classificatori:

- LDA nel caso di feature concatenate con quelle estratte dalle reti AlexNet e Vgg che permette di ottenere valori di accuratezza rispettivamente dell'81,0% e dell'83,9% sul Test set;
- KNN nel caso di feature concatenate con quelle estratte dalla rete ResNet che permette di ottenere valori di accuratezza sul Test set pari all'84,1%;
- SVM nel caso di feature concatenate con quelle estratte dalla rete DenseNet che permette di ottenere valori di accuratezza sul Test set pari all'85,4%.

#### 4.2.1 Feature selection

Vista la numerosità dei vettori concatenati si è deciso di effettuare feature selection per limitare il tempo di apprendimento degli algoritmi e per diminuire una eventuale rumorosità data dal numero elevato di dati di input. Viene quindi applicato l'algoritmo random forest che permette di ottenere una sessantina di feature per ogni vettore concatenato. Questi nuovi vettori sono stati dati in input agli stessi classificatori del caso precedente per analizzarne le performance. Nella Figura 4.11 riportata di seguito, vengono confrontati i risultati di classificazione in termini di accuratezza, precisione e AUC nel caso in cui si effettua feature selection e nel caso di vettori originali.

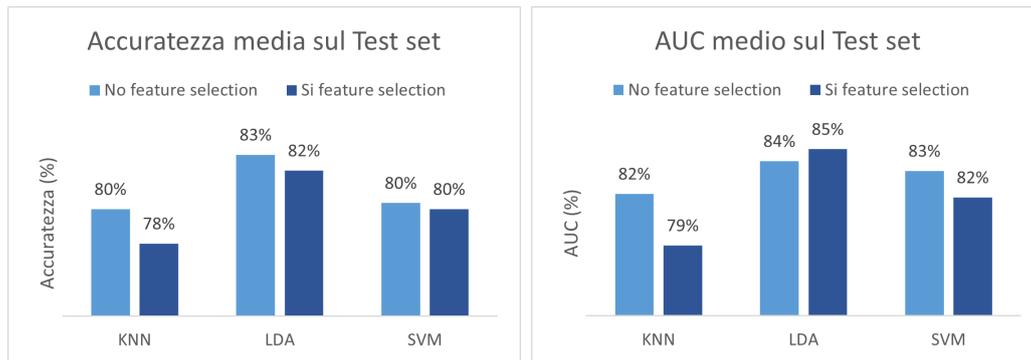


Figura 4.11. Accuratezza sui diversi set di dati per ogni combinazione di vettore di feature concatenato e classificatore utilizzata

Si può notare che tutti i classificatori hanno performance migliori quando non si effettua la selezione delle funzionalità. Si è per questo deciso di non introdurla per i successivi confronti.

### 4.3 Confronto tra classificazioni

Ottenute le performance delle due classificazioni appena analizzate (con solo feature locali e con le feature concatenate), si è effettuato un confronto tra i risultati dei migliori classificatori per determinare quale approccio portasse alle migliori prestazioni. Nella Figura 4.12 viene riportata l'accuratezza mediata sui cinque fold per i due allenamenti migliori corrispondenti ai due tipi di classificazioni.

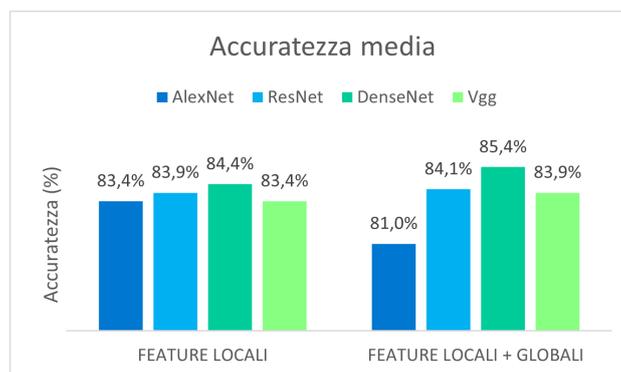


Figura 4.12. Accuratezza sul Test set mediata sui cinque fold nel caso in cui si classifichino solo le feature locali o nel caso in cui si classifichi la concatenazione di feature locali e globali

I dati non mostrano un andamento comune per tutte le architetture. Si osserva che le performance ottenute attraverso le sole feature locali sono molto simili tra le diverse reti (con una variazione di accuratezza pari all'1%), mentre se si utilizzano le feature concatenate l'accuratezza varia in modo più evidente tra l'81,0% e

l'85,4%. Le feature globali sembrerebbero dare un contributo e quindi migliorare l'accuratezza di classificazione quando già le performance sulle sole feature locali sono elevate, questo è l'esempio dell'architettura DenseNet che guadagna l'1% di accuratezza quando si utilizza il vettore concatenato. Per quanto riguarda le altre reti, che già presentavano performance meno elevate, attraverso l'introduzione delle feature globali ottengono un miglioramento contenuto. Solo le feature estratte attraverso l'architettura AlexNet risentono di un peggioramento dell'accuratezza del classificatore quando vengono aggiunte le feature manuali.

In questo contesto si può quindi determinare l'architettura, il tipo di feature e il classificatore che portano alle performance migliori. Si tratta della combinazione tra le feature globali e le feature estratte attraverso l'architettura DenseNet, successivamente classificate senza l'applicazione di feature selection attraverso il classificatore SVM. Vengono riportate nella tabella seguente tutte le metriche calcolate per ogni set di dati ottenute da questa combinazione di tecniche.

	ACC	TPR	TNR	PPV	NPV	F1	AUC
Training set	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Validation set	86,3%	93,8%	68,3%	87,8%	82,7%	90,7%	87,6%
Test set	85,4%	94,5%	70,0%	86,2%	83,6%	90,1%	86,6%

Tabella 4.1. Performance del classificatore migliore

## 4.4 Analisi delle misclassificazioni

Al fine di approfondire le limitazioni e gli sviluppi futuri del presente studio, sono state analizzate le immagini non correttamente classificate. L'analisi si è eseguita in collaborazione con gli specialisti del reparto di Senologia delle Molinette.

Inizialmente sono state esaminate quante e quali immagini sono state misclassificate dalle quattro architetture di deep learning, i cui risultati vengono descritti di seguito:

- il 48% dei falsi negativi è risultato comune ad almeno due architetture: il 10% dei casi viene classificato erroneamente da tre architetture su quattro, mentre il 13% dei casi non viene classificato correttamente da nessuna architettura;
- il 72% dei falsi positivi è risultato comune ad almeno due architetture: il 25% dei casi viene classificato erroneamente da tre architetture su quattro, mentre il 13% dei casi non viene classificato correttamente da nessuna architettura.

La stessa analisi è stata effettuata sulle misclassificazioni ottenute dagli algoritmi di machine learning allenati sulle feature concatenate, i cui risultati vengono descritti nuovamente di seguito:

- il 42% dei falsi negativi è risultato comune ad almeno due architetture: il 12% dei casi viene classificato erroneamente da tre architetture su quattro, mentre l'8% dei casi non viene classificato correttamente da nessuna architettura;

- l'87% dei falsi positivi è risultato comune ad almeno due architetture: il 23% dei casi viene classificato erroneamente da tre architetture su quattro, mentre il 42% dei casi non viene classificato correttamente da nessuna architettura.

Dal momento che un gran numero di falsi classificati risultano essere gli stessi per tutti gli algoritmi, si può concludere che alcune immagini risultano essere particolarmente difficili da riconoscere indipendentemente dalla tecnica utilizzata.



## Capitolo 5

# Conclusioni e Sviluppi futuri

I risultati migliori, in termini di variabilità e valore medio, si ottengono con il classificatore SVM applicato al vettore di feature concatenato. L'unione delle feature globali, descrittive della natura del tessuto mammario, con quelle estratte mediante gli algoritmi di deep learning, sembrano portare un leggero aumento delle prestazioni rispetto alla classificazione ottenuta attraverso le sole feature locali. Dall'84,4% di accuratezza ottenuto attraverso l'architettura DenseNet, si raggiunge l'85,4% quando si uniscono i due tipi di feature. Anche se il miglioramento non è elevato, dimostra che informazioni ulteriori oltre a quelle estraibili dalla sola lesione possono contribuire all'analisi del tumore. Questo approccio è proprio della valutazione clinica da parte dei medici che combinano le informazioni visive ricavate dalla zona tumorale con quelle relative alla natura del seno e della cartella clinica del paziente. Per un possibile sviluppo futuro si può quindi pensare di integrare l'analisi delle lesioni effettuata attraverso apprendimento profondo con ulteriori dati riguardanti sia la natura del tessuto della mammella che la storia clinica del paziente.

Questo progetto è la continuazione di un precedente lavoro che aveva come obiettivo sempre la classificazione della malignità delle lesioni, ma che si è concentrato sullo sviluppo di algoritmi di machine learning addestrati su feature manuali. Le tecniche introdotte nel nuovo lavoro hanno portato a prestazioni migliori rispetto ai semplici algoritmi di machine learning dimostrando che l'apprendimento profondo, come spesso riportato in letteratura, permette di raggiungere risultati migliori grazie al maggiore livello di astrazione delle caratteristiche.

A seguito dell'analisi dei falsi classificati condotta in collaborazione con le dottoresse dal reparto di Senologia delle Molinette si sono effettuate diverse considerazioni. In prima battuta, i classificatori non riescono a riconoscere lesioni che non sono state ben rappresentate nel dataset di Training. È bene quindi creare un dataset con una buona rappresentabilità, ovvero è necessaria la presenza di un buon numero di esempi per ogni tipologia di tumore. Così facendo gli algoritmi non devono riconoscere elementi molto diversi da quelli su cui si sono allenati. Ovviamente non è possibile

introdurre nel dataset qualsiasi tipo di lesione esistente (a causa dell'enorme variabilità anatomica), ma più il set di dati sarà completo e migliori saranno le prestazioni di classificazione.

Un'ulteriore considerazione riguarda le dimensioni delle lesioni. Si è notato che spesso gli algoritmi tendono a non classificare nel modo corretto lesioni con dimensioni molto piccole o molto grosse. Nel primo caso (ovvero quando le lesioni sono piccole) l'errore è stato reputato accettabile dai medici in quanto loro stessi di fronte a lesioni di questo tipo richiedono ulteriori accertamenti. Questo avviene perché non sempre, a causa delle dimensioni ridotte, si riescono ad identificare le caratteristiche salienti che ne permettono la discriminazione. Nel secondo caso (quando si considerano lesioni molto grandi) l'errore è stato reputato importante in quanto molto spesso questo tipo di lesioni appare più facilmente classificabile ai loro occhi. Dal momento che le dimensioni dei tumori sembrano giocare un ruolo cruciale nell'attendibilità dei risultati, sarà quindi necessario introdurre tale variabile in un ipotetico studio futuro. Per evitare di ridimensionare tutte le lesioni ad una dimensione comune, si possono attuare diverse strategie tra cui l'analisi di piccole zone dell'immagine quando il formato originale delle lesioni è molto grande oppure la costruzione e l'allenamento di classificatori diversi a seconda della grandezza dei tumori.

Il software implementato in questo studio si basa sull'individuazione e sul contornamento del tumore da parte dei medici all'interno dell'immagine. Sono state effettuate delle prove per valutare se piccole variazioni nella creazione della bounding box incidessero sulle performance del classificatore. Si è ottenuto che gli algoritmi sono indipendenti da tali variazioni, aspetto particolarmente importante che determina un'elevata usabilità dello strumento.

# Ringraziamenti

Al termine di questo progetto desidero ringraziare tutte le persone che in modo diverso hanno contribuito al mio percorso di crescita personale e professionale.

In primo luogo, un ringraziamento al Professore Filippo Molinari per aver suscitato in me sempre grande interesse verso le materie che trattava. Un ringraziamento particolare al Professore Massimo Salvi per avermi dato l'opportunità di realizzare questo progetto di tesi e per il suo costante supporto durante tutto il lavoro. Grazie per tutti gli utili consigli, le spiegazioni e gli insegnamenti di questi mesi. Un grazie a Matilde, che in qualità di rappresentante dell'azienda Tecnologie Avanzate, mi ha affiancato con la sua infinita disponibilità e pazienza.

Ringrazio tutti gli amici e in modo particolare il gruppo LAB, sempre presente in questi ultimi due anni per tutti i progetti di laboratorio e non solo. Un ringraziamento speciale anche a Giorgio, Virginia e Camilla che mi sono sempre stati accanto nei momenti di difficoltà e in quelli di gioia.

La mia più profonda gratitudine va, però, alla mia famiglia, a cominciare dai loro due pilastri: mia mamma e mio papà. Non sarei quello che sono oggi senza di voi e non avrei avuto la possibilità di raggiungere questo grande traguardo. Grazie per avermi insegnato che la più alta forma di libertà è perseguire i propri sogni nonostante il sacrificio e il coraggio necessari per affrontare tutte le avversità della vita. Un grazie, infine, a mio fratello Luca che mi ha aiutato a crescere e ha migliorato la mia vita dal momento in cui è nato.



# Bibliografia

- [1] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, “Representation learning for mammography mass lesion classification with convolutional neural networks,” *Computer methods and programs in biomedicine*, vol. 127, pp. 248–257, 2016.
- [2] A. I. R. Tumori, “I numeri del cancro in italia. 2020,” 2012.
- [3] X. Qin, G. Lu, I. Sechopoulos, and B. Fei, “Breast tissue classification in digital tomosynthesis images based on global gradient minimization and texture features,” in *Medical Imaging 2014: Image Processing*, vol. 9034, pp. 473–480, SPIE, 2014.
- [4] C. J. Vyborny and M. L. Giger, “Computer vision and artificial intelligence in mammography,” *AJR. American journal of roentgenology*, vol. 162, no. 3, pp. 699–708, 1994.
- [5] Berloco, “Anatomia del seno.” <https://marcoberloco.com/anatomia-del-seno>, 2016.
- [6] V. Volpicelli, “Mammografia..” <http://www.fertilitycenter.it/mammella/mammografia>, 2017.
- [7] “Specifiche tecniche..” <https://womenshealth-emea.fujifilm.com/ita/specifiche-tecniche>, 2022.
- [8] “Mammografia digitale.” <https://diagnosticaothman.it/servizi/mammografia-digitale>, 2022.
- [9] v. p. i. p. Manuale MSD, “Figure: Struttura tipica di una cellula nervosa..” <https://www.msmanuals.com/it-it/casa/multimedia/figure/struttura-tipica-di-una-cellula-nervosa>.
- [10] G. Buttazzo, “Reti neurali in grado di apprendere..” <http://retis.sssup.it/~giorgio/paps/2020/Ithaca20-Buttazzo.pdf>, 2020.
- [11] M. H. Esfe, S. A. Eftekhari, M. Hekmatifar, and D. Toghraie, “A well-trained artificial neural network for predicting the rheological behavior of mwent–al2o3 (30–70%)/oil sae40 hybrid nanofluid,” *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

- [12] G. Gullo, “Deep learning svelato: ecco come funzionano le reti neurali artificiali.” <https://medium.com/professionai/deep-learning-svelato-ecco-come-funzionano-le-reti-neurali-artificiali-f7dced21>, 2018.
- [13] N. Manchev, “Convolutional neural networks with pytorch | domino data lab.” <https://www.dominodatalab.com/blog/gpu-accelerated-convolutional-neural-networks-with-pytorch>, 2022.
- [14] A. V. Pratik Choudhari, “Understanding ‘convolution’ operations in cnn.” <https://medium.com/analytics-vidhya/understanding-convolution-operations-in-cnn-1914045816d4>, 2020.
- [15] A. Griguolo, “Tumore al seno: cause, sintomi, sopravvivenza e prevenzione.” <https://www.my-personaltrainer.it/salute/tumore-seno.html>, 2022.
- [16] A. B. D. Colcavagno, “Cenni di anatomia della ghiandola mammaria.” <https://www.idoctors.it/articolo/cenni-di-anatom/290/62>, 2013.
- [17] A. Marrazzo, P. Taormina, M. David, L. Casà, D. Lo Gerfo, A. Noto, *et al.*, “The role of sentinel lymph-node biopsy (slnb) in the treatment of breast cancer,” *Chir Ital*, vol. 58, no. 3, pp. 299–304, 2006.
- [18] “Tumore al seno: sintomi, prevenzione, cause, diagnosi.” <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-del-seno>, 2021.
- [19] “Tumore al seno - aimac - associazione italiana malati di cancro.” <https://www.aimac.it/schede-tumori/tumore-seno>, 2021.
- [20] “Tumore al seno domande e risposte dalla diagnosi al dopo cura.” <https://www.fondazioneveronesi.it/magazine/tools-della-salute/download/i-manuali/tumore-al-seno-domande-e-risposte-dalla-diagnosi-al-dopo-cura>, 2019.
- [21] “Cancer.” [https://www.who.int/health-topics/cancer#tab=tab\\_1](https://www.who.int/health-topics/cancer#tab=tab_1), 2019.
- [22] M. Masud, A. E. Eldin Rashed, and M. S. Hossain, “Convolutional neural network-based models for diagnosis of breast cancer,” *Neural Computing and Applications*, pp. 1–12, 2020.
- [23] “Screening per il cancro del seno.” <https://www.salute.gov.it/portale/tumori/dettaglioContenutiTumori.jsp?lingua=italiano&id=5542&area=tumori&menu=screening>, 2022.
- [24] “Screening per il tumore del seno.” <https://www.airc.it/cancro/prevenzione-tumore/guida-agli-screening/seno-mammografia>, 2022.

- [25] M. Løberg, M. L. Lousdal, M. Bretthauer, and M. Kalager, “Benefits and harms of mammography screening,” *Breast Cancer Research*, vol. 17, no. 1, pp. 1–12, 2015.
- [26] S. S. Nazari and P. Mukherjee, “An overview of mammographic density and its association with breast cancer,” *Breast cancer*, vol. 25, no. 3, pp. 259–267, 2018.
- [27] P. S. Sujlana, M. Mahesh, S. Vedantham, S. C. Harvey, L. A. Mullen, and R. W. Woods, “Digital breast tomosynthesis: Image acquisition principles and artifacts,” *Clinical imaging*, vol. 55, pp. 188–195, 2019.
- [28] N. Tirada, G. Li, D. Dreizin, L. Robinson, G. Khorjekar, S. Dromi, and T. Ernst, “Digital breast tomosynthesis: physics, artifacts, and quality control considerations,” *Radiographics*, vol. 39, no. 2, pp. 413–426, 2019.
- [29] L. Morra, D. Sacchetto, M. Durando, S. Agliozzo, L. A. Carbonaro, S. Delsanto, B. Pesce, D. Persano, G. Mariscotti, V. Marra, *et al.*, “Breast cancer: computer-aided detection with digital breast tomosynthesis,” *Radiology*, vol. 277, no. 1, pp. 56–63, 2015.
- [30] K. J. Geras, R. M. Mann, and L. Moy, “Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives,” *Radiology*, vol. 293, no. 2, p. 246, 2019.
- [31] P. Skaane, A. I. Bandos, E. B. Eben, I. N. Jebsen, M. Krager, U. Haakenaasen, U. Ekseth, M. Izadi, S. Hofvind, and R. Gullien, “Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images,” *Radiology*, vol. 271, no. 3, pp. 655–663, 2014.
- [32] A. Smith, “Synthesized 2d mammographic imaging theory and clinical performance motivation for synthesized 2d images,”
- [33] B. Zeng, K. Yu, L. Gao, X. Zeng, and Q. Zhou, “Breast cancer screening using synthesized two-dimensional mammography: A systematic review and meta-analysis,” *The Breast*, vol. 59, pp. 270–278, 2021.
- [34] “Screening per la prevenzione del tumore della mammella..” <https://www.ats-brescia.it/screening-per-la-prevenzione-del-tumore-della-mammella>, 2022.
- [35] R. J. McDonald, K. M. Schwartz, L. J. Eckel, F. E. Diehn, C. H. Hunt, B. J. Bartholmai, B. J. Erickson, and D. F. Kallmes, “The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload,” *Academic radiology*, vol. 22, no. 9, pp. 1191–1198, 2015.
- [36] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, “Introduction to machine learning, neural networks, and deep learning,” *Translational Vision Science & Technology*, vol. 9, no. 2, pp. 14–14, 2020.

- [37] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, “Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review,” *Expert Systems with Applications*, vol. 167, p. 114161, 2021.
- [38] D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE conference on computer vision and pattern recognition*, pp. 3642–3649, IEEE, 2012.
- [39] N. M. Pasquale Rosati, Roberto Colombo, *Istologia*. Milano: Edi.Ermes, 2006.
- [40] B. Q. Huynh, H. Li, and M. L. Giger, “Digital mammographic tumor classification using transfer learning from deep convolutional neural networks,” *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016.
- [41] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, S. Jastrzębski, T. Févry, J. Katsnelson, E. Kim, *et al.*, “Deep neural networks improve radiologists’ performance in breast cancer screening,” *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1184–1194, 2019.
- [42] C. Zhang, J. Zhao, J. Niu, and D. Li, “New convolutional neural network model for screening and diagnosis of mammograms,” *Plos one*, vol. 15, no. 8, p. e0237674, 2020.
- [43] X. Li, G. Qin, Q. He, L. Sun, H. Zeng, Z. He, W. Chen, X. Zhen, and L. Zhou, “Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification,” *European radiology*, vol. 30, no. 2, pp. 778–788, 2020.
- [44] H. Kibriya, R. Amin, A. H. Alshehri, M. Masood, S. S. Alshamrani, and A. Alshehri, “A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers,” *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [45] H. U. Rashid, T. Ibrikci, S. Paydaş, F. Binokay, and U. Çevik, “Analysis of breast cancer classification robustness with radiomics feature extraction and deep learning techniques,” *Expert Systems*, p. e13018, 2022.
- [46] Y. Wu, J. Wu, Y. Dou, N. Rubert, Y. Wang, and J. Deng, “A deep learning fusion model with evidence-based confidence level analysis for differentiation of malignant and benign breast tumors using dynamic contrast enhanced mri,” *Biomedical Signal Processing and Control*, vol. 72, p. 103319, 2022.
- [47] J. Alyami, T. Sadad, A. Rehman, F. Almutairi, T. Saba, S. A. Bahaj, and A. Alkhurim, “Cloud computing-based framework for breast tumor image classification using fusion of alexnet and glm texture features with ensemble multi-kernel support vector machine (mk-svm),” *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

- [48] S. Shi, “A novel hybrid deep learning architecture for predicting acute kidney injury using patient record data and ultrasound kidney images,” *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 1329–1345, 2021.
- [49] M. Yousefi, A. Krzyżak, and C. Y. Suen, “Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning,” *Computers in biology and medicine*, vol. 96, pp. 283–293, 2018.
- [50] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, “Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets,” *IEEE transactions on medical imaging*, vol. 38, no. 3, pp. 686–696, 2018.
- [51] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, “Deep convolutional neural networks for mammography: advances, challenges and applications,” *BMC bioinformatics*, vol. 20, no. 11, pp. 1–20, 2019.
- [52] R. Hao, K. Namdar, L. Liu, M. A. Haider, and F. Khalvati, “A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted mri using convolutional neural networks,” *Journal of Digital Imaging*, vol. 34, no. 4, pp. 862–876, 2021.
- [53] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine learning for medical imaging,” *Radiographics*, vol. 37, no. 2, p. 505, 2017.
- [54] H. Li, X. Meng, T. Wang, Y. Tang, and Y. Yin, “Breast masses in mammography classification with local contour features,” *Biomedical engineering online*, vol. 16, no. 1, pp. 1–12, 2017.