

POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Biomedica

Tesi di Laurea Magistrale

**Sviluppo di un modello basato
sulla combinazione di reti neurali
convoluzionali per il
riconoscimento e la localizzazione
del carcinoma polmonare in
vetrini istologici**



Relatore
Prof. Filippo Molinari
Ing. Massimo Salvi

Laureando
Desirè CASORELLI

MARZO 2021

Indice

Elenco delle figure	5
Elenco delle tabelle	8
Sommario	9
1 Introduzione	11
1.1 Anatomia respiratoria e polmone	11
1.2 Cancro del polmone	17
1.2.1 Cancro del polmone a piccole cellule	18
1.2.2 Cancro del polmone non a piccole cellule	18
1.2.3 Stadiazione	19
1.3 Diagnosi e analisi istologica del tumore al polmone	21
1.3.1 Anamnesi	21
1.3.2 Test di imaging	21
1.3.3 Biopsia	23
1.3.4 Analisi istologica	25
2 Materiali e metodi	31
2.1 Dataset	31
2.2 Architettura	32
2.3 Preprocessing	34
2.3.1 Estrazione manuale delle Tiles	34
2.3.2 Estrazione smart delle patch	35
2.4 Reti neurali convoluzionali CNN	45
2.4.1 Introduzione al Deep Learning e alle CNN	46
2.4.2 Architettura reti: VGG, ResNet50,DenseNet	50
2.4.3 Allenamento delle CNN	55
2.4.4 Fusione output delle reti con media ponderata	58
2.5 Testing: Creazione Heatmap	60
2.6 Metriche di validazione	62
2.6.1 Metriche di validazione delle performance delle reti	62

2.6.2	Metriche di validazione delle Heatmap	67
3	Risultati	69
3.1	Validazione delle reti	69
3.1.1	Risultato dell'allenamento della CNN	69
3.1.2	Risultati sul Training set	72
3.1.3	Risultati sul Test set	79
3.1.4	Risultati con fusione mediante media pesata	85
3.2	Validazione delle Heatmap	88
4	Conclusioni e sviluppi futuri	97
4.1	Conclusioni	97
4.2	Sviluppi futuri	98
	Bibliografia	101

Elenco delle figure

1.1	Organizzazione dell'apparato respiratorio [2]	12
1.2	Epitelio colonnare pseudostratificato [7]	15
1.3	Schemi semplificati di organizzazione	16
1.4	Principali cellule dell'epitelio alveolare [11]	16
1.5	Tipologie di carcinoma polmonare [15]	17
1.6	Posizione e morfologia del cancro del polmone a piccole cellule SCLC [15]	18
1.7	Posizione e morfologia del cancro NSCLC	19
1.8	Stadiazione del carcinoma polmonare [19]	21
1.9	Diagnosi mediante radiografia [20]	22
1.10	Diagnosi mediante TAC [20]	22
1.11	Diagnosi mediante PET [20]	23
1.12	Agobiopsia polmonare [20]	24
1.13	Biopsia broncoscopica [20]	24
1.14	Biopsia toracoscopica [21]	24
1.15	Toracotomia [22]	25
1.16	Ecoendoscopia [23]	25
1.17	Schema di preparazione del campione istologico [28]	26
1.18	Scanner Leica Biosystems e monitor per la visualizzazione delle WSI [31]	28
1.19	Esempi di artefatti istologici	29
2.1	Software ASAP: WSI con annotazioni manuali del patologo	32
2.2	Schema totale dell'algoritmo proposto	33
2.3	Pipeline per l'estrazione manuale delle Tiles	34
2.4	Creazione Bounding box con software ASAP	35
2.5	Tile estratta dal dataset	35
2.6	Pipeline di estrazione smart delle patch	36
2.7	Caricamento immagine e ottimizzazione	37
2.8	Blue Ratio image	37
2.9	Nuclei detection	38
2.10	White detection	39

2.11	Conversione immagine da RGB a HSV	40
2.12	Tissue fold enhancement	40
2.13	Grafico per la scelta della soglia ottimale	41
2.14	Maschera finale dell’algoritmo di tissue fold detection	42
2.15	Creazione mappe di focus	43
2.16	Mappa finale di focus	44
2.17	Patch estratte: in verde patch benigne, in arancione patch maligne, in rosso annotazione manuale	45
2.18	Schema di una generica rete neurale [44]	46
2.19	Struttura di una CNN [45]	47
2.20	Tre tipologie di transfer learning [47]	48
2.21	Possibili situazioni di training [48]	49
2.22	Skip connection [50]	52
2.23	Rappresentazione architettura DenseNet [52]	54
2.24	Architettura DenseNet [52]	54
2.25	Esempio di patch date in input alle reti	55
2.26	Pipeline creazione Heatmap	60
2.27	Heatmap su un campo della WSI	61
2.28	Confusion matrix per classificazione binaria [55]	63
2.29	Esempi di curve ROC [58]	65
2.30	Esempio di t-SNE [61]	66
2.31	Esempio di saliency map per identificare le parti più rilevanti per una CNN [66]	67
3.1	Curve di training VGG	70
3.2	Curve di training ResNet50	71
3.3	Curve di training DenseNet	71
3.4	Confusion matrix sul training con rete VGG	72
3.5	Curve ROC e PR della VGG	73
3.6	Distribuzione t-SNE del training set	73
3.7	Saliency map con rete VGG	74
3.8	Confusion matrix sul training con rete ResNet50	74
3.9	Curve ROC e PR della ResNet	75
3.10	Distribuzione t-SNE del training set	75
3.11	Saliency map con rete ResNet	76
3.12	Confusion matrix sul training con rete DenseNet	76
3.13	Curve ROC e PR della DenseNet	77
3.14	Distribuzione t-SNE del training set	78
3.15	Saliency map con rete DenseNet	78
3.16	Confusion matrix sul test set con rete VGG	79
3.17	Curve ROC e PR della VGG	80
3.18	Distribuzione t-SNE del test set	80

3.19	Confusion matrix sul test set con rete ResNet	81
3.20	Curve ROC e PR della VGG	82
3.21	Distribuzione t-SNE del test set	82
3.22	Confusion matrix sul test set con rete DenseNet	83
3.23	Curve ROC e PR della ResNet	84
3.24	Distribuzione t-SNE del test set	84
3.25	Confusion matrix con modello di insieme	85
3.26	Curve PR per il modello di insieme	87
3.27	Curve ROC per il modello di insieme	87
3.28	Metriche di validazione Heatmap con VGG considerando solo il tessuto	89
3.29	Metriche di validazione Heatmap con VGG con i pixel del vetrino .	89
3.30	Metriche di validazione Heatmap con ResNet50 considerando solo il tessuto	91
3.31	Metriche di validazione Heatmap con ResNet50 con i pixel del vetrino	91
3.32	Metriche di validazione Heatmap con DenseNet considerando solo il tessuto	92
3.33	Metriche di validazione Heatmap con DenseNet con i pixel del vetrino	93
3.34	Metriche di validazione Heatmap con fusione dei modelli consideran- do solo il tessuto	94
3.35	Metriche di validazione Heatmap con fusione dei modelli e con i pixel del vetrino	94
3.36	Maschere del modello finale di insieme	95

Elenco delle tabelle

2.1	Architettura VGG-16 e VGG-19	51
2.2	Architettura ResNet50	53
2.3	Suddivisione Dataset in training, validation e test.	55
2.4	Parametri di training VGG	56
2.5	Parametri di training ResNet50	57
2.6	Parametri di training ResNet50	58
3.1	Performance VGG sul training set	72
3.2	Performance ResNet sul training set	75
3.3	Performance DenseNet sul training set	77
3.4	Performance VGG sul test set	79
3.5	Performance ResNet sul test set	81
3.6	Performance DenseNet sul test set	83
3.7	Pesi delle reti sull'output finale	85
3.8	Performance modello di insieme sul training set e sul test set	86
3.9	Metriche Heatmap VGG	90
3.10	Metriche Heatmap ResNet50	92
3.11	Metriche Heatmap DenseNet	93
3.12	Metriche Heatmap con fusione delle reti	95

Sommario

Il cancro al polmone, oggi, a causa dei nuovi agenti patogeni, rappresenta la prima causa di decessi per neoplasia al mondo ed il secondo tumore per incidenza; dopo il cancro alla prostata per gli uomini e quello al seno per le donne. Nonostante le terapie sviluppate per la cura, il carcinoma polmonare rimane una neoplasia a prognosi quasi sempre sfavorevole con una sopravvivenza a 5 anni dei pazienti approssimativamente pari al 18%. Di conseguenza, la diagnosi precoce si rileva di fondamentale importanza per il successo dei trattamenti a cui sarà sottoposto il paziente. Il *golden standard* per la diagnosi del cancro è la biopsia polmonare con la successiva ispezione del tessuto effettuata da un anatomopatologo esperto. Tuttavia, degli studi precedenti hanno dimostrato che esiste una variazione inter-operatore nella classificazione del tessuto polmonare modesta, con valori di concordanza compresi tra 0.65-0.81. In aggiunta alle difficoltà di avanzamento di una diagnosi accurata, il carico di lavoro in termini di tempo e il numero ridotto di patologi, rappresentano problematiche non indifferenti. In quest'ottica, un sistema basato sull'intelligenza artificiale potrebbe porsi come soluzione o come supporto al medico. La digital pathology con l'utilizzo di scanner ad alta risoluzione, ha consentito la creazione di WSI su cui fosse molto più semplice osservare le caratteristiche tissutali direttamente dal monitor di un computer e inserire delle annotazioni manuali per la rilevazione del tumore. Tutto ciò ha facilitato lo sviluppo di algoritmi automatici basati su reti neurali per la classificazione delle immagini.

Lo scopo principale del progetto di tesi consiste nella realizzazione di un sistema automatico fondato su CNN per la detection del carcinoma polmonare. Il dataset utilizzato è quello della challenge ACDC@ LUNGHP (Automatic Cancer Detection and Classification in Whole-slide Lung histopathology), che si è tenuta in The IEEE International Symposium on Biomedical Imaging (ISBI), costituito da 200 WSI con colorazione H&E. Più precisamente sono state allenate tre architetture di reti neurali convoluzionali: VGG16, ResNet50 e DenseNet, e, in ultimo, combinate le loro predizioni con l'operazione di media pesata in funzione dell'accuratezza di ciascun modello. I training delle reti sono stati eseguiti su delle patch di dimensione 500x500 estratte tramite un algoritmo automatico di estrazione smart che tenga conto dei nuclei cellulari, dei pixel bianchi del vetrino e della presenza di eventuali artefatti. Le prestazioni del classificatore finale si sono rilevate molto incoraggianti

con un' *AUC* di 96.1% sul training set e di 95.7% sul test set. Infine, a partire dal modello di insieme, sono state realizzate le Heatmap sulle WSI appartenenti al test set con un *Dice score* medio del 80%. Il sistema sviluppato, al netto di successivi miglioramenti, può dunque essere sfruttato come supporto al patologo per una seconda opinione al fine di ridurre i tempi di analisi e giungere ad una diagnosi altamente accurata.

Capitolo 1

Introduzione

Nel capitolo di introduzione vengono presentate le caratteristiche anatomiche dell'apparato respiratorio ed in particolare del polmone con un focus sull'istologia, oggetto delle immagini su cui si erge il progetto di tesi. Segue una descrizione del carcinoma polmonare, specificandone gli aspetti epidemiologici (incidenza, mortalità e fattori di rischio) e le principali tipologie. Vengono successivamente presentati i metodi di diagnosi più utilizzati per la stadiazione con particolare attenzione alla biopsia polmonare. In ultimo, viene mostrato l'iter di preparazione del campione istologico con le principali colorazioni fino al conseguimento dell'immagine digitale (WSI).

1.1 Anatomia respiratoria e polmone

La funzione principale del sistema respiratorio è la respirazione, ovvero il processo di scambio di gas tra il corpo e l'esterno tale da garantire la corretta ossigenazione del sangue e l'eliminazione dell'anidride carbonica in eccesso. In generale la respirazione comprende due livelli distinti [1]

- **Respirazione interna:** detta anche respirazione cellulare, ha come funzione principale la generazione di ATP sfruttando l'ossigeno all'interno dei mitocondri (fosforilazione ossidativa) e producendo anidride carbonica (CO₂) come prodotto di scarto;
- **Respirazione esterna:** è il processo di scambio di ossigeno e anidride carbonica tra l'ambiente e i tessuti attraverso quattro stadi:
 1. Ventilazione polmonare con le fasi di inspirazione ed espirazione.
 2. Scambio diffusivo a livello delle cavità aeree.
 3. Trasporto dell'ossigeno e dell'anidride carbonica tra i polmoni e i tessuti per mezzo del sangue.

4. Scambio diffusivo dei gas tra il sangue e i tessuti.

Organizzazione dell'apparato respiratorio

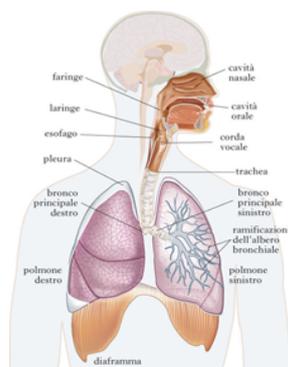


Figura 1.1: Organizzazione dell'apparato respiratorio [2]

L'apparato respiratorio si divide in due componenti principali: **le vie aeree superiori** (tratto respiratorio superiore) e **le vie aeree inferiori** (tratto respiratorio inferiore). In particolare, le vie aeree superiori includono i passaggi per l'aria collocati nella zona della testa e del collo. L'ingresso dell'aria nel corpo umano può avvenire a livello della cavità nasale e/o della cavità orale. Il naso con le sue cavità nasali rappresenta la principale apertura verso l'esterno per il flusso gassoso e il filtraggio dell'aria inalata. D'altra parte, la bocca, apertura secondaria dell'apparato respiratorio, ha il compito di sostenere e in alcuni casi di sostituire il naso nel processo di inalazione ed espulsione dell'aria. Entrambe le cavità conducono l'aria alla *faringe*, componente muscolo-membranosa della gola situata in fondo alla bocca. La faringe fa parte dell'apparato digerente e del respiratorio consentendo il transito del cibo e dell'aria. A sua volta, la faringe si divide in tre comportamenti dall'alto verso il basso:

1. Rinofaringe: sede delle tonsille adenoidi, si estende dalla base del cranio e termina all'altezza della superficie superiore del palato molle.
2. Orofaringe: sede delle tonsille linguali, palatine e di parte della lingua.
3. Ipofaringe: tratto inferiore della faringe, posizionato al di sotto dell'epiglottide, rappresenta la porzione della faringe in cui transita esclusivamente il cibo.

Dalla faringe le vie per il cibo e l'aria divergono: il cibo giunge all'esofago mentre l'aria entra nella laringe, ultimo tratto delle vie aeree superiori. Con la laringe inizia il tratto respiratorio che include tutte le vie di transito dell'aria sino ai polmoni. Da un punto di vista funzionale, il tratto respiratorio si compone di due zone distinte per lo spessore delle pareti:

1. **Zona di conduzione:** permette la conduzione dell'aria della laringe ai polmoni.
2. **Zona respiratoria:** comprende tutti i siti di scambio gassoso all'interno dei polmoni.

La laringe, primo organo della zona di conduzione, è un condotto di forma tubulare che include componenti cartilaginee come l'epiglottide e ospita le corde vocali. Durante la deglutizione l'epiglottide è mantenuta abbassata sulla glottide (apertura della laringe) impedendo al cibo e all'acqua di entrare nella trachea. La laringe è coinvolta nei processi di fonazione per mezzo della vibrazione delle corde vocali, e nella respirazione conducendo l'aria alla trachea [3]. Con la trachea iniziano le vie aeree inferiori che comprendono anche i *bronchi*, i *bronchioli*, i *polmoni* e i *muscoli della respirazione*. La trachea è un condotto elastico di 2.5 cm di diametro e 10 cm di lunghezza che decorre parallelamente e anteriormente all'esofago. A differenza di quest'ultimo che risulta sempre collassato, tranne che per la fase di deglutizione, la trachea rimane sempre aperta grazie alla presenza dei suoi 15-20 anelli cartilaginei a forma di 'C' che le conferiscono rigidità strutturale. L'epitelio che riveste la laringe e la trachea contiene cellule caliciformi che secernono muco e cellule ciliate. Le ciglia delle cellule ciliate si muovono al fine di spingere il muco con le particelle inglobate verso la faringe per essere deglutito [1].

In corrispondenza della cavità toracica, la trachea si divide nei bronchi primari extra polmonari sinistro e destro che conducono l'aria in ciascun polmone. Gli anelli cartilaginei dei bronchi si estendono per l'intera circonferenza a differenza della trachea. Superati i bronchi primari, all'interno di ogni polmone i bronchi si dividono in piccoli condotti detti bronchi secondari: tre bronchi secondari per i lobi del polmone destro e due per il sinistro. A loro volta, i secondari si dividono in terziari ed infine in bronchioli (dal diametro inferiore al millimetro), ottenendo circa 20-23 livelli di ramificazione [1]. I bronchioli dal punto di vista istologico non presentano una componente cartilaginea e per tale ragione possono collassare; al ridursi del diametro aumenta invece la componente muscolare liscia. Essi infine conducono ai *bronchioli terminali*. La parete interna dei bronchi e dei bronchioli è rivestita da epitelio ciliare e in misura minore da cellule mucose caliciformi che secernono muco al fine di catturare le particelle estranee dell'aria inalata.

Dai bronchioli terminali inizia la zona respiratoria dove hanno luogo gli scambi gassosi a livello polmonare. Le prime strutture sono i *bronchioli respiratori* che conducono, per mezzo dei *dotti alveolari*, agli *alveoli* (sede primaria degli scambi). Gli alveoli sono raggruppati tra loro in modo da formare una struttura a grappolo. Nel corpo umano ci sono circa 300 milioni di alveoli capaci di produrre 140 m² di superficie di scambio. Il diametro di un alveolo è compreso tra 200 e 500 μm . A livello degli alveoli polmonari avvengono gli scambi respiratori tra i gas polmonari dell'aria inalata e quelli del sangue. La diffusione gassosa è consentita dal sottile

spessore dell'epitelio alveolare all'interno del quale si riconoscono due tipologie di cellule epiteliali, dette pneumociti di tipo I e di tipo II.

All'interno delle superfici lumenali degli alveoli e nei dotti alveolari sono anche presenti i macrofagi, deputati alla rimozione delle sostanze dannose come il pulviscolo ambientale, i batteri e le particelle inquinanti. Gli alveoli comunicano tra loro mediante i pori di Kohn, presenti nella parete alveolare, in modo da mantenere in equilibrio la pressione dentro i polmoni [4].

Il polmone

I polmoni sono organi a forma di cono costituiti da tessuto spugnoso e collocati all'interno della cavità toracica. Pesano circa 1.3 kg insieme (680 g il destro e 620 g il sinistro): il polmone sinistro, più piccolo del destro, condivide lo spazio nel petto con il cuore e presenta un'impressione cardiaca per ospitare l'apice del cuore [5]. Insieme, i polmoni contengono circa 2400 km di vie aeree e sono divisi tra loro dal cuore e dal mediastino. Ogni polmone ha una base appoggiata sul diaframma e un apice arrotondato stretto che si estende superiormente raggiungendo l'estremità sternale dalla prima costola. Essi si estendono dalla spina dorsale nella cavità toracica fino alla parte anteriore del torace e procedendo dall'alto al basso, dall'estremità inferiore della trachea fino al diaframma. La superficie interna della parete toracica e quella esterna dei polmoni sono coperte rispettivamente da due pleure polmonari: parietale, più robusta e ricca di fibre, e viscerale, più sottile e trasparente. Tra le due è presente la cavità pleurica contenente un sottile strato di fluido pleurico lubrificante che consente agli strati di scivolare l'uno sull'altro nel corso della respirazione. Le pleure sono costituite da uno strato di cellule epiteliali e da tessuto connettivo elastico, formato da un intreccio di fibre di collagene ed elastina [6].

Ciascun polmone è diviso in lobi con la propria vascolarizzazione e ventilazione. I lobi sono tre a destra (superiore, medio ed inferiore) e due a sinistra (superiore ed inferiore) separati da scissure. Il polmone sinistro presenta una caratteristica analoga al lobo medio chiamata *lingula* (piccola lingua). Le scissioni sono due per il polmone destro: obliquo, divide il lobo superiore dall'inferiore, e orizzontale che divide il lobo medio da quello inferiore. Il polmone sinistro ha una sola scissione obliqua che divide il superiore dall'inferiore. I polmoni hanno un duplice apporto di sangue fornito dalla circolazione bronchiale e polmonare. La prima fornisce il sangue ossigenato mediante le arterie bronchiali (due a sinistra e una a destra) che lasciano l'aorta alle vie aeree dei polmoni. La circolazione polmonare trasporta, invece, il sangue deossigenato dal cuore ai polmoni e il sangue ossigenato al cuore. La circolazione polmonare inizia dall'arteria polmonare, che riceve il sangue povero di ossigeno dal ventricolo destro e poi si divide nelle due arterie, una per ogni polmone. Il sangue ossigenato torna all'atrio sinistro tramite le vene polmonari.

I polmoni sono innervati dal sistema nervoso autonomo. I nervi del polmone sono formati dai rami del nervo vago che contengono fibre broncostrittrici e vasodilatatrici, e dell'ortosimpatico con fibre broncodilatatrici e vasostrittrici. Nella fase dell'inspirazione prevale l'azione vasodilatatrice del vago e broncodilatatore dell'ortosimpatico, viceversa per l'espiazione [5].

Istologia polmonare

La zona di conduzione all'interno dei polmoni inizia nella trachea e si estende fino ai bronchioli terminali. Le cellule coinvolte in questa regione provvedono al riscaldamento dell'aria e alla rimozione delle particelle. Tutto questo tratto respiratorio è rivestito dall'epitelio respiratorio che per la maggior parte è un epitelio colonnare pseudostratificato ciliato costituito da cinque tipologie di cellule [7].

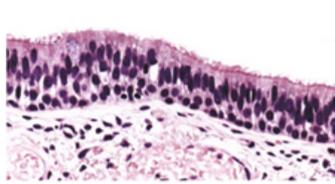


Figura 1.2: Epitelio colonnare pseudostratificato [7]

Cellule ciliate: sono la categoria più numerosa all'interno dell'epitelio. Le cellule ciliate spingono le particelle estranee verso la faringe per eliminarle attraverso la tosse o la deglutizione, costituendo il primo meccanismo di difesa dei polmoni. Questa tipologia cellulare è molto vulnerabile ad agenti nocivi come il fumo di sigaretta [1].

Cellule caliciformi: così chiamate per il loro aspetto a forma di calice, sono riempite con granuli di mucina sulla loro superficie apicale con il nucleo rivolto verso lo strato basale. Al ridursi del diametro dell'albero respiratorio vengono progressivamente sostituite dalle cellule di Clara.

Cellule basali: sono cellule triangolari con un vertice attaccato alla membrana basale e costituiscono l'attacco delle due tipologie cellulari precedenti. A differenza delle altre tipologie le cellule basali sopravvivono alle lesioni della mucosa per garantirne la sua ricostruzione.

Cellule a spazzola: dette anche pneumociti di tipo III. Possono essere cellule colonnari provviste di microvilli nello strato apicale. Si pensa che siano dei chemorecettori adibiti al monitoraggio della qualità dell'aria inalata.

Cellule neuroendocrine: dette anche cellule di Kulchitsky e costituiscono solo lo 0.5% della mucosa. Possono produrre serotonina, dopamina, noradrenalina e prodotti polipeptidici.

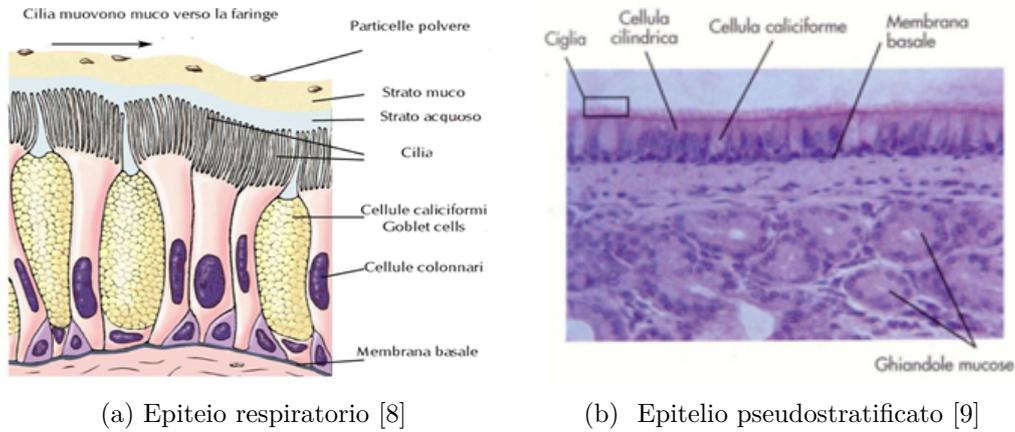


Figura 1.3: Schemi semplificati di organizzazione

L'epitelio alveolare è composto da due tipi di cellule [10]:

Cellule alveolari o pneumociti di tipo I: cellule squamose, sottili e piatte costituiscono il 90% - 95% dell'epitelio alveolare. Dato il loro spessore, si lasciano attraversare facilmente dai gas respiratori garantendo una maggiore facilità di scambio tra sangue ed aria per diffusione.

Cellule alveolari o pneumociti di tipo II: cellule dalla forma cuboidale costituiscono solo il 5% - 10% della superficie. Contengono al loro interno organelli secretori fusi con la membrana cellulare che rilasciano tensioattivo polmonare per ridurre la tensione superficiale. In mancanza di questo rivestimento, gli alveoli collasserebbero. Inoltre, gli pneumociti di tipo II sono coinvolti nella riparazione dell'epitelio alveolare danneggiato. Oltre a queste tipologie, sono inoltre presenti le cellule a spazzola, le cellule di Clara e i macrofagi come in figura 1.4.

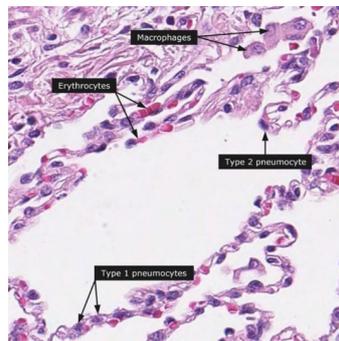


Figura 1.4: Principali cellule dell'epitelio alveolare [11]

1.2 Cancro del polmone

Il cancro al polmone all'inizio del '900 era una malattia rara ma oggi a causa dei nuovi agenti patogeni rappresenta la principale causa di morte in Europa per tumore essendo responsabile del 20% dei decessi. Nel mondo, nel 2012, si sono registrati 1.8 milioni di casi risultando in 1.6 milioni di morti. In Europa la sopravvivenza media a 5 anni è del 11.2% negli uomini e 13,9% nelle donne [12]. Seconda neoplasia in ordine di frequenza dopo il cancro alla prostata, il carcinoma polmonare manifesta i suoi sintomi dopo molto tempo. La causa maggiore di questo cancro, responsabile del 85% dei casi, è il fumo di sigaretta. Tuttavia, il 10%-15% dei casi riguarda persone che non hanno mai fumato [13]. In questi casi i fattori di rischio possono essere principalmente genetici, esposizione a sostanze cancerogene (come amianto, radon, arsenico ecc.), fumo passivo o agenti atmosferici inquinanti [14]. Inoltre, si sospetta la correlazione con altre malattie polmonari come la fibrosi polmonare, la broncopneumopatia cronica ostruttiva e il deficit di alfa-1 antitripsina. Il cancro del polmone, anche chiamato carcinoma polmonare, origina tipicamente nelle cellule epiteliali che rivestono i bronchi, i bronchioli o gli alveoli. La crescita incontrollata di queste cellule può condurre alla formazione di metastasi nei tessuti circostanti o altri parti del corpo. Viene comunemente distinto in due macro categorie:

1. **Cancro del polmone a piccole cellule (SCLC)**
2. **Cancro del polmone non a piccole cellule (NSCLC)**

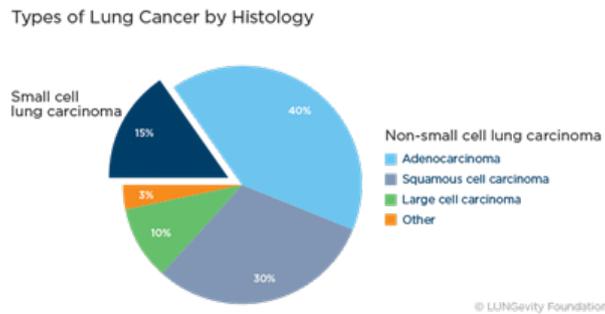


Figura 1.5: Tipologie di carcinoma polmonare [15]

I sintomi più comuni di tutte le tipologie di cancro del polmone sono [13]:

- Tosse persistente.
- Presenza di sangue nell'espettorato.
- Infezione polmonare che non si risolve o continua a ripresentarsi.
- Dolore persistente al torace o alla spalla.
- Difficoltà a respirare/affanno.
- Raucedine o abbassamento del tono della voce.
- Respiro sibilante.

1.2.1 Cancro del polmone a piccole cellule

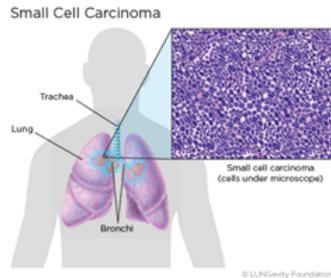


Figura 1.6: Posizione e morfologia del cancro del polmone a piccole cellule SCLC [15]

Meno comune rispetto alle altre forme, rappresenta circa il 15 % di tutte le tipologie tumorali. Si caratterizza per la ridotta dimensione delle cellule di cui è costituito e per il luogo di origine. Lo sviluppo avviene principalmente nella sottomucosa in sede ilare o centrale con coinvolgimento dei bronchi e con comparsa di piccole cellule ovali, rotonde o affusolate con margini ben definiti le cui dimensioni sono molto inferiori rispetto alle cellule sane e alle altre tipologie tumorali [15]. SCLC è molto aggressivo con crescita rapida, e si manifesta quasi sempre nei fumatori. Nella fase di diagnosi l'80% dei pazienti ha già sviluppato metastasi.

1.2.2 Cancro del polmone non a piccole cellule

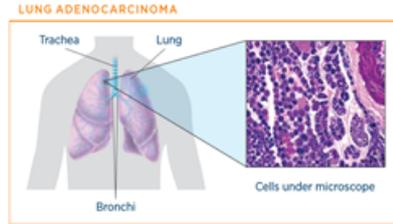
Rappresenta l'80%-90% di tutte le neoplasie polmonari maligne. NSCLC è più variabile in base al tipo istologico ed il 40% dei pazienti ha sviluppato metastasi al momento della diagnosi. I tre principali sottotipi istologici di NSCLC sono [16]:

Adenocarcinoma polmonare (Figura 1.7a): rappresenta il 40% di tutti i tumori polmonari, origina dalle cellule secernenti muco delle vie aeree. Solitamente si sviluppa nella porzione periferica del polmone. Nella maggior parte dei casi, il tumore si sviluppa a causa di infezioni croniche o in presenza di aree necrotiche. L'origine del tumore risiede probabilmente nelle cellule che compongono le ghiandole mucipare.

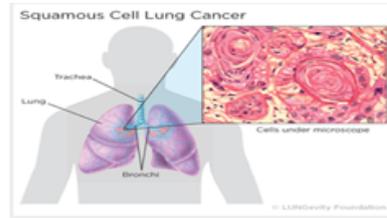
Il carcinoma a cellule squamose (Figura 1.7b): rappresenta il 25%-30% di tutti i tumori, origina dalle cellule epiteliali danneggiate a causa del fumo, solitamente nella porzione centrale del polmone. Il carcinoma a cellule squamose è classificato come tale in base all'aspetto a squama di pesce delle cellule al microscopio con la presenza di cheratinizzazione e ponti intercellulari [17].

Il carcinoma a grandi cellule (indifferenziato) (Figura 1.7c): rappresenta il 10%-15% di tutti i tumori e può comparire preferenzialmente nella zona periferica del polmone con coinvolgimento subpleurico. Si distingue dalle altre tipologie tumorali per la presenza di cellule poligonali con grossi nuclei circondati da scarso

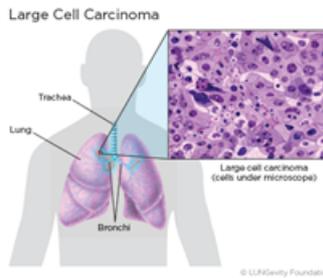
citoplasma. La diagnosi avviene normalmente per esclusione essendo scarsamente differenziato.



(a) Adenocarcinoma polmonare [18]



(b) Carcinoma polmonare a cellule squamose [15]



(c) Carcinoma a grandi cellule [15]

Figura 1.7: Posizione e morfologia del cancro NSCLC

1.2.3 Stadiazione

Il sistema di stadiazione per i tumori polmonari non a piccole cellule è il TNM [19] dove:

T: indica le dimensioni e il grado di estensione del tumore principale (da T1 a T4);

N: rappresenta il grado di coinvolgimento dei linfonodi vicini (da N0 a N3);

M: indica la presenza di metastasi (M0 a M1b).

Per la valutazione della terapia si considerano tre stadi progressivi: malattia iniziale (corrispondente agli stadi I e II), malattia localmente avanzata (stadio III) e malattia avanzata o metastatica (stadio IV). In particolare, gli stadi TNM del carcinoma polmonare sono:

- **Stadio 0:** viene anche chiamato malattia in situ dal momento che il cancro non ha invaso i tessuti circostanti.
- **Stadio 1:** le cellule tumorali non hanno coinvolto i linfonodi vicini e sono limitate ad un solo polmone. In base alle dimensioni del tumore si possono distinguere due sotto stadi:
 - IA in cui il tumore principale presenta una dimensione inferiore a 3 cm.

- IB in cui la dimensione è compresa tra 3 e 5 cm.
- **Stadio 2:** le cellule tumorali sono ancora presenti in unico polmone ma sono di dimensioni maggiori e potrebbe aver coinvolto o no altre aree circostanti o i linfonodi vicini. Non sono presenti metastasi. Si divide in due sotto stadi:
 - IIA in cui la dimensione del tumore è inferiore a 5 cm con coinvolgimento dei linfonodi vicini o in cui la dimensione è compresa tra 5 cm e 7 cm ma senza coinvolgimento dei linfonodi.
 - IIB in cui la dimensione è compresa tra 5 cm e 7 cm con coinvolgimento dei linfonodi oppure in cui il cancro non ha coinvolto i linfonodi vicini ma è presente una delle seguenti condizioni: dimensione superiore a 7 cm, tumore nei bronchi principali o tumore nel diaframma (o altri organi circostanti).
- **Stadio 3:** il cancro è ancora presente in un solo polmone, ha dimensioni superiori, potrebbe o no invadere organi circostanti o coinvolgere i linfonodi vicini, senza metastasi. Si divide in due sotto stadi:
 - IIIA in cui il tumore ha qualsiasi dimensione invade altri organi e coinvolge i linfonodi vicini oppure il cancro ha coinvolto linfonodi anche più lontani ed è presente una delle seguenti condizioni: dimensione superiore a 7 cm, tumore nei bronchi principali o tumore nel diaframma (o in altri organi).
 - IIIB in cui il tumore, di qualsiasi dimensione, ha coinvolto i linfonodi anche più lontani nel petto e ha invaso altri organi oppure in cui il tumore ha coinvolto linfonodi controlaterali (sia all'interno che all'esterno del polmone) ed è presente una delle seguenti condizioni: dimensione superiore a 7 cm, tumore nei bronchi principali, tumore nel diaframma (o altri organi circostanti) o tumore nel cuore.
- **Stadio 4:** i tumori sono di qualsiasi dimensione e hanno coinvolto l'altro polmone con metastasi.

	T	N	M
Carcinoma occulto	TX	N0	M0
Stadio 0	Tis (carcinoma in situ)	N0	M0
Stadio IA	T1a,b	N0	M0
Stadio IB	T2a	N0	M0
	T2b	N0	M0
Stadio IIA	T1a,b	N1	M0
	T2a	N1	M0
	T2b	N1	M0
Stadio IIB	T3	N0	M0
	T1a,b, T2a,b	N2	M0
Stadio IIIA	T3	N1, N2	M0
	T4	N0, N1	M0
	T4	N2	M0
Stadio IIIB	Qualunque T	N3	M0
Stadio IV	Qualunque T	Qualunque N	M1

■ Malattia iniziale
■ Malattia localmente avanzata*
■ Malattia metastatica

* Alcune classificazioni inquadrano lo stadio II come malattia localmente avanzata

Figura 1.8: Stadiazione del carcinoma polmonare [19]

1.3 Diagnosi e analisi istologica del tumore al polmone

La diagnosi del carcinoma polmonare può avvenire mediante numerose modalità al fine di pianificare adeguatamente il trattamento. Gli step da perseguire per la diagnosi includono: anamnesi, test di imaging, test di laboratorio, biopsia e test per verificare la presenza di biomarcatori [20]. I test di imaging sono necessari per rilevare la localizzazione e la diffusione della massa tumorale ma solo con la biopsia è possibile ottenere una diagnosi definitiva.

1.3.1 Anamnesi

L'anamnesi, o storia clinica, è la raccolta dei sintomi e delle informazioni utili di carattere medico per ottenere una diagnosi provvisoria, da approfondire con altri esami. Il medico prenderà in considerazione una serie di fattori, quali [20]:

- Storia personale del paziente: come esposizione a fattori di rischio, fumo o fumo passivo, occupazioni che includono l'utilizzo di metalli pesanti o lavoro in miniera ecc.
- Storia familiare di carcinoma polmonare o altri tumori epiteliali.
- Storia medica del paziente: malattie passate o correnti, assunzione di farmaci, integratori e vitamine.
- Studio dei sintomi come tosse, mancanza di respiro, perdita di peso, raucedine ecc. ed esame fisico per rilevare ad esempio la presenza di linfonodi ingrossati, suoni anomali del respiro, gonfiore delle mani, piedi, viso o caviglie ecc.

1.3.2 Test di imaging

I test di imaging forniscono informazioni utili per ricreare l'intero quadro clinico del paziente. Le tecniche di imaging vengono utilizzate per individuare eventuali

aree sospette, diffusione del cancro, per monitorare il trattamento polmonare e il post trattamento [20]. Le principali tecniche di imaging includono la radiografia al torace, la tomografia computerizzata (TC), la risonanza magnetica (MRI), la tomografia ad emissioni di positroni (PET) e la scansione ossea.

Radiografia al torace: solitamente è il primo test che viene condotto quando si ha un sospetto di carcinoma polmonare. Tuttavia, la radiografia al torace non è in grado di rilevare il 20% o più dei tumori polmonari soprattutto se sono di piccole dimensioni come nella fase iniziale. Considerando che il tempo medio di raddoppio del carcinoma polmonare può essere di 136 giorni, il ritardo nella diagnosi, causato da falsi negativi, potrebbe significare perdere la possibilità di operare chirurgicamente nelle fasi iniziali. In questa fase, infatti, i tumori polmonari sono facilmente curabili e localizzati. Nelle immagini a raggi X, la polmonite può oscurare ulteriormente il cancro. Molto spesso, inoltre, il tumore viene erroneamente diagnosticato come tubercolosi essendo molto simili ai raggi X.

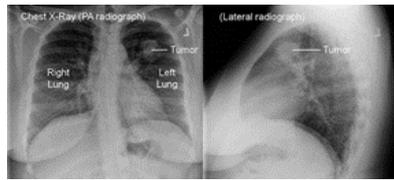


Figura 1.9: Diagnosi mediante radiografia [20]

Tomografia computerizzata: a differenza di una radiografia tradizionale, che scatta una singola foto, uno scanner TC scatta più foto ruotando intorno al paziente in modo da ricostruire un volume. Una TAC è in grado di ricostruzione le dimensioni, la forma e la posizione di masse o noduli polmonari. Inoltre, è in grado di rilevare linfonodi o masse sospette in altri organi. Tuttavia, per confermare la presenza o meno di un cancro polmonare si necessita di un esame istologico prelevato da un campione polmonare. Oggi viene spesso utilizzata una LDCT low dose-CT scan (dose pari ad un quinto dello scanner normale) per lo screening polmonare e i follow up per monitorare l'efficacia del trattamento.

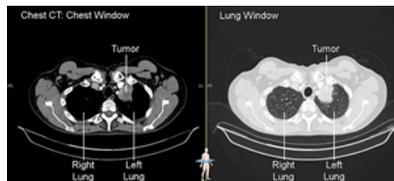


Figura 1.10: Diagnosi mediante TAC [20]

Risonanza magnetica (MRI): è considerata superiore alla TC ma viene più spesso utilizzata quando il cancro è stato già diagnosticato. Viene utilizzato soprattutto per verificare che il cancro non si sia diffuso in altri organi come il cervello o il midollo spinale. La MRI, tuttavia, è scarsamente utilizzata per le parti del corpo in movimento come i polmoni.

Tomografia ad emissione di positroni: questa tecnica aiuta a rilevare dei noduli polmonari a partire da immagini TC e radiografie sospette, ma viene comunemente utilizzata per fare la stadiazione di un tumore. In particolare, l'esame PET si rivela molto utile per rintracciare il coinvolgimento dei linfonodi in sede o in altre regioni. Viene iniettata per via endovenosa una piccola quantità di zucchero radioattivo. Questo componente viene maggiormente assorbito dalle cellule tumorali piuttosto che da quelle sane. Normalmente la PET è combinata con uno scanner TC per rilevare aree sospette con una qualità dell'immagine superiore.

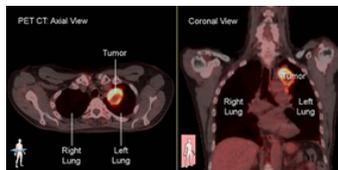


Figura 1.11: Diagnosi mediante PET [20]

Scansione ossea: utilizza un tracciante radioattivo, iniettato per via endovenosa, per osservare l'interno delle ossa per mezzo di una telecamera. L'osso sano appare più chiaro e le aree di lesione, come quelle causate dal cancro, spiccano sull'immagine. La PET ha, tuttavia, sostituito le scansioni ossee per trovare NSCLC che si è diffuso nelle ossa.

1.3.3 Biopsia

La biopsia è un esame medico che consiste nel prelievo di una porzione di un tessuto per l'analisi al microscopio o con tecniche di microbiologia o biologia molecolare [20]. L'obiettivo è di arrivare ad una diagnosi definitiva circa il sospetto di un carcinoma polmonare e di capirne la tipologia per l'iter terapeutico. Attualmente, le biopsie sono l'unico metodo per confermare la diagnosi di cancro ai polmoni. Le procedure di biopsia includono:

L'agobiopsia polmonare: un radiologo interventista inserisce un ago attraverso la parete toracica tra due costole con l'ausilio di una TC del torace o un'ecografia dopo aver eseguito l'anestesia locale. Viene quindi aspirato un campione della massa e inviato al laboratorio per le analisi istologica al microscopio. I vantaggi dell'agobiopsia sono l'anestesia locale e la rapidità della procedura senza il ricovero ospedaliero. Gli svantaggi principali sono dovuti alla possibilità di perdere qualche nodulo di piccole dimensioni o di non avere a disposizione un campione sufficiente per la diagnosi. Inoltre, è possibile che si verifichi lo pneumotorace.



Figura 1.12: Agobiopsia polmonare [20]

Biopsia polmonare broncoscopica: il chirurgo o lo pneumologo inserisce un tubo flessibile e sottile con una luce, nella bocca o nel naso fino a raggiungere i passaggi respiratori polmonari e, mediante degli strumenti fatti passare nel broncoscopio vengono prelevati dei campioni. Prima dell'intervento viene spruzzato un medicinale paralizzante in bocca. Non prevede alcun ricovero ospedaliero, ma solo un breve periodo di osservazione della durata di 1-2 ore.

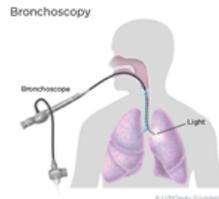


Figura 1.13: Biopsia broncoscopica [20]

Biopsia toracoscopica: si tratta di un'operazione chirurgica eseguita in anestesia totale. Il chirurgo esegue un piccolo taglio nella pelle della parete toracica e inserisce uno strumento dotato di una videocamera a fibra ottica collegata ad un monitor per esaminare i polmoni. Essendo invasiva non è la prima scelta tra le procedure eseguite per la diagnosi, mentre viene spesso eseguita nella fase di trattamento della patologia.

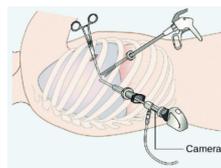


Figura 1.14: Biopsia toracoscopica [21]

Biopsia a cielo aperto o toracotomia : viene effettuata solo se altri metodi non hanno funzionato per ottenere campioni polmonari. Si tratta di un intervento chirurgico vero e proprio, e come tale richiede l'anestesia totale. Il chirurgo esegue un taglio dal petto e sotto le braccia alla schiena per prelevare una porzione cellulare maggiore e poi effettua una sutura con punti riassorbibili. Al termine, il chirurgo deve praticare un drenaggio pleurico per l'espansione del polmone collassato. È richiesto il ricovero di qualche giorno per il monitoraggio dei parametri vitali.



Figura 1.15: Toracotomia [22]

Ecoendoscopia con agoaspirazione : dopo aver effettuato l'anestesia locale, viene inserito il broncoscopio flessivo dotato di un dispositivo ad ultrasuoni lungo la trachea. All'interno del broncoscopio viene poi fatto passare un ago attraverso il bronco e il linfonodo mediastinico per prelevare un campione cellulare.



Figura 1.16: Ecoendoscopia [23]

Toracentesi o toracocentesi [24]: la toracentesi consiste nel prelievo di liquido attraverso l'inserimento di un ago nel margine superiore della costa, dopo un'anestesia locale. Il fluido viene in seguito inviato in laboratorio per verificare la presenza di cellule tumorali.

Mediastinoscopia e mediastinotomia : questo esame permette la visualizzazione dell'area toracica compresa tra i due polmoni (mediastino) attraverso una sonda a fibre ottiche [25]. Il chirurgo esegue una piccola incisione in anestesia totale appena sopra lo sterno per raggiungere il mediastino mediante un sottile tubo cavo con una luce. Attraverso l'apertura vengono poi prelevati dei campioni dai linfonodi lungo la trachea. La mediastinotomia ottiene tessuto anche dall'area dei polmoni con un'incisione maggiore.

1.3.4 Analisi istologica

L'istologia è una branca della biologia che studia la composizione dei tessuti vegetali e animali. L'istologia riveste un ruolo fondamentale nell'ambito dell'anatomia patologica attraverso l'esame di campioni cellulari prelevati mediante biopsia e fissati su un vetrino al microscopio [26]. Prima dell'analisi al microscopio ottico è

necessario ottenere dei campioni dallo spessore di pochi μm che presentano le stesse caratteristiche fisiologiche ma con funzioni biologiche arrestate. La preparazione del campione istologico è costituita da una serie di step [27]:

1. Prelievo di frammenti dell'organo
2. Fissazione
3. Inclusione
4. Colorazione

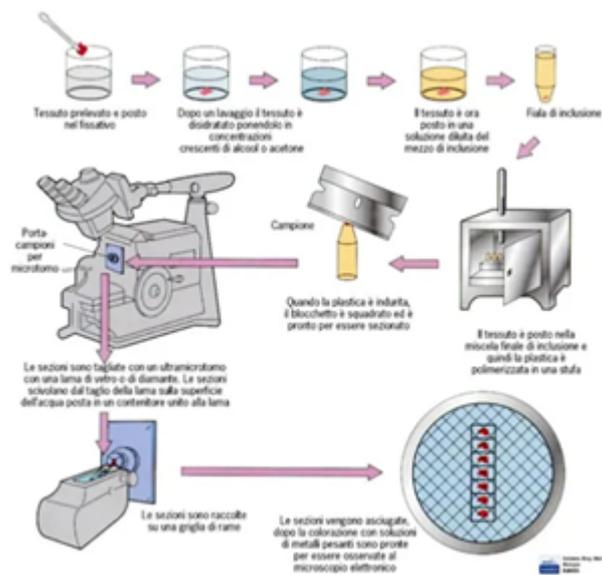


Figura 1.17: Schema di preparazione del campione istologico [28]

Il prelievo va condotto rapidamente sul materiale fresco ottenuto mediante biopsia polmonare. La fissazione consiste nel trattamento del frammento mediante procedimenti chimici al fine di stabilizzare i costituenti del tessuto e inattivare gli enzimi autolitici. Questa fase dura da pochi minuti a massimo due giorni in base alle dimensioni del frammento. I fissativi più utilizzati sono la formaldeide (o formalina) al 10%, che consente di realizzare i più comuni metodi di colorazione e l'alcool etilico a 90°. Segue l'inclusione che consente il taglio in sezioni sottili di pochi μm in modo da rendere possibile il passaggio della luce attraverso il campione. L'inclusione viene realizzata lasciando permeare nel tessuto una sostanza idrofoba, la paraffina, in grado di solidificare a temperatura ambiente. Prima della penetrazione è necessario allontanare l'acqua dal tessuto per mezzo di un disidratante. Per tale operazione viene normalmente impiegato l'alcool etilico a gradazione crescente che verrà poi rimosso tramite l'utilizzo di sostanze solubili in paraffina. Una volta

che la penetrazione è avvenuta, il pezzo viene raffreddato in modo da ottenere la consistenza della paraffina solida. A questo punto il tessuto viene tagliato con un microtomo a lama d'acciaio per ottenere fette sottilissime che verranno montate su un vetrino porta oggetto. Per la fase della colorazione, viene allontanata la paraffina e riportato il tessuto al suo stato di idratazione. La colorazione mette in evidenza i singoli componenti strutturali. I coloranti impiegati sono in grado di legarsi alle strutture aumentando il contrasto dell'immagine. Al termine del processo di colorazione, il campione viene sigillato tra due vetrini e viene analizzato al microscopio.

Colorazioni delle immagini istologiche

In ambito istologico esistono attualmente numerose tecniche di colorazione, le principali sono [29]:

- Ematossilina Eosina (E-E)
- PAS (acido periodico-reattivo di Schiff)
- Ematossilina ferrica
- Tricotomia di Mallory
- Tricotomia di Heidenhain
- Tricotomia di Masson
- Ighosti

Le immagini a disposizione per la tesi sono caratterizzate dalla colorazione con Ematossilina Eosina (in inglese H&E). Si tratta di una colorazione bicromica con un colorante basico (ematossilina) e uno acido (eosina). L'Ematossilina o emallume di Mayer è un colorante vegetale in grado di colorare di blu-viola i componenti cellulari carichi negativamente come i nuclei. L'Eosina colora invece di rosa i componenti carichi positivamente come il citoplasma e il tessuto connettivo.

Istologia digitale

Con il termine istologia digitale ci si riferisce al processo completo di acquisizione, gestione, condivisione e interpretazione dei dati. In un primo momento, il termine si riferiva esclusivamente al processo di digitalizzazione con cui i campioni posti sul vetrino venivano convertiti in slide digitali attraverso l'utilizzo di uno scanner ad alta risoluzione. Oggi, l'accezione del termine è più ampia e include anche gli approcci di machine learning per la rilevazione, la segmentazione e l'analisi delle immagini [30]. La patologia digitale è stata approvata dalla FDA per la diagnosi primaria nel 2017.

Le slide digitali, dette WSI (Whole Slide Images), indicano delle rappresentazioni di un intero campione sul vetrino alla risoluzione del microscopio. Le slide digitali possono essere visualizzate e analizzate sul monitor di un computer ed essere condivise con più esperti anche su internet per ottenere una diagnosi più accurata. L'analisi dei campioni al microscopio è molto più complessa e fortemente soggettiva. Grazie ai software di analisi automatica è possibile facilitare l'interpretazione delle immagini digitali mediante l'aggiunta di annotazioni, note o la misurazione



Figura 1.18: Scanner Leica Biosystems e monitor per la visualizzazione delle WSI [31]

di precise strutture all'interno dell'immagine. L'introduzione delle slide digitali o WSI consente una migliore reperibilità dei dati e la possibilità di confronto tra le varie porzioni del tessuto o con casi precedenti [32]. Negli ultimi anni lo sviluppo di sistemi di image processing in ambito medico e l'introduzione del machine learning ha condotto molti ricercatori e studiosi a creare software e metodi sempre più precisi per la diagnosi.

Artefatti delle immagini istologiche

La preparazione del campione istologico è un processo molto complesso che richiede molta attenzione ed esperienza in laboratorio. Molto spesso i preparati istologici risentono della presenza di alterazioni che possono condurre ad interpretazioni errate o inconcludenti. Tutti i cambiamenti indesiderati introdotti sui campioni di tessuto e/o sulle slide digitali durante qualsiasi fase del processo istologico si definiscono 'artefatti'. Gli artefatti costituiscono la principale causa di problemi diagnostici e per tale ragione è necessario riconoscerli per evitare errori. Tra gli innumerevoli artefatti, i più ricorrenti ed importanti sono [33]:

- **Contaminazioni:** può capitare che il bagno d'acqua venga contaminato da capelli, polvere o cellule residue di altre sezioni ecc. alterando l'immagine esaminata al microscopio (Figura 1.19a).
- **Tissue folds/ Pieghe del tessuto:** questa tipologia di artefatto si manifesta durante la fase di taglio a cause del ripiegamento del tessuto su sé stesso per posizionamento impreciso. Le regioni ripiegate su sé stesse presentano una colorazione più scura rispetto al resto del preparato istologico (Figura 1.19b).
- **Bolle d'aria:** le bolle d'aria, per disidratazione incompleta, possono rimanere intrappolate al di sotto del tessuto alterandone la colorazione una volta asciutte (Figura 1.19c).
- **Strappi:** si presentano come delle interruzioni striate nelle sezioni (a volte vere e proprie rotture) durante la fase di taglio, dovute ad eccessiva durezza del campione o vibrazioni della lama o angolo di gioco elevato (Figura 1.19d).

- **Sfocatura dell'immagine:** l'artefatto di sfocatura (o blur artifact) a differenza dei precedenti avviene nella fase di acquisizione e/o digitalizzazione dei campioni. La causa principale risiede in un allineamento impreciso delle porzioni con il piano focale del microscopio (Figura 1.19e).

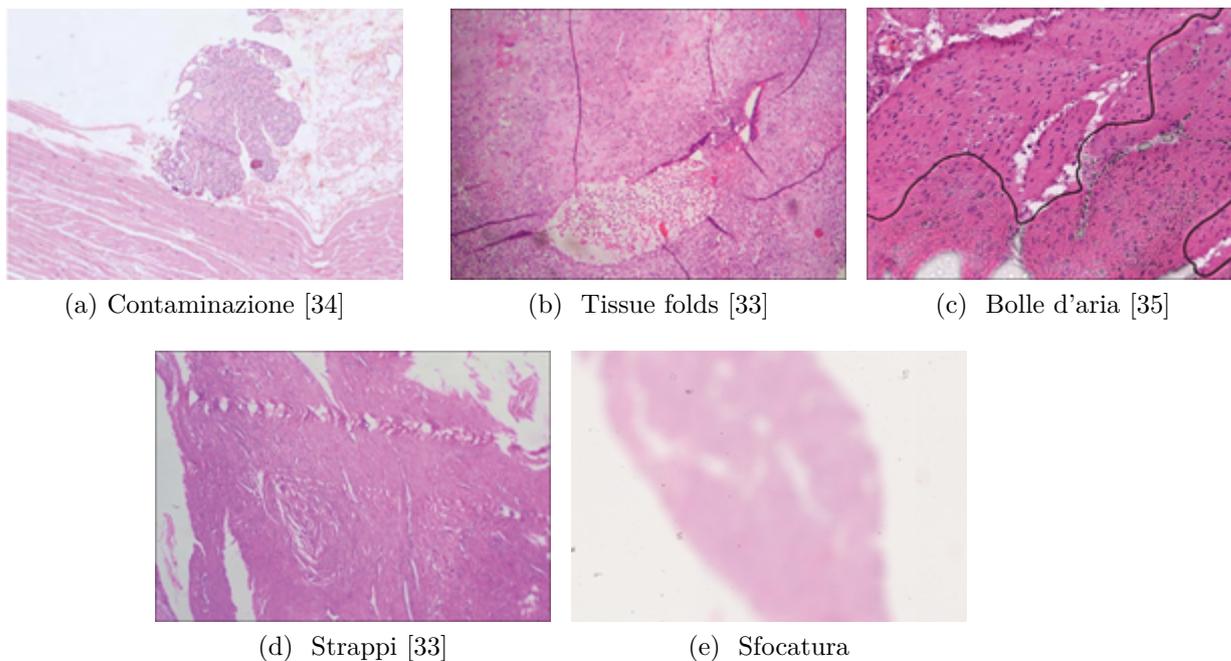


Figura 1.19: Esempi di artefatti istologici

Capitolo 2

Materiali e metodi

In questo capitolo vengono mostrati i metodi e i materiali impiegati per la realizzazione dell’algoritmo proposto nel progetto di tesi. In primo luogo, viene presentato il dataset per lo sviluppo e la validazione dell’algoritmo in termini di fonte, numerosità e caratteristiche principali. Segue la presentazione completa dell’architettura del metodo proposto per la classificazione di immagini istologiche nelle sue diverse fasi:

- Preprocessing
- Allenamento
- Postprocessing

Nello specifico verranno presentate le varie tipologie di reti neurali utilizzate nella fase di allenamento e le diverse metriche impiegate per ottenere i risultati della validazione.

2.1 Dataset

Per lo svolgimento della tesi è stato utilizzato il dataset pubblico della challenge ACDC@ LUNGHP (Automatic Cancer Detection and Classification in Whole-slide Lung histopathology) che si è tenuta in The IEEE International Symposium on Biomedical Imaging (ISBI) a Venezia ad Aprile 2019 [36]. Il dataset completo consta di 200 WSI ottenute da 200 pazienti con cancro al polmone (sito C34.1-C34.9 e tipo istologico: adenocarcinoma, carcinoma a cellule squamose e carcinoma a piccole cellule in rapporto 6:3:1) reclutati dal Department of Pulmonary Oncology in the First Hospital of Changsha, a partire da Gennaio 2016 e fino a Novembre 2017. Tutti i pazienti reclutati presentano un’età compresa tra i 30 e i 90 anni, hanno effettuato biopsia chirurgica e nessuna radioterapia precedente all’intervento. Sono stati esclusi, invece, candidati con tumori multipli primari, con metastasi, pazienti con immunodeficienza o storia di trapianti e/o pazienti che non hanno

aderito al consenso informato [37]. I campioni sono stati sottoposti alla colorazione con ematossilina ed eosina (H&E) e sono stati digitalizzati utilizzando lo scanner digitale 3DHISTECH Panoramic 250 con risoluzione 20x ($0.242 \mu\text{m}/\text{pixel}$). Delle 200 WSI, 150 con le relative annotazioni sono state utilizzate come training mentre le altre 50 (sprovviste delle annotazioni) come test. Tutti i 200 campioni sono stati accuratamente annotati dal direttore del dipartimento di patologia mentre per i 50 campioni del test è stata chiesta un'ulteriore annotazione ad un secondo anatomopatologo con 20 anni di esperienza. Le 150 WSI del training sono state ulteriormente suddivise in:

- Training 1: 100 WSI
- Training 2: 50 WSI

Le dimensioni delle WSI ad alta risoluzione (livello zero) sono di circa 200000×80000 pixel. Le immagini sono state salvate e rilasciate in formato TIFF e le annotazioni manuali in formato XML. All'interno della stessa WSI sono presenti più campioni della stessa biopsia con diversa orientazione. Per la visualizzazione delle WSI e delle relative annotazioni è stato utilizzato il software ASAP (Automated Slide Analysis Platform), Figura 2.1, che offre la possibilità di creare e visualizzare annotazioni di vario tipo (punti, rettangoli, poligoni o annotazioni libere).

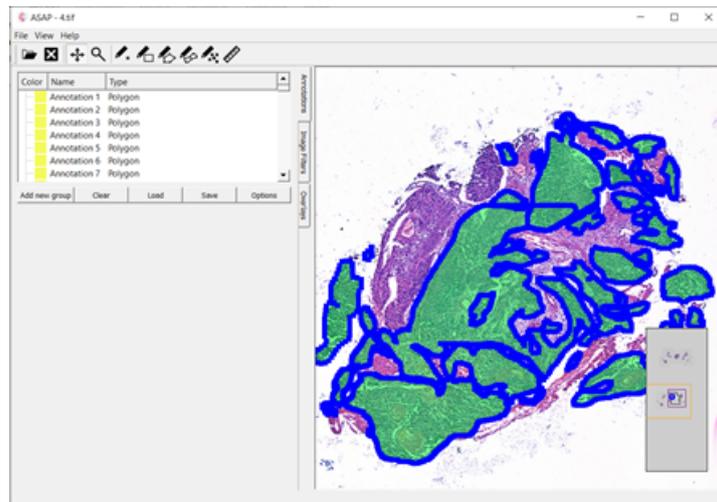


Figura 2.1: Software ASAP: WSI con annotazioni manuali del patologo

2.2 Architettura dell'algoritmo

L'architettura totale dell'algoritmo proposto è schematizzata in figura 2.2 Il primo step dell'algoritmo consiste nel caricamento delle WSIs costituenti il dataset ACDC@ LUNGHP in colorazione H&E. Successivamente si procede con la fase di

preprocessing con l'obiettivo di estrarre le patch di dimensione 500x500 da utilizzare come input per il training delle reti neurali. Per ciascuna WSI vengono estratte manualmente 10 Tiles di dimensione 3000x3000 che costituiranno il punto di partenza dell'algoritmo di patch extraction. Le patch ottenute in questa fase rappresentano, quindi, il dataset di allenamento delle CNN. Le fasi di ottimizzazione e di estrazione smart delle patch, nello step di preprocessing, vengono realizzate sfruttando i metodi sviluppati in lavori precedenti [38] [39]. Il corpo centrale dell'algoritmo si compone dell'allenamento delle tre diverse architetture neurali al fine di ottenere tre modelli in grado di apprendere pattern diversi del dataset e di assemblarne le predizioni per migliorare l'accuratezza della classificazione. Lo step di postprocessing consiste invece nella creazione finale delle Heatmaps a partire dalle WSI del test. Per le Heatmaps viene fatta progressivamente scorrere una sliding window 500x500 di cui viene ricavata la predizione totale fondendo i risultati dei tre modelli allenati.

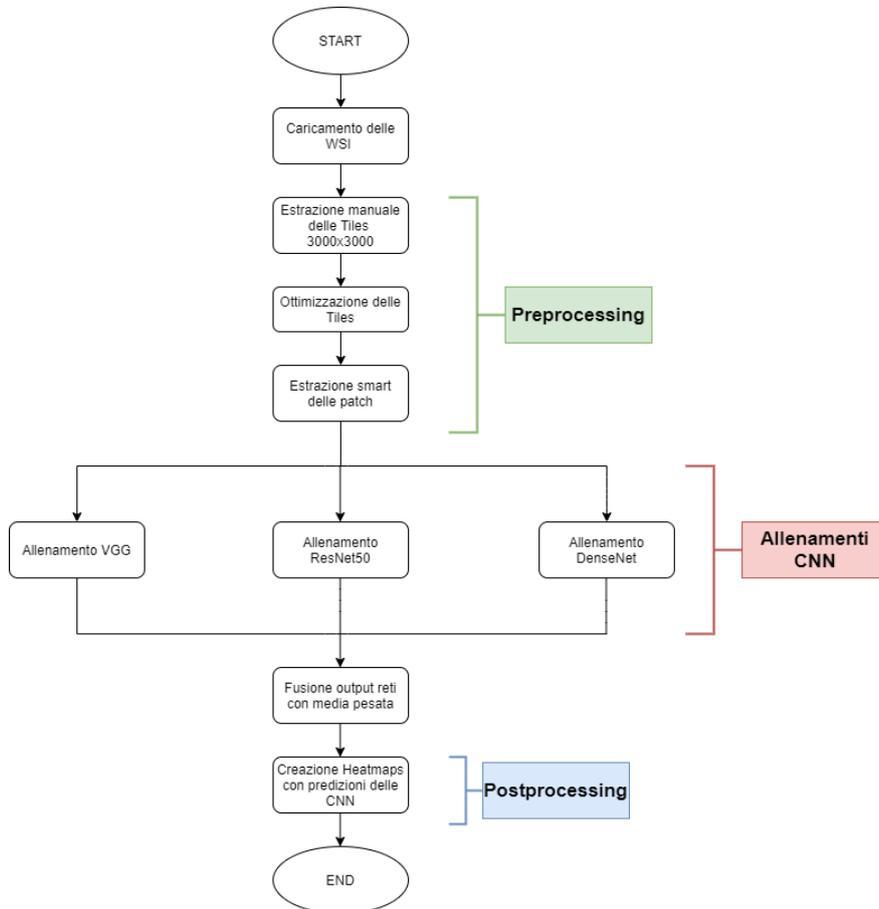


Figura 2.2: Schema totale dell'algoritmo proposto

2.3 Preprocessing

Lo step di preprocessing dell'algoritmo si articola in due fasi principali come descritto nei paragrafi seguenti 2.3.1 e 2.3.2. Nella prima parte vengono ottenute manualmente 10 Tiles per ogni WSI con il software ASAP mediante la creazione di bounding box di dimensione 3000x3000. Segue l'algoritmo di ottimizzazione e standardizzazione delle Tiles per migliorarne e uniformarne l'aspetto. Sulle Tiles ottimizzate viene quindi eseguito l'algoritmo automatico di estrazione delle patch 500x500 che costituiranno il dataset per il training delle reti.

2.3.1 Estrazione manuale delle Tiles

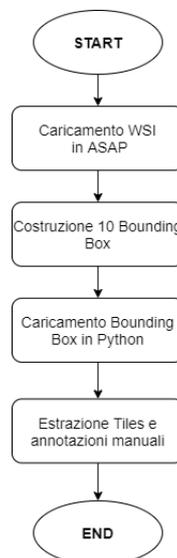


Figura 2.3: Pipeline per l'estrazione manuale delle Tiles

Il primo passaggio della fase di preprocessing consiste nell'estrazione manuale delle tiles e delle relative annotazioni dalle 150 WSI costituenti il training. Le tiles sono state estratte con il software ASAP mediante la creazione di bounding box di dimensione 3000x3000 come in Figura 2.4. Le quattro coordinate dei vertici dei bounding box realizzati sono state salvate in formato XML e caricate in Python per l'estrazione della regione corrispondente della WSI (a livello 0) e dell'annotazione manuale del patologo. Per ogni WSI sono state estratte 10 tiles non consecutive, avendo cura di non selezionare zone affette dagli artefatti descritti nel paragrafo 1.3.4 (artefatti delle immagini istologiche). In totale si sono ottenute 1500 Tiles di cui 30 (provenienti dalle WSI 8-15 del Training 1 e dalla WSI 38 del Training 2) sono state eliminate, dopo un'attenta analisi, per scarsa risoluzione. L'estrazione delle

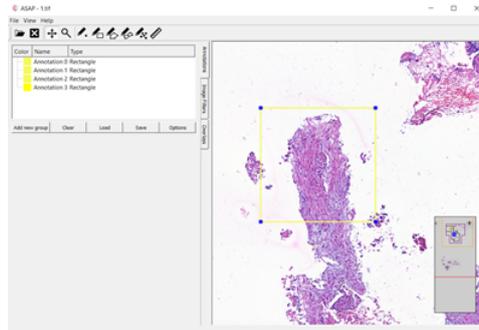


Figura 2.4: Creazione Bounding box con software ASAP

Tiles è stata eseguita in modalità manuale in modo da individuare le parti della WSI con maggiore risoluzione e minore presenza di artefatti con lo scopo di costruire un classificatore accurato. Inoltre, con l'obiettivo di realizzare un modello più robusto possibile, sono state appositamente scelte Tiles comprendenti sia tessuto tumorale che sano. Le Tiles come in Figura 2.5 presentano le seguenti caratteristiche:

- Numero: 1470
- Dimensioni: 3000x3000
- Risoluzione: livello 0 (20x)
- Formato: .png

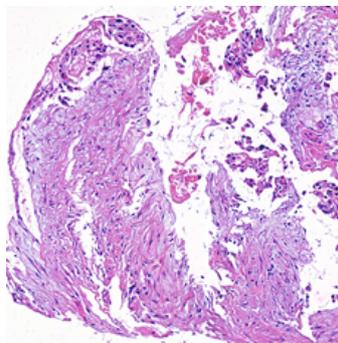


Figura 2.5: Tile estratta dal dataset

2.3.2 Estrazione smart delle patch

L'estrazione delle patch è un procedimento fondamentale per l'allenamento delle reti neurali. Le patch, con le loro label, rappresentano l'input delle CNN, motivo per cui i risultati del training sono largamente influenzati dall'informazione proveniente dalle immagini. L'utilizzo di patch estratte a griglia a partire dalla WSI

genererebbe un gran numero di immagini non significative nel dataset per l'allenamento. Basti pensare all'eccessiva presenza di bianco su una WSI in relazione al tessuto di interesse. Si rende quindi necessario un algoritmo in grado di discernere tra parti più o meno rilevanti all'interno della WSI al fine di garantire la creazione di un modello capace di predire correttamente la classe.

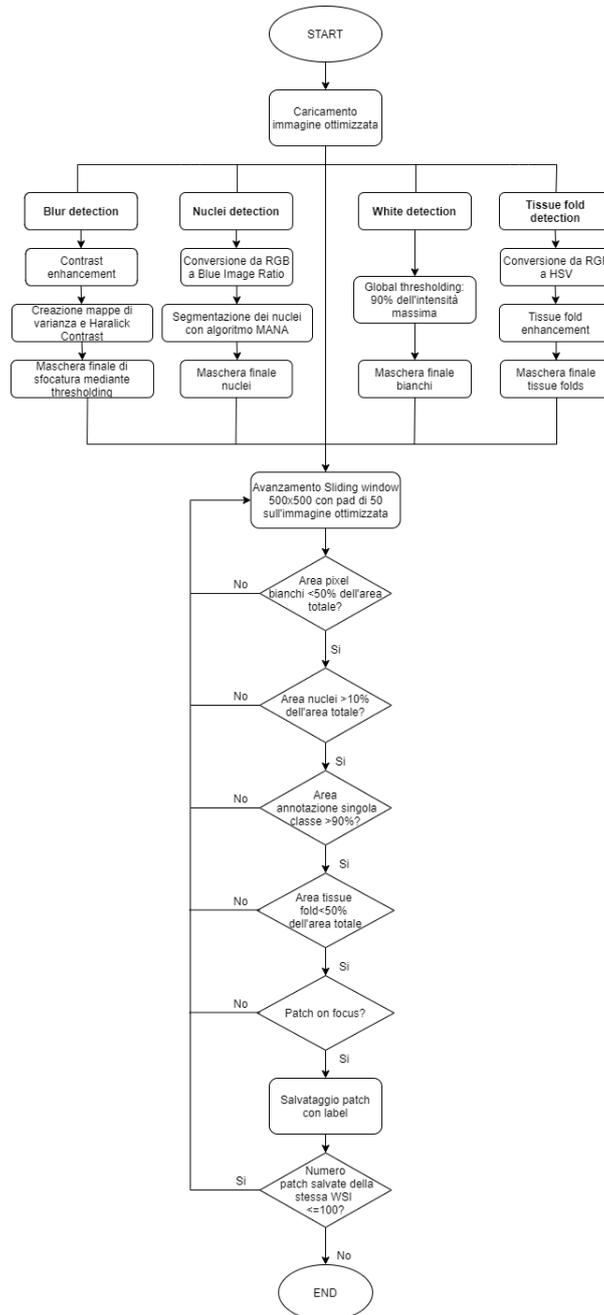
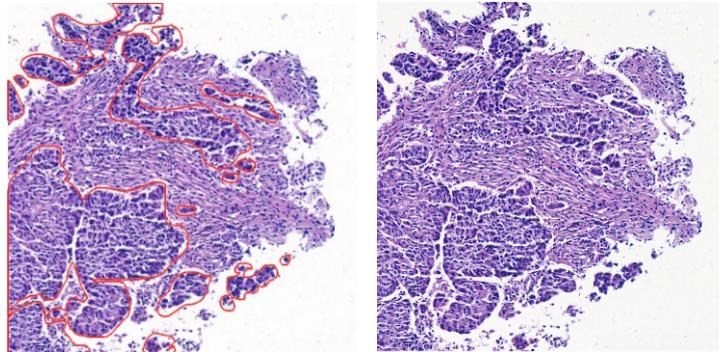


Figura 2.6: Pipeline di estrazione smart delle patch

Prima di procedere con l'estrazione smart delle patch, la Tile viene ottimizzata utilizzando algoritmi precedentemente implementati [38] (figura 2.7).



(a) Immagine originale con annotazione in rosso (b) Immagine ottimizzata

Figura 2.7: Caricamento immagine e ottimizzazione

Segue la creazione delle maschere dei nuclei, dei bianchi, dei tissue folds e delle zone sfuocate. Le maschere ottenute verranno utilizzate come criterio di estrazione delle patch.

Nuclei detection

Il primo step consiste nella conversione dell'immagine ottimizzata da RGB in Blue Ratio Image con la seguente formula [40]:

$$BlueRatio = \frac{100 \cdot B}{1 + R + G} \cdot \frac{256}{1 + R + G + B} \quad (2.1)$$

dove R rappresenta l'intensità del canale rosso dell'immagine, B l'intensità del canale blu e G l'intensità del canale verde.

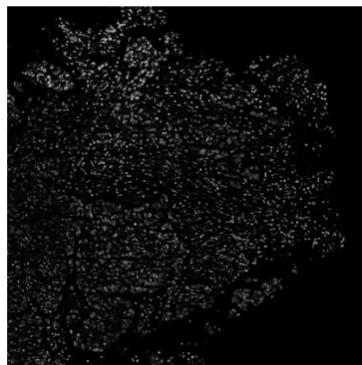


Figura 2.8: Blue Ratio image

L'obiettivo della conversione è quello di evidenziare gli oggetti ad elevata luminosità nell'immagine nel canale blu. In effetti, il canale blu dell'immagine è in grado di accentuare la mitosi e quindi i nuclei delle cellule in un'immagine istologica.

A seguito della conversione, sulla Blue Ratio Image (figura 2.8) viene eseguito l'algoritmo MANA di segmentazione dei nuclei [39]. In primo luogo, sull'immagine così ottenuta viene calcolato l'istogramma e a partire da questo, la *progressive weighted mean curve* così definita:

$$PWM_{CURVE} = \frac{\sum_{i=1}^P w_i x_i}{\sum_{i=1}^P x_i} \quad (2.2)$$

dove P indica la classe generica dell'istogramma, w_i l'occorrenza per i-esima classe e x_i la localizzazione del bin dell'istogramma. La PWM_{CURVE} viene quindi calcolata per tutte le classi dell'istogramma al fine di individuarne i punti di flesso come possibili valori di soglia per la segmentazione dei nuclei. L'immagine viene quindi segmentata utilizzando tutti i punti di flesso e per ogni segmentazione viene calcolata l'area media degli oggetti trovati. La soglia iniziale è quella che porta alla segmentazione di oggetti con area media maggiore. In seguito, vengono eliminate le aree troppo piccole per essere considerate nuclei (area minore del 25% dell'area media), e vengono separate le aree troppo grandi utilizzando progressivamente valori di soglia più bassi. L'ultimo step dell'algoritmo di MANA consiste in un'ulteriore separazione dei nuclei sulla base della *marker-based watershed* in cui i seed dei nuclei vengono scelti sulla base dei massimi locali della *distance transform*. A questo punto, viene ottenuta la maschera di segmentazione dei nuclei da utilizzare per la selezione delle patch (figura 2.9).

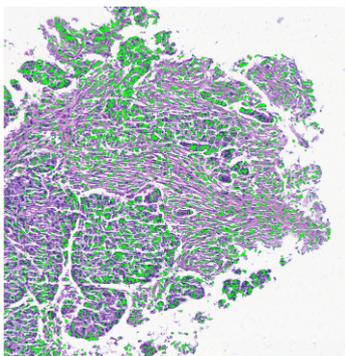


Figura 2.9: Nuclei detection

White detection

Prima dell'estrazione smart vera e propria, viene infine effettuata una *white detection* sulla Tile con un *global thresholding* pari al 90% del massimo di intensità dell'immagine. La white detection ha come scopo l'esclusione delle parti bianche

del vetrino nella selezione delle patch per l'allenamento della rete. La white detection produce in uscita una maschera dei pixel bianchi dell'immagine (figura 2.10) da utilizzare come criterio di esclusione per l'estrazione delle patch.

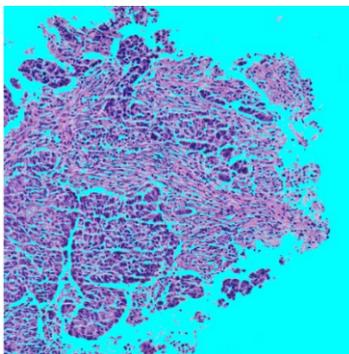


Figura 2.10: White detection

Tissue fold detection

L'algoritmo di tissue fold detection ha come obiettivo l'individuazione degli artefatti di piegatura a partire dalle Tiles. In uscita si ottiene quindi la maschera dei tissue folds utilizzata come criterio di esclusione delle patch. Una WSI con colorazione H&E contiene tre regioni principali: nuclei, citoplasma e artefatti di tissue folds. Le tre regioni differiscono principalmente per il canale di saturazione (S) e di value (V) della codifica colore HSV [41].

- **Nuclei:** regioni blu-viola con bassa intensità o value.
- **Citoplasma:** regioni rosa con alta intensità.
- **Tissue folds:** regioni con elevata saturazione e bassa intensità.

Di conseguenza, essendo le diverse strutture più evidenti nella codifica HSV, il primo step dell'algoritmo consiste nella conversione dell'immagine da RGB in HSV come mostrato in figura 2.11 . A questo punto, viene valutato, attraverso l'istogramma delle luminosità, il numero di pixel del canale di saturazione con intensità superiore a 0.8 (double) o 204 (uint8) secondo la seguente formula:

$$Counts\ max_{saturation} = \frac{\frac{\sum_{i=m}^n w_i}{n - m}}{\frac{\sum_{i=1}^n w_i}{n}} \quad (2.3)$$

dove w_i indica il numero di pixel per i -esima classe dell'istogramma, n la massima intensità pari a 256 e m pari a 204. Nel caso in cui $Counts\ max_{saturation}$ sia maggiore a 0.6, il canale di saturazione dell'immagine viene scalato tra 0 e 0.5. L'immagine viene quindi riconvertita prima in RGB e poi nuovamente in HSV. Si procede con la

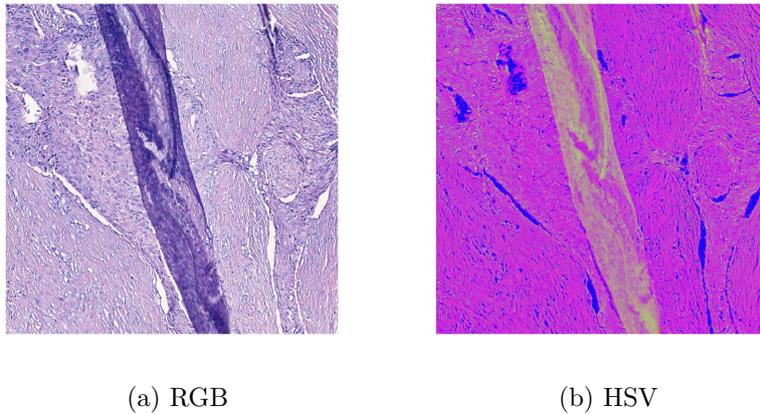


Figura 2.11: Conversione immagine da RGB a HSV

definizione della mappa di tissue folds enhancement D ottenuta nel modo seguente:

$$D(x, y) = H(x, y) + 2 \cdot (S(x, y) - V(x, y)) \quad (2.4)$$

dove $H(x,y)$, $S(x,y)$ e $V(x,y)$ costituiscono i tre canali dell'immagine nella codifica HSV. All'immagine così ottenuta viene effettuata un'operazione di erosione mediante elemento morfologico strutturale circolare di raggio 5 (figura 2.12). L'immagine mette in evidenza gli elementi che presentano quindi un'elevata saturazione e bassa intensità come i tissue fold e in alcuni casi anche i nuclei.

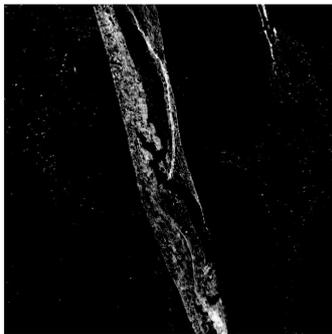


Figura 2.12: Tissue fold enhancement

L'immagine viene quindi segmentata mediante global thresholding con varie soglie in un range compreso tra $t \in \{-0.5, +1\}$ al fine di individuare il valore di soglia ottimale. Per valori elevati di soglia, vengono messi in evidenza unicamente pochi elementi connessi (principalmente tissue folds); per soglie intermedie vengono

invece segmentati più elementi connessi (nuclei e tissue folds); per soglie basse vengono invece presi grandi elementi connessi che comprendono anche il citoplasma cellulare. La soglia ottimale per la segmentazione viene quindi ricavata in base al numero di oggetti connessi segmentati per quella determinata soglia. Viene quindi rappresentata la curva degli elementi connessi per ciascuna soglia come in figura 2.13.

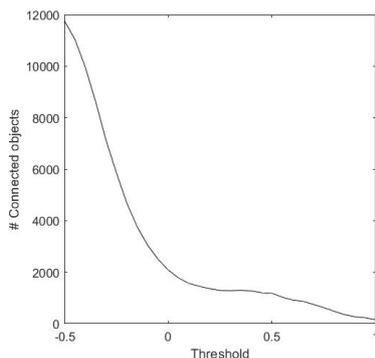


Figura 2.13: Grafico per la scelta della soglia ottimale

Per la segmentazione vengono utilizzate due soglie, definite a partire dalla curva rappresentante il numero di elementi connessi individuati per ogni soglia:

- **Hard threshold:** è la soglia più elevata delle due quindi più specifica. Viene utilizzata per individuare la presenza o meno di artefatti nell'immagine ed è così definita:

$$t_{hard} = \max(t(C_t \geq 0.8 \cdot \text{mean}(C_t))) \quad (2.5)$$

dove C_t indica il numero di elementi connessi per la soglia t . L'immagine viene quindi preliminarmente segmentata con la soglia 'hard'. Se l'area maggiore individuata nella maschera presenta un numero di pixel superiore a $3 \cdot 10^4$ si procede con la segmentazione utilizzando la soglia 'soft' giacché è stato identificato un artefatto all'interno dell'immagine. In caso contrario, la maschera di segmentazione finale sarà completamente nera.

- **Soft threshold:** è la soglia minore delle due quindi più sensibile. Viene utilizzata per la segmentazione finale dell'artefatto di tissue fold nell'immagine ed è così definita:

$$t_{soft} = \text{mean}(t(C_t \geq 0.8 \cdot \text{mean}(C_t))) \quad (2.6)$$

Dopo la segmentazione con la *soft threshold* si procede con il postprocessing. Vengono pertanto eliminate tutte le aree inferiori a 10^4 e chiusi i buchi della maschera. Successivamente, vengono rimosse tutte le aree inferiori al 5% dell'area massima individuata al fine di eliminare aree piccole non di interesse. Infine, vengono eseguite

in ordine delle operazioni di chiusura e apertura rispettivamente con elemento morfologico strutturale circolare di raggio 6 e 4 ed eliminate nuovamente le aree inferiori a 10^4 create a seguito delle operazioni morfologiche. La maschera finale ottenuta (figura 2.14) viene quindi utilizzata come criterio di esclusione per l'estrazione smart delle patch.

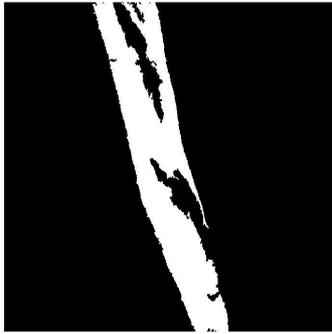


Figura 2.14: Maschera finale dell'algoritmo di tissue fold detection

Blur detection

L'algoritmo di blur detection ha come scopo primario l'individuazione di zone sfocate e on-focus all'interno delle Tiles da utilizzare come ulteriore criterio di estrazione delle patch. L'artefatto di sfocatura può infatti influenzare negativamente il training della rete e le successive analisi del tessuto. Per il riconoscimento di parti out-of-focus nelle immagini si utilizzano due criteri: la varianza del Laplaciano e il contrasto di Haralick.

Il primo step dell'algoritmo consiste nel padding a specchio dell'immagine (*mirror padding*). L'immagine così ottenuta viene quindi letta utilizzando delle sliding window di 500x500 con un passo di 250. Segue la creazione della mappa di focus mediante il criterio della varianza [42]. La patch 500x500 viene convoluta con un filtro Laplaciano con kernel così definito:

$$h = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (2.7)$$

Una volta ottenuta l'immagine filtrata ne viene calcolata la varianza ed il suo valore viene salvato nel crop centrale 250x250 della patch. Facendo nuovamente scorrere la sliding window si ripete il procedimento fino alla creazione della prima mappa di focus come in figura 2.15b.

Si procede con la creazione della seconda mappa di focus a partire dalla patch 500x500 considerando come criterio il contrasto di Haralick [43]. In primo luogo, l'immagine viene convertita da RGB in toni di grigio e ne vengono saturati l'1% dei

pixel più scuri e l'1% dei pixel più chiari. Viene quindi calcolata la matrice delle co-occorrenze **GLCM**.

Data un'immagine I di dimensione $n \times m$ la GLCM con offset $(\Delta x, \Delta y)$ è definita come:

$$C_{\Delta x \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (2.8)$$

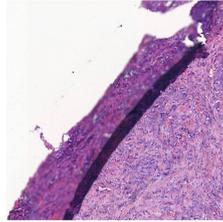
La sua normalizzata P si ottiene dividendo C per la somma degli elementi di C :

$$P_{\Delta x \Delta y}(i, j) = \frac{C_{\Delta x \Delta y}(i, j)}{\sum_{k,l} C_{\Delta x \Delta y}(k, l)} \quad (2.9)$$

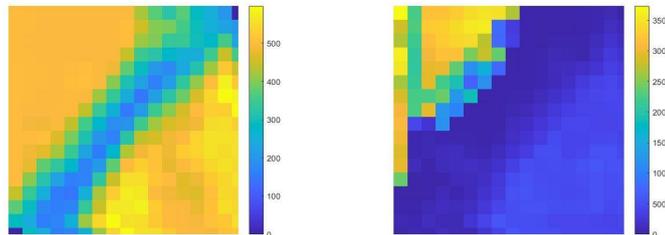
Il contrasto di Haralick HC viene conseguentemente calcolato come:

$$HC_{\Delta x \Delta y}(i, j) = \sum_{i=0}^L \sum_{j=0}^L (i - j)^2 P_{\Delta x \Delta y}(i, j) \quad (2.10)$$

dove per $(\Delta x, \Delta y)$ si è utilizzata la coppia (1,1). Analogamente al procedimento eseguito per la varianza, il valore di HC viene sostituito nel crop centrale 250x250 e si prosegue con lo scorrimento della patch fino al conseguimento della seconda mappa di focus (figura 2.15c).



(a) Immagine iniziale



(b) Mappa basata sulla varianza (c) Mappa basata sul contrasto di Haralick

Figura 2.15: Creazione mappe di focus

Per l'ottenimento della mappa finale di focus (figura 2.16) vengono quindi scelte delle soglie fisse per entrambi i criteri sui rispettivi crop centrali della patch.

$$quality = \begin{cases} 1, & \text{if } HC \geq 150 \text{ and } \sigma^2 \geq 400 \\ 0, & \text{if } HC < 150 \text{ and } \sigma^2 < 400 \\ 0.5, & \text{otherwise} \end{cases} \quad (2.11)$$

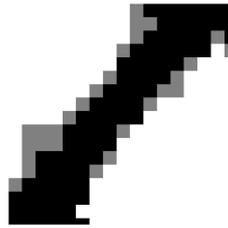


Figura 2.16: Mappa finale di focus

Estrazione finale delle patch

Ottenute quindi le maschere dei nuclei cellulari, delle porzioni bianche dell'immagine, dei tissue folds e la mappa di focus; le patch vengono estratte utilizzando una *sliding window* con le seguenti caratteristiche:

- Dimensione patch = 500x500
- Stride = 50

La selezione avviene se sono validi i seguenti requisiti all'interno della patch:

1. Area delle regioni bianche inferiore al 50% dell'area totale della patch;
2. Area dei nuclei segmentati almeno pari al 10% dell'area totale della patch;
3. Area dei tissue fold inferiore al 50% dell'area totale della patch;
4. Patch on focus: 50% dei pixel della mappa di focus con intensità superiore a 0.5.
5. Annotazione manuale (benigna o tumorale) almeno pari al 90% dell'intera patch.

Se la patch soddisfa tutte le richieste viene quindi salvata insieme alla label della classe considerata come unica all'interno della patch (figura 2.17):

- Label 0: classe benigna (colore verde della patch)
- Label 1: classe tumorale (colore arancione della patch)

Inoltre, allo scopo di ottenere un dataset eterogeneo, con patch provenienti da più WSI e quindi da pazienti diversi viene limitato il numero totale di patch estratte

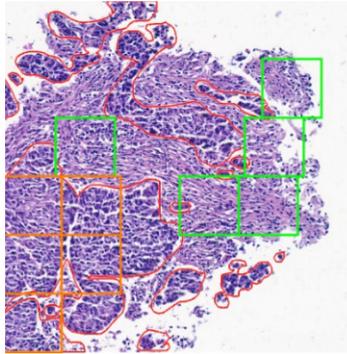


Figura 2.17: Patch estratte: in verde patch benigne, in arancione patch maligne, in rosso annotazione manuale

da ogni WSI per ciascuna delle due classi. In tale modo la rete sarà capace di riconoscere pattern della stessa classe anche in immagini diverse. In conclusione, sono state ottenute 23716 patch di dimensione 500x500 così distribuite:

- Patch benigne: 11336;
- Patch maligne: 12380;

2.4 Reti neurali convoluzionali CNN

Il corpo centrale dell'algoritmo, nonché obiettivo principale del progetto di tesi, consiste nell'allenamento dei seguenti tre modelli neurali:

1. VGG16;
2. ResNet50;
3. DenseNet121;

Il training delle tre reti ha lo scopo di distinguere accuratamente il carcinoma polmonare dal tessuto sano all'interno delle patch selezionate nella fase di preprocessing. Il risultato delle reti costituisce la base per la realizzazione delle heatmap finali, fondendo la predizione dei modelli tra loro.

Prima di procedere con la descrizione dei parametri per l'allenamento dei tre modelli proposti, si rende di fatto necessaria un'introduzione sulla teoria e sul funzionamento di una *Rete Neurale Convoluzionale* generica nel paragrafo 2.4.1. In seguito, nel paragrafo 2.4.2, vengono presentate le architetture delle tre reti con le relative caratteristiche.

2.4.1 Introduzione al Deep Learning e alle CNN

Il *deep learning* è una tecnica di Machine Learning basata su reti neurali artificiali per l'apprendimento di task di varia natura. Gli algoritmi di deep learning simulano il processo di apprendimento del cervello umano fondato sull'esperienza. L'obiettivo principale di questi modelli è il riconoscimento di oggetti di diverso tipo come immagini, testi, suoni ecc. a partire da esempi di riferimento. I modelli di questo tipo sono quindi in grado di immagazzinare delle conoscenze per poi renderle disponibili all'uso.

Le reti neurali artificiali sono allenate su set di dati grandi dimensioni normalmente già classificati da altri. L'apprendimento di questo tipo viene definito supervisionato ed è volto alla minimizzazione degli errori tra la predizione del modello e la realtà. Come detto, le reti neurali imitano il processo di apprendimento umano caratterizzato da neuroni e sinapsi. La struttura che ne consegue è quella di un grafo in cui i nodi rappresentano i neuroni del cervello e gli archi che li collegano le sinapsi umane, come mostrato in Figura 2.18.

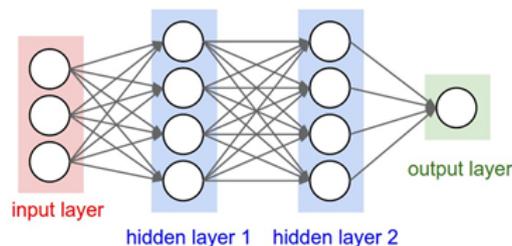


Figura 2.18: Schema di una generica rete neurale [44]

Una rete neurale generica si compone di tre livelli detti layer:

- Input layer: livello che riceve in input le informazioni dall'ambiente esterno;
- Hidden layer: livello di collegamento volto all'apprendimento delle features dei dati;
- Output layer: livello finale di produzione dell'output della rete;

Ciascun livello è formato da molti neuroni con la propria funzione di attivazione collegati tra loro da connessioni pesate.

Le reti neurali convoluzionali o CNN (*Convolutional Neural Network*) rappresentano l'architettura di rete più utilizzata per l'analisi delle immagini. Le CNN sono delle reti di 'Feed Forward', così chiamate perché il flusso delle informazioni avviene in un'unica direzione. Ogni neurone è caratterizzato da pesi W e da bias b che vengono modificati nel corso dell'allenamento mediante un processo di retropropagazione dell'errore (*backpropagation*). L'apprendimento avviene infatti minimizzando, ad ogni iterazione, la cost function data dall'errore tra l'output atteso e

quello generato dalla rete:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}_i, y_i) \quad (2.12)$$

dove L è la funzione di perdita che può essere di vario tipo, \hat{y}_i è l'output generato dall' i -esimo input e y_i è l'output atteso. I blocchi principali di una CNN sono: blocco convoluzionale o feature learning e il blocco di classificazione come mostrato in Figura 2.19.

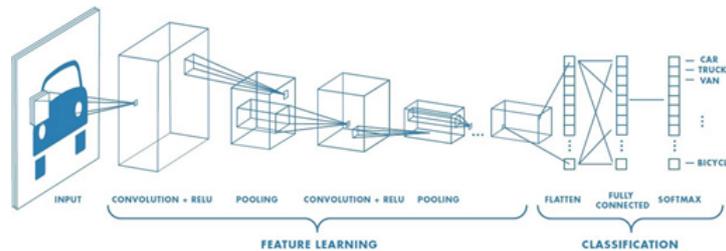


Figura 2.19: Struttura di una CNN [45]

Il primo blocco è adibito al riconoscimento di schemi all'interno dell'immagine via via più complessi con l'aumentare dei livelli. Al primo blocco appartengono i seguenti layer caratteristici [46]:

- *Convolutional Layer* – Strato convoluzionale: layer principale di una CNN, da cui ne deriva il nome, volto al riconoscimento di pattern nell'immagine di input. Questo layer attua un'operazione matematica di convoluzione tra l'immagine ed un filtro, kernel, che scorre su di essa ottenendone un'immagine di uscita. La porzione dell'input analizzata ad ogni spostamento del kernel viene chiamata 'campo ricettivo'. La dimensione dell'immagine di uscita, detta *mappa di attivazione*, dipende dalla scelta di quattro iperparametri:
 1. Numero di filtri: da cui dipende la profondità del volume.
 2. Dimensione del kernel e stride: definiscono l'altezza e la larghezza della mappa di attivazione. Lo stride indica il passo con cui il kernel si sposta sull'immagine di input per ogni convoluzione.
 3. Padding: aggiunge una cornice di pixel all'immagine al fine di evitare di perdere informazioni al passaggio da un livello ad un altro. Solitamente si attua lo zero-padding con cui vengono aggiunti degli 0.

Dopo l'operazione di convoluzione normalmente viene applicata una funzione di attivazione per modificare l'output dello strato precedente. La più utilizzata è la ReLu che impone a 0 tutti i valori negativi dell'output, ma ne esistono di vario tipo.

- *Pooling Layer* – Strato di raggruppamento: viene utilizzato per semplificare l'output del livello precedente riducendo il numero di parametri della rete.

Anche per questo layer è necessario specificare le dimensioni del filtro. Le tipiche operazioni sono quelle di max pooling o di average pooling. In questo modo l'output presenterà una dimensione ridotta preservandone la profondità.

Il secondo blocco si occupa della classificazione vera e propria ed è costituito da:

- *Fully connected layers*: sono presenti solitamente come ultimi livelli di un'architettura neurale e sono costituiti da neuroni connessi a tutti i neuroni degli altri layers come nelle generiche reti neurali.

Transfer Learning

L'allenamento di una rete può avvenire sostanzialmente in due modi:

- Training from Scratch: allenamento da zero
- Transfer Learning: utilizzo di un modello pre-allenato per un task come base per un task simile.

Il transfer learning costituisce un approccio di allenamento molto utilizzato di recente poiché permette una riduzione delle risorse computazionali e la creazione di modelli anche con esigui dataset. Per rendere il modello capace di eseguire il nuovo task è necessario rimuovere il layer di classificazione e la softmax ed adattarla al nuovo obiettivo (*fine-tuning*). A questo punto si può procedere in vari modi, come in Figura 2.20 [47] :

1. Training della rete totale: richiede grandi dimensioni del dataset;
2. Training di alcuni layer lasciandone altri congelati: il numero dei layer da allenare aumenta all'aumentare della dimensione del dataset;
3. Training del solo layer di classificazione: si può ricorrere a questa strategia se si hanno pochi dati e/o poca potenza oppure se il modello pre-allenato è simile al task richiesto.

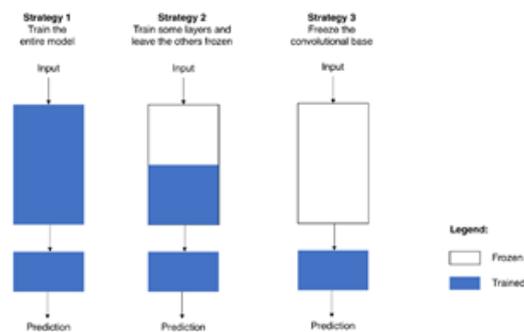


Figura 2.20: Tre tipologie di transfer learning [47]

Problemi del training: overfitting e underfitting

Nella fase di allenamento risulta fondamentale comprendere quali possono essere i possibili errori di predizione che provocano una scarsa capacità di generalizzazione della rete. In tal senso è bene conoscere i concetti di bias e di varianza dal momento che tra questi esiste il cosiddetto problema del *tradeoff* nel cercare di minimizzarli [48].

- **Bias:** è la differenza tra la previsione media e il valore atteso. Rappresenta l'errore sistematico non derivante dalla casualità. Un modello con bias elevato presenta delle scarse prestazioni sul training set.
- **Varianza:** è la variabilità di previsione del modello. Rappresenta la sensibilità del modello alla casualità dei dati. Un modello con elevata varianza non è in grado di generalizzare e pertanto risente di scarse prestazioni sul validation set.

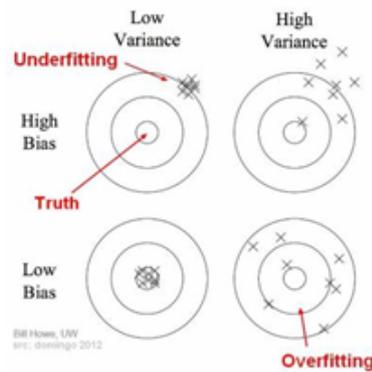


Figura 2.21: Possibili situazioni di training [48]

In Figura 2.21 sono riportate le situazioni che possono presentarsi in fase di training. Il modello perfetto è caratterizzato da bassa varianza e basso bias (in basso a sinistra) mentre il modello peggiore presenta valori elevati sia di varianza che di bias (in alto a destra). Tra queste due casistiche si presentano altre due situazioni intermedie che possono essere corrette adottando varie soluzioni.

1. **Underfitting:** si verifica quando il modello presenta elevato bias e bassa varianza. In questa condizione, la rete, a causa della sua semplicità, non riesce a comprendere i pattern dei dati presentati. Le possibili soluzioni sono:
 - Utilizzo di reti con più layer;
 - Modifica dell'architettura della rete;
 - Training più lunghi.

2. **Overfitting**: si verifica quando il modello presenta basso bias ed elevata varianza. In questa situazione, la rete, a causa della sua sensibilità al rumore dei dati, non è in grado di generalizzare adeguatamente. Le possibili soluzioni sono:

- Aumento dei dati a disposizione per il training. L' incremento dei dati può essere realizzato anche in modo 'fittizio' attraverso tecniche di *data augmentation* che eseguono rotazioni, flip, distorsioni, zoom, modifiche di colore ecc sui dati del training.
- Suddivisione del dataset in *Training set* e *Validation set*. Questo può essere eseguito mediante cross-validation per visualizzare il comportamento della rete quando vengono presentati esempi diversi durante il training.
- Early Stopping: l'allenamento viene interrotto nel momento in cui la rete inizia a sovra utilizzare i dati del training set.
- Dropout: nel corso dell'allenamento vengono casualmente spenti alcuni neuroni in modo da rendere la rete meno sensibile al peso di alcuni neuroni a discapito di altri.
- Regolarizzazione: consiste nell'aggiunta di un termine alla funzione di costo. Tra le più utilizzate ci sono la regolarizzazione L1 e L2. La funzione di costo con la regolarizzazione L2 viene così modificata:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}_i, y_i) + \lambda \|w^2\| \quad (2.13)$$

dove λ rappresenta il parametro di regolarizzazione.

2.4.2 Architettura reti: VGG, ResNet50, DenseNet

VGG

L'architettura della VGG [49] è stata realizzata nel 2014 da Karen Simonyan and Andrew Zisserman della Visual Geometry Group nell'Università di Oxford. Le versioni più utilizzate della VGG sono la VGG-16 e la VGG-19 che si distinguono per il numero di layer. La rete si ispira alla precedente AlexNet ma presenta un numero maggiore di layer e dei kernel con campi ricettivi più piccoli (3x3 con uno stride di 1 contro 11x11 o 5x5 dell'AlexNet). Ne deriva un numero di parametri pari a 138 milioni.

La dimensione dell'input di una VGG è di 224x224x3. I kernel impiegati presentano una dimensione di 3x3 con stride 1 e, in alcuni layer, 1x1 con stride 1 per realizzare una sorta di trasformazione lineare. L'architettura della rete, come mostrato Tabella 2.1, presenta: 5 blocchi di layer convoluzionali 3x3 con numero di kernel crescenti dal blocco 1 al blocco 4 e costante per il blocco 5. Il numero di kernel è sempre una potenza di 2 compresa tra 64 e 512. Ci sono sostanzialmente 2 layer

convoluzionali nei primi 2 blocchi e 3 (nella VGG-16) o 4 (nella VGG-19) negli ultimi 3. Tra un blocco e l'altro sono inseriti dei layer di maxpooling. In ultimo, è presente un blocco con 3 fully connected layers di rispettivamente 4096, 4096 e 1000 neuroni.

Tabella 2.1: Architettura VGG-16 e VGG-19

VGG16	VGG19
input 224x224	input 224x224
Conv 3x3,64	Conv 3x3,64
Conv 3x3,64	Conv 3x3,64
maxpool	maxpool
Conv 3x3,128	Conv 3x3,128
Conv 3x3,128	Conv 3x3,128
maxpool	maxpool
Conv 3x3,256	Conv 3x3,256
Conv 3x3,256	Conv 3x3,256
Conv 3x3,256	Conv 3x3,256
	Conv 3x3,256
maxpool	maxpool
Conv 3x3,512	Conv 3x3,512
Conv 3x3,512	Conv 3x3,512
Conv 3x3,512	Conv 3x3,512
	Conv 3x3,512
maxpool	maxpool
Conv 3x3,512	Conv 3x3,512
Conv 3x3,512	Conv 3x3,512
Conv 3x3,512	Conv 3x3,512
	Conv 3x3,512
Maxpool	Maxpool
FC-4096	FC-4096
FC-4096	FC-4096
FC-1000	FC-1000
softmax	softmax

ResNet50

La ResNet [50] è una CNN profonda molto utilizzata nell'ambito della classificazione di immagini. Nel 2015 la ResNet si è imposta al primo posto nella competizione *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) per la classificazione del dataset ImageNet con oltre un milione di immagini in 1000 categorie diverse. Esistono diverse varianti in base al numero di layers. La versione utilizzata nella tesi è la ResNet50.

La ResNet è stata introdotta per cercare di risolvere il problema del *vanishing gradient* (scomparsa del gradiente) che si verifica per reti molto profonde. Per queste reti l'aggiornamento dei pesi dei layer vicini all'input (o in alcuni casi di tutti i layer) durante il backpropagation è molto ridotto, quasi tendente a zero. A causa di questo problema, i layer iniziali non apprendono effettivamente le features. Con la ResNet è stato introdotto il concetto di skip connections o identity connection. Le skip connection, o blocchi residui, come mostrato in Figura 2.22, consistono nell'aggiungere l'input x all'output di alcuni layer successivi.

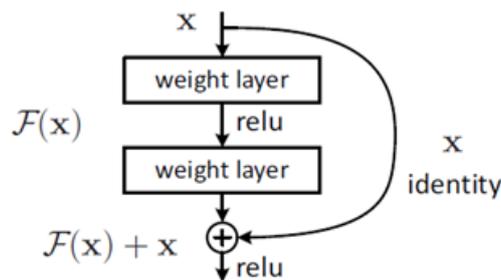


Figura 2.22: Skip connection [50]

Per realizzare queste connessioni è necessario che l'input e $F(x)$ abbiano la stessa dimensione, ad esempio usando il padding dell'input. A questo punto, durante il backpropagation i gradienti, grazie alle skip connection, non devono incontrare necessariamente dei 'livelli di peso' e quindi vengono preservati da qualsiasi modifica. Il gradiente salta i blocchi e può raggiungere gli strati iniziali. La dimensione dell'input di una ResNet50 è di $224 \times 224 \times 3$. Il primo blocco è costituito da un layer convoluzionale con 64 kernel di dimensioni 7×7 e stride di 2 e un max pooling layer con kernel 3×3 e stride 2. La rete si articola sostanzialmente in 4 stages con rispettivamente 9 layer, 12 layer, 18 layer e 9 layer. I blocchi residui (in numero 3,4,6,3 per ogni stage) contengono sempre 3 layer ciascuno. I 3 layer convoluzionali di ogni blocco residuo sono così costituiti:

- Layer convoluzionale 1×1 : con numero di kernel per stage 64, 128, 256, 512
- Layer convoluzionale 3×3 : con numero di kernel per stage 64, 128, 256, 512
- Layer convoluzionale 1×1 : con numero di kernel per stage 256, 512, 1024, 2048

Tra uno stage ed il successivo il layer convoluzionale utilizza uno stride di 2. Dopo l'ultimo stage, si ha un blocco con un layer di average pooling e un fully connected layer con 1000 neuroni. Infine, si ottiene la softmax. In totale si hanno 50 layer : 1+ 9+12+18+9+1 come semplificato in Tabella 2.2, per un totale di 23 milioni di parametri.

Tabella 2.2: Architettura ResNet50

Layer_name	50-layer
Conv1	7x7, 64, stride 2 3x3 max pool, stride 2
Conv2	$\begin{bmatrix} 1x1 & 64 \\ 3x3 & 64 \\ 1x1 & 256 \end{bmatrix} \times 3$
Conv3	$\begin{bmatrix} 1x1 & 128 \\ 3x3 & 128 \\ 1x1 & 512 \end{bmatrix} \times 4$
Conv4	$\begin{bmatrix} 1x1 & 256 \\ 3x3 & 256 \\ 1x1 & 1024 \end{bmatrix} \times 6$
Conv4	$\begin{bmatrix} 1x1 & 512 \\ 3x3 & 512 \\ 1x1 & 2048 \end{bmatrix} \times 3$
	Average pooling 1000 fc softmax

DenseNet

La rete di DenseNet [52] è una delle innovazioni nel campo del machine learning. Risulta molto simile alla struttura della ResNet ma invece di utilizzare le identità, sommando 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. La DenseNet ha quindi accesso ai gradienti dalla funzione di perdita e dall'input. 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 dei seguenti layer:

- Batch Normalization
- ReLU
- Layer Convoluzionale 3x3

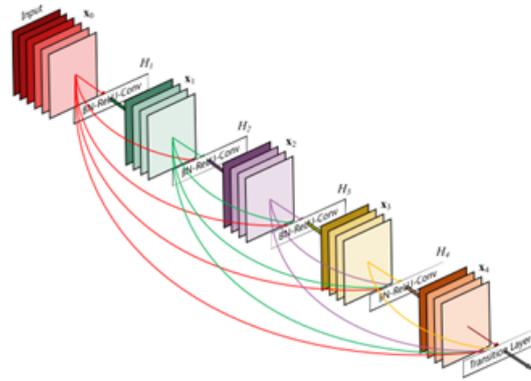


Figura 2.23: Rappresentazione architettura DenseNet [52]

Dal momento che le mappe di feature vengono concatenate tra loro è necessario che abbiano la stessa dimensione ma, d'altro canto, è anche fondamentale il down-sampling, Per cui tra due blocchi di Dense viene inserito un layer di transizione:

- Batch Normalization
- Layer Convoluzionale 1x1
- Average pooling 2x2

L'architettura per diverse tipologie di DenseNet è mostrata in Figure 2.24 .

Layers	Output Size	DenseNet-121($k = 32$)	DenseNet-169($k = 32$)	DenseNet-201($k = 32$)	DenseNet-161($k = 48$)
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	1×1 conv 3×3 conv $\times 6$	1×1 conv 3×3 conv $\times 6$	1×1 conv 3×3 conv $\times 6$	1×1 conv 3×3 conv $\times 6$
Transition Layer (1)	56×56 28×28	1×1 conv 2×2 average pool, stride 2			
Dense Block (2)	28×28	1×1 conv 3×3 conv $\times 12$	1×1 conv 3×3 conv $\times 12$	1×1 conv 3×3 conv $\times 12$	1×1 conv 3×3 conv $\times 12$
Transition Layer (2)	28×28 14×14	1×1 conv 2×2 average pool, stride 2			
Dense Block (3)	14×14	1×1 conv 3×3 conv $\times 24$	1×1 conv 3×3 conv $\times 32$	1×1 conv 3×3 conv $\times 48$	1×1 conv 3×3 conv $\times 36$
Transition Layer (3)	14×14 7×7	1×1 conv 2×2 average pool, stride 2			
Dense Block (4)	7×7	1×1 conv 3×3 conv $\times 16$	1×1 conv 3×3 conv $\times 32$	1×1 conv 3×3 conv $\times 32$	1×1 conv 3×3 conv $\times 24$
Classification Layer	1×1	7×7 global average pool 1000D fully-connected, softmax			

Figura 2.24: Architettura DenseNet [52]

2.4.3 Allenamento delle CNN

Le CNN di classificazione sono state allenate in ambiente Python mediante la piattaforma online Kaggle. Il dataset per la classificazione del tessuto polmonare per tutte le reti prevede 23716 patch di dimensione 500x500. Al fine di ottenere un modello robusto, il dataset è stato diviso in training, validation e test in base alla WSI di appartenenza della patch. In questo modo, a differenza di una suddivisione del dataset completamente casuale, si evita che nel test ci siano immagini provenienti da vetrini del training. Questa scelta permette di ottenere dei modelli più robusti in grado di riconoscere effettivamente il tessuto tumorale in immagini completamente sconosciute alla rete. Per il training sono quindi state utilizzate patch provenienti da 110 WSI (78 del Training 1 e 32 del Training 2), per il validation 10 WSI (7 del Training 1 e 3 del Training 2) e per il test 30 (15 del Training 1 e 15 de Training 2).

La suddivisione finale in termini di patch prevede quindi:

Tabella 2.3: Suddivisione Dataset in training, validation e test.

Training	Validation	Test
16583	1371	5762
8283 Ben e 8300 Mal	623 Ben e 748 Mal	2305 Ben e 3457 Mal

L'allenamento è supervisionato, ovvero vengono fornite in input alla rete le immagini con le rispettive label in formato binario:

- *classe 0*: benigna o sana.
- *classe 1*: maligna o tumorale.

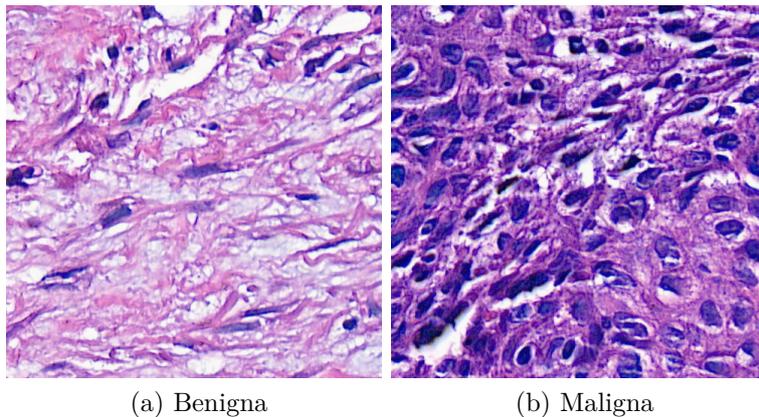


Figura 2.25: Esempio di patch date in input alle reti

Allenamento VGG

L'allenamento delle patch è stato eseguito utilizzando la VGG-16 a cui sono stati sostituiti i *fully connected layers* con rispettivamente un layer di *Global Average Pooling 2D*, un layer di *Dropout* con rate 0.65 e due layer di Dense: il primo con 128 neuroni e funzione di attivazione ReLU ed il secondo con un neurone e funzione di attivazione sigmoide per la classificazione binaria. I parametri utilizzati per il training sono descritti in Tabella 2.4.

Tabella 2.4: Parametri di training VGG

Metodo di ottimizzazione	Nadam (learning rate= $5e^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 0.1$)
Funzione di perdita	Binary Crossentropy
Numero di epoche	200
Input size	224x224x3
Batch size	128
Data augmentation	Flip orizzontale e verticale, rotazione con range di 180°
Callbacks	Early Stopping se dopo 15 epoche l'accuracy sul validation non aumenta di 0.01
Numero di seeds	5

Le immagini vengono preliminarmente divise in training set e validation set. Inoltre i pesi del Dense layer vengono moltiplicati per un fattore dipendente dal numero delle immagini per ogni classe. Questa scelta deriva dalla maggior numerosità della classe benigna rispetto alla maligna. Le immagini di training vengono successivamente aumentate effettuando operazioni di flip lungo l'asse x e lungo l'asse y e rotazione in un range di 180° . In seguito, alle patch 500x500 viene effettuato un resize per adattarle alla dimensione di input standard della VGG-16 pari a 224x224. Segue l'allenamento per il quale è stato scelto come algoritmo di ottimizzazione Nadam (Nesterov-accelerated Adaptive Moment Estimation) in cui i pesi della rete vengono aggiornati in modo da ridurre la funzione d'errore o di perdita. L'aggiornamento dei pesi segue la formula:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} (\beta_1 \hat{m}_t + \frac{(1 - \beta_1)g_t}{1 - \beta_1^t}) \quad (2.14)$$

dove w_t rappresenta il peso all'iterazione corrente, w_{t+1} il peso alla nuova iterazione, η il learning rate, \hat{m}_t e \hat{v}_t sono degli stimatori corretti del momento primo (m) e secondo (v) del gradiente, ϵ è un termine di smoothing aggiunto ai fini della stabilità numerica e β_1 e β_2 (contenuto nel valore di \hat{v}_t) sono iperparametri che controllano il decadimento esponenziale e g_t rappresenta il gradiente della funzione di costo J all'iterazione corrente.

$$g_t = \nabla J_t(w_t) \quad (2.15)$$

La funzione di perdita utilizzata è la *binary cross entropy* definita come :

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (2.16)$$

Si è scelto, inoltre, di utilizzare dei mini-batch di dimensione 128. L'epoca indica il passaggio di tutto il set di training mentre l'iterazione del solo mini batch. Per tale ragione, il numero di iterazioni per ogni epoca, usando questa dimensione per il mini batch corrisponde a 130. Il numero di epoche massimo per l'allenamento è pari a 200 ed al termine di ciascuna i dati del training e del validation vengono rimescolati. Per il training si è utilizzato il fine-tuning per il quale vengono congelati i primi 6 layer, con i pesi della rete pre-allenata su Imagenet, e allenati i successivi. Viene inoltre adottato l'early stopping come criterio per fermare l'allenamento nel momento in cui non si verichi più un miglioramento dell'accuratezza del validation set per 15 epoche consecutive rispetto al valore massimo raggiunto.

Allenamento ResNet50

Come per la VGG, per l'allenamento della ResNet sono stati sostituiti i *fully connected layers* con rispettivamente un layer di *Global Average Pooling 2D*, un layer di *Dropout* con rate 0.6 e due layer di Dense: il primo con 128 neuroni e funzione di attivazione ReLU ed il secondo con un neurone e funzione di attivazione sigmoide per la classificazione binaria. I parametri utilizzati per il training sono descritti in Tabella 2.5.

Tabella 2.5: Parametri di training ResNet50

Metodo di ottimizzazione	Nadam (learning rate= $5e^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 0.1$) con learning rate decay di un fattore pari a $e^{-0.1}$ a partire dall'epoca 10
Funzione di perdita	Binary Crossentropy
Numero di epoche	200
Input size	224x224x3
Batch size	128
Data augmentation	Flip orizzontale e verticale, rotazione con range di 90° , Zooming compreso tra -50% e +50% Modifica luminosità compresa tra -50% e +50%
Callbacks	Early Stopping se dopo 15 epoche l'accuracy sul validation non aumenta di 0.01
Numero di seeds	5

L'allenamento viene condotto seguendo i passi implementati per la rete VGG. Le uniche differenze risiedono nell'adozione, in questo caso, di maggiori tecniche di

data augmentation e del learning rate decay a partire dall'epoca 10. Inoltre per la ResNet50 vengono congelati i primi 100 layer e allenati i successivi.

Allenamento DenseNet

Anche per la DenseNet vengono seguiti gli stessi passi di implementazione delle altre reti. La versione utilizzata per il training è la DenseNet121 in cui i *fully connected layers* finali vengono sostituiti rispettivamente con un layer di *Global Average Pooling 2D*, un layer di *Dropout* con rate 0.4 e due layer di *Dense*: il primo con 128 neuroni e funzione di attivazione ReLU ed il secondo con un neurone e funzione di attivazione sigmoide per la classificazione binaria. I parametri utilizzati per il training sono presentati in Tabella 2.6.

Tabella 2.6: Parametri di training ResNet50

Metodo di ottimizzazione	Nadam (learning rate= $3e^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 0.1$) con learning rate decay di un fattore pari a $e^{-0.1}$ a partire dall'epoca 10 e fino all'epoca 20 poi rimane costante
Funzione di perdita	Binary Crossentropy
Numero di epoche	200
Input size	224x224x3
Batch size	64
Data augmentation	Flip orizzontale e verticale, rotazione con range di 90°
Callbacks	Early Stopping se dopo 15 epoche l'accuracy sul validation non aumenta di 0.01
Numero di seeds	5

Per questa architettura, a differenza della VGG e della ResNet50, è stato utilizzata una dimensione del mini-batch minore e pari a 64. In questo modo per ogni epoca vengono condotte 260 iterazioni. Inoltre, come per la ResNet, vengono congelati i primi 100 layer e si procede con il training dei successivi.

2.4.4 Fusione output delle reti con media ponderata

Una volta ottenuti i modelli con le diverse architetture è necessario combinarne i risultati per ottenere un unico output o previsione. Per la fusione dell'output delle reti si è scelto di utilizzare un algoritmo in grado di ricercare i pesi migliori da attribuire a ciascuna rete al fine di incrementare l'accuratezza totale del modello di insieme. Infatti, la scelta di combinare gli output delle reti in egual misura, con la classica operazione di media, spesso risulta in prestazioni non di molto superiori al miglior modello e in alcuni casi anche inferiori [53]. Ciò che succede più di frequente è che certi modelli sono migliori e si vorrebbe che contribuissero maggiormente nella definizione della predizione finale di insieme ed altri, che pur essendo peggiori,

potrebbero risultare utili ma dovrebbero pesare meno. La media ponderata permette a ciascun modello di contribuire alla predizione finale in proporzione alla sua performance sulle immagini di training. I pesi per il modello di insieme vengono trovati attraverso una ricerca a griglia. In primo luogo, viene definito un vettore di pesi compresi tra $\{0,1\}$ a passi di 0.1. A questo punto, vengono generati con il prodotto cartesiano tutti i possibili vettori di pesi con un numero di elementi pari a quello delle reti allenate. Ogni vettore creato in questo modo viene forzato ad avere somma unitaria dei pesi. In particolare viene calcolata la norma L1. Dato un vettore x :

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \quad (2.17)$$

La norma L1 è definita come:

$$|x|_1 = \sum_{i=1}^n |x_i| \quad (2.18)$$

Il vettore generato viene quindi diviso per la norma L1, in modo che la somma dei pesi sia uguale a 1. A questo punto i vettori dei pesi generati vengono moltiplicati per i vettori delle predizioni di ciascuna rete sulle patch di training. Se il valore i -esimo del vettore è uguale o superiore a 0.5 allora viene considerato come patch tumorale e quindi il valore di probabilità viene sostituito con un 1. In caso contrario la patch è benigna e quindi il valore viene sostituito con 0. Infine, il vettore delle predizioni viene comparato con il vettore contenente le classi reali delle patch, salvando progressivamente la combinazione di pesi che consente di ottenere la migliore accuratezza.

2.5 Testing: Creazione Heatmap

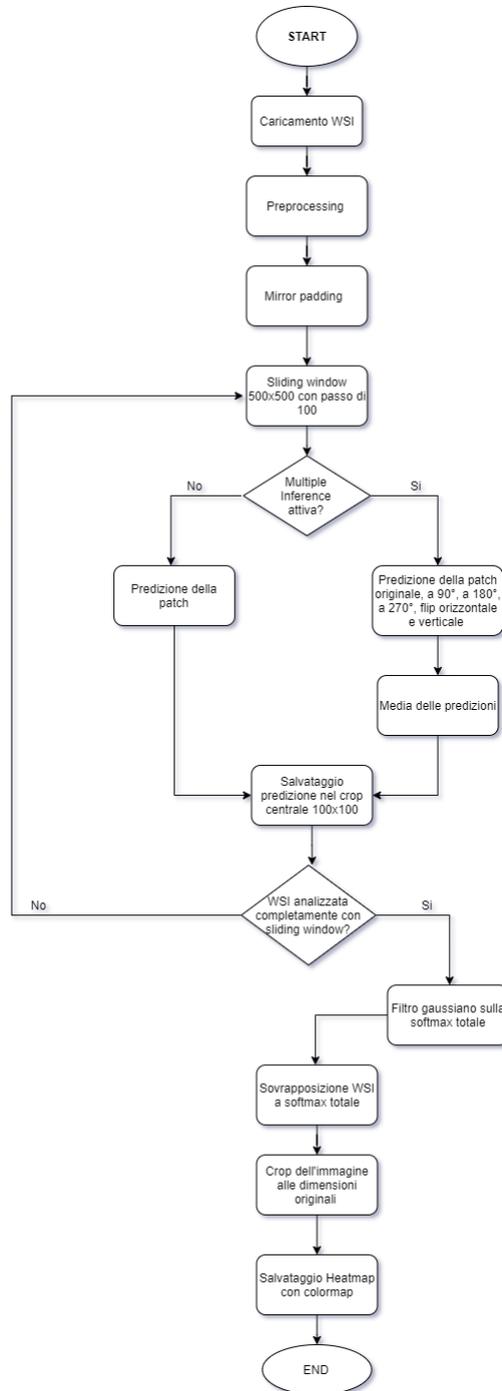


Figura 2.26: Pipeline creazione Heatmap

Nella fase di testing, vengono caricate le immagini dalla cartella di interesse contenente i campi delle WSI e preprocessate secondo il metodo di preprocessing della rete di interesse.

L'immagine viene quindi letta utilizzando delle sliding window di 500x500 con un passo di 100. La patch letta viene classificata dalla rete e al termine viene ricavata la softmax. Per la generazione delle heatmap la predizione della rete viene unicamente sostituita nel quadrato centrale della patch di dimensione 100x100. Con questo procedimento la mappa di probabilità ottenuta risente di una mancanza di informazione ai bordi. Si è proceduto quindi al padding, ampliando l'immagine originale mediante riflesso a specchio (*mirror padding*) di 200x200 pixel in ogni direzione. Sull'immagine estesa viene quindi fatta scorrere la finestra ricavandone la predizione, e di quest'ultima viene mantenuta solo la porzione centrale con un crop size di 100. In questa fase è possibile attivare l'opzione di Multiple Inference per ottenere una predizione più accurata. Se attiva, la patch verrà classificata per sei volte:

1. Patch originale;
2. Patch ruotata di 90°;
3. Patch ruotata di 180°;
4. Patch ruotata di 270°;
5. Patch invertita orizzontalmente;
6. Patch invertita verticalmente;

La predizione finale sarà data dalla media delle predizioni. Alla softmax, ricavata mediante lo scorrimento della finestra su tutta la WSI, è applicato un filtro gaussiano in modo da smussarne i contorni. Infine, l'immagine viene tagliata e riportata alla dimensione originale. L'heatmap si ottiene sovrapponendo la softmax alla WSI originale (come in figura 2.27) e modificando la colormap di default.

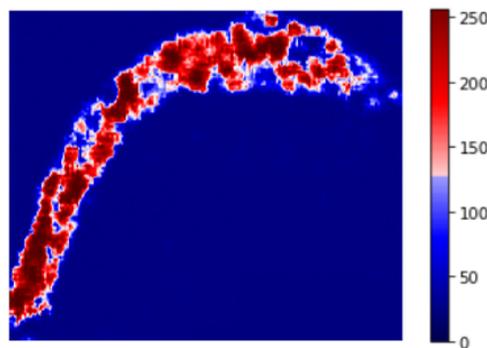


Figura 2.27: Heatmap su un campo della WSI

2.6 Metriche di validazione

Le metriche di validazione vengono utilizzate per quantificare, mediante indicatori numerici, la bontà dei risultati e valutare, quindi, le performances dell'algoritmo proposto. I risultati del metodo esposto nell'ambito del progetto di tesi vengono confrontati con le annotazioni manuali, chiamate *ground truth*, realizzate da un patologo con più di 30 anni di esperienza. La validazione avviene in due fasi. In primo luogo, vengono valutate le prestazioni dei classificatori ottenuti a seguito dell'allenamento sulle patch 500x500 estratte dal dataset ACDC@ LUNGHP con metriche definite *Object-based*. In seguito, viene valutata la bontà delle heatmaps sulle WSI di test del dataset ACDC@ LUNGHP con metriche definite *Pixel-based*.

2.6.1 Metriche di validazione delle performance delle reti

Le metriche di validazione utilizzate in questa prima fase hanno come scopo primario quello di quantificare le prestazioni delle CNN ottenute a seguito del training. Nei problemi di classificazione, le metriche confrontano la classe attesa o reale con quella predetta dal modello o, in alternativa, valutano la probabilità di appartenenza ad una determinata classe. A tale scopo vengono estratte delle patch di dimensione 500x500 con una label che descrive la classe reale di appartenenza:

- 0: classe normale o non tumorale
- 1: classe tumorale

Le patch vengono in seguito classificate dal modello al fine di ottenerne la classe predetta. I vettori contenenti le label reali e quelle predette vengono infine confrontati tra loro in termini di Confusion Matrix e curve ROC. Inoltre, per meglio comprendere la modalità di classificazione messa in atto dalla rete vengono utilizzate le tecniche della t-SNE e delle Saliency map.

Confusion matrix

Nell'ambito dell'intelligenza artificiale e nello specifico in un problema di classificazione, la metrica di validazione più impiegata per valutare le prestazioni del modello predittivo è la *confusion matrix* o tavola di contingenza. Una confusion matrix è una tabella in cui le righe rappresentano lo stato effettivo, detto classe reale, mentre le colonne ne rappresentano la previsione o classe predetta (e viceversa) [54]. Gli elementi disposti sulla diagonale principale della matrice costituiscono i corretti classificati, mentre gli altri vengono denominati misclassificati. Le dimensioni della matrice dipendono dal numero di classi totali del dataset. In generale una confusion matrix ha dimensioni $N \times N$ dove N è il numero delle classi. Il caso più comune è quello di classificazione binaria (Vero/Falso, 1/0, Si/No) con dimensioni della matrice 2×2 , come mostrato in Figura 2.28.

		Predicted class	
		<i>P</i>	<i>N</i>
Actual Class	<i>P</i>	True Positives (TP)	False Negatives (FN)
	<i>N</i>	False Positives (FP)	True Negatives (TN)

Figura 2.28: Confusion matrix per classificazione binaria [55]

Nelle classificazioni binarie la classe di interesse viene indicata come ‘positiva’ ovvero 1 e l’altra come ‘negativa’ o 0. Nell’ambito del progetto di tesi la classe positiva è sempre riferita a quella tumorale mentre quella negativa alla benigna o normale. Per meglio definire la bontà della classificazione messa in atto dal metodo di interesse vengono definiti i seguenti termini [56]:

- *True Positive TP – Veri Positivi*: casi in cui gli elementi sono positivi (patch tumorali) e vengono classificati correttamente come positivi;
- *True Negative TN – Veri Negativi*: casi in cui gli elementi sono negativi (patch normali) e vengono classificati correttamente come negativi;
- *False Positive FP – Falsi Positivi*: casi in cui gli elementi sono negativi (patch normali) e vengono classificati erroneamente come positivi (patch tumorali), sono anche detti falsi allarmi;
- *False Negative FN – Falsi Negativi*: casi in cui gli elementi sono positivi (patch tumorali) e vengono classificati erroneamente come negativi (patch normali);

Definiti gli elementi che compongono la matrice, è possibile calcolare i seguenti indicatori numerici:

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

$$\text{Accuratezza} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.19)$$

- **Recall/ Sensibilità/ True positive rate (TPR)**: rappresenta la proporzione delle previsioni positive corrette (TP) sul totale delle istanze positive reali:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.20)$$

- **Precisione/ Valore predittivo positivo (PPV)**: rappresenta la proporzione delle previsioni positive corrette (TP) sul totale delle previsioni positive:

$$Precisione(PPV) = \frac{TP}{TP + FP} \quad (2.21)$$

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

$$Specificità = \frac{TN}{TN + FP} \quad (2.22)$$

- **Valore predittivo negativo (NPV)**: rappresenta la proporzione delle previsioni negative corrette (TN) sul totale delle previsioni negative:

$$NPV = \frac{TN}{TN + FN} \quad (2.23)$$

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

$$F1score = \frac{2 \cdot Recall \cdot Precisione}{Recall + Precisione} \quad (2.24)$$

Curve ROC

Le curve ROC (*Receiver Operating Characteristics*) rappresentano un'altra metrica fortemente impiegata per la valutazione delle prestazioni di un classificatore binario al variare delle soglie impiegate. Si tratta di una metodica sviluppata durante la II Guerra Mondiale per l'analisi delle immagini radar. La curva ROC viene realizzata legando la probabilità di ottenere un risultato vero positivo (TP) alla probabilità di ottenere un risultato falso-positivo (FP) [57]. In altri termini, la curva ROC viene costruita ponendo sull'asse delle ordinate la Sensibilità o *True Positive Rate* e sull'asse delle ascisse il *False positive rate* (1-specificità) che quantifica la probabilità dei cosiddetti falsi allarmi come in Figura 2.29.

In Figura 2.29, vengono inoltre riportati tre esempi di curve ROC. La curva ROC peggiore possibile è la diagonale blu, detta *linea di non discriminazione*, che corrisponderebbe ad una classificazione del tutto casuale (50% di probabilità di appartenere ad una o all'altra classe). I punti immediatamente al di sopra della diagonale rappresentano dei buoni risultati per la classificazione mentre al di sotto della curva risultano esserci dei risultati peggiori della totale casualità.

L'algoritmo di predizione perfetto possibile produrrebbe un punto in corrispondenza della coordinata (0,1) ovvero nel vertice superiore sinistro dello spazio ROC. In questo punto, la sensibilità avrebbe valore massimo e il *False Positive Rate* valore minimo (o specificità massima). Di conseguenza non ci sarebbero né falsi negativi e né falsi positivi [59].

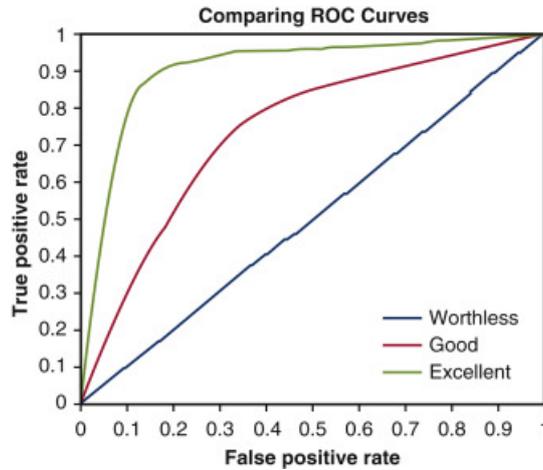


Figura 2.29: Esempi di curve ROC [58]

Ottenuta la curva ROC, è poi possibile calcolarne l'area sottesa alla curva o AUC (*Area Under Curve*), un valore compreso tra 0 e 1 che indica quanto un modello di predizione è in grado di distinguere tra due classi.

Di seguito una possibile interpretazione dei valori di AUC [60]:

- $AUC=0.5$ → il classificatore non è informativo (casuale);
- $0.5 < AUC \leq 0.7$ → il classificatore è poco accurato;
- $0.7 < AUC \leq 0.9$ → il classificatore è moderatamente accurato;
- $0.9 < AUC < 1$ → il classificatore è molto accurato;
- $AUC=1$ → classificatore perfetto.

In sintesi, un modello viene considerato adeguato se presenta un AUC almeno superiore all'80%.

t-SNE

t-SNE (*t-distributed stochastic neighbor embedding*) è una tecnica non supervisionata e non lineare di riduzione della dimensionalità sviluppata da Hinton e Laurens van der Maaten nel 2008. La t-SNE viene ampiamente impiegata per la visualizzazione di dati di grandi dimensioni in uno spazio costituito generalmente da due o tre dimensioni. In sintesi, la tecnica consente di modellare ogni oggetto con un punto organizzando oggetti simili come vicini in un grafico a dispersione e oggetti distinti come lontani, preservandone la struttura locale come in Figura 2.30.

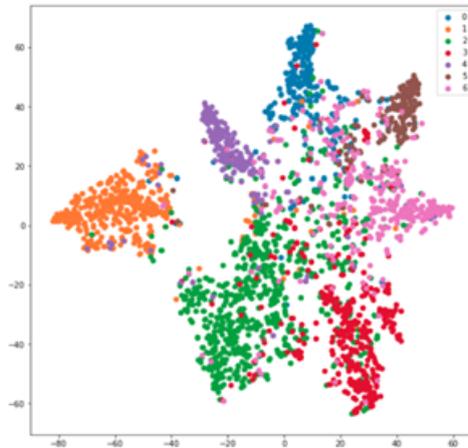


Figura 2.30: Esempio di t-SNE [61]

Dal punto di vista matematico l'algoritmo si articola in due fasi distinte. Durante la prima, viene calcolata una distribuzione di probabilità tra coppie di oggetti ad alta dimensionalità, assegnando una probabilità alta per elementi simili e bassa per elementi diversi. In seconda istanza, viene calcolata una nuova distribuzione di probabilità sui punti a ridotta dimensionalità. I punti nello spazio vengono organizzati, quindi, minimizzando la divergenza di Kullback-Leibler tra le due distribuzioni ottenute [62].

Nel campo dell'apprendimento supervisionato, la t-SNE si è rivelata essere uno strumento utile per comprendere come un modello ragioni. Le reti neurali costituiscono delle black box, modelli in grado di apprendere schemi e/o caratteristiche senza poter comprendere cosa avviene al loro interno. In aggiunta in modelli di questo tipo il numero di neuroni (o feature) è generalmente molto alto. Applicando un algoritmo di riduzione della dimensionalità come il t-SNE all'ultimo layer nascosto della CNN è possibile ottenere una distribuzione grafica della feature apprese. In questo modo è sia possibile visualizzare come le reti modificano la rappresentazione dei dati di input passando attraverso i vari layer, sia rintracciare le osservazioni che la rete misclassifica [63]. La presenza di cluster distinti e separati, ognuno rappresentante di una categoria, indica una buona classificazione, motivo per cui viene utilizzato come metrica di validazione per una CNN.

Saliency map

Le *saliency map* o *occlusion sensitivity map* è una mappa delle parti di un'immagine più rilevanti ai fini della classificazione. Questa tecnica consiste nel misurare la sensibilità di una rete all'occlusione di diverse porzioni dell'immagine di ingresso. Il metodo si basa quindi sulla sostituzione di piccole parti dell'immagine con una maschera grigia opaca, in genere un quadrato. La maschera viene progressivamente fatta scorrere sull'immagine di input e viene calcolata la probabilità di

appartenenza ad una determinata classe per ogni posizione della maschera. Viene quindi calcolata la variazione dei punteggi di predizione in funzione della variazione dell'input secondo la formula seguente [64]:

$$\frac{\partial \text{output}}{\partial \text{input}} \quad (2.25)$$

La formula permette di capire quanto i valori di output si modificano in relazione a piccole variazioni dell'input. In questo modo, la mappa di occlusione mostra le parti dell'immagine che danno un maggiore e minore contributo alla classificazione, come in Figura 2.31. Le parti rilevanti se oscurate portano ad una drastica riduzione della predizione della rete. Viceversa, porzioni meno significative dell'immagine se oscurate non portano ad una modifica netta della capacità di predizione o, in alternativa, possono aumentare le prestazioni per un'altra classe. Spesso, infatti, queste aree sono rilevanti per altre classi e possono confondere la rete [65].

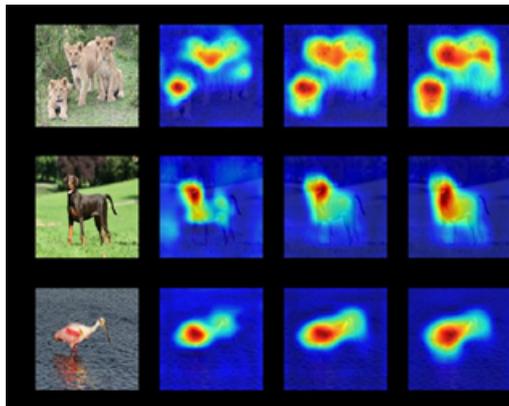


Figura 2.31: Esempio di saliency map per identificare le parti più rilevanti per una CNN [66]

2.6.2 Metriche di validazione delle Heatmap

Le metriche di validazione utilizzate durante la seconda fase di testing hanno come obiettivo la quantificazione delle performance dell'algorithm di creazione delle heatmap. A tale scopo vengono classificate le 30 WSI appartenenti al test utilizzando delle sliding window di dimensione 500x500 e con passo di 100. A partire dalla classificazione viene salvato solo il crop centrale di dimensione 100x100 ottenendo le cosiddette mappe di probabilità o heatmap. A questo punto, alle heatmap viene applicata una soglia (o *threshold*) per segmentarle e ottenere delle annotazioni in bianco e nero. Le maschere binarie delle heatmap vengono confrontate con le annotazioni (o *ground truth*) realizzate da un patologo esperto. A differenza della

validazione delle reti, in questo caso il confronto tra ground truth e annotazione automatica avviene a livello del singolo pixel dell'immagine.

Per tale motivo i termini della Figura 2.28 assumono un significato basato sulla classificazione dei pixel e non più sulla classe della patch:

- *True Positive TP – Veri Positivi*: somma dei pixel dell'intersezione tra l'annotazione manuale e quella automatica (pixel classificati correttamente come tumorali);
- *True Negative TN – Veri Negativi*: somma dei pixel dell'intersezione tra l'annotazione manuale negata e quella automatica negata (pixel classificati correttamente come normali);
- *False Positive FP – Falsi Positivi*: somma dei pixel dell'intersezione tra l'annotazione manuale negata e quella automatica (pixel classificati erroneamente come tumorali);
- *False Negative FN – Falsi Negativi*: somma dei pixel dell'intersezione tra l'annotazione manuale e quella automatica negata (pixel classificati erroneamente come normali);

Gli indicatori numerici impiegati in questa fase sono i seguenti:

- **Dice score**: anche definito come Sørensen–Dice index, misura il grado di sovrapposizione tra due immagini binarie [67]:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (2.26)$$

- **Accuratezza**: come espresso in Equazione 2.19
- **Recall**: come espresso in Equazione 2.20
- **Precisione**: come espresso in Equazione 2.21
- **Specificità**: come espresso in Equazione 2.22

Capitolo 3

Risultati

In questo capitolo vengono presentati i risultati dei metodi proposti nel capitolo 2. In primo luogo, vengono mostrati i grafici degli allenamenti delle reti in termini di *accuratezza* e di *loss* (o perdita) sul *training set* e *validation set*. Segue la validazione sulle patch di training e di test con le metriche descritte nel paragrafo 2.6.1 per ciascuno dei modelli proposti e, in ultimo, per il modello ottenuto dalla combinazione delle tre reti singole con media pesata in funzione dell'accuratezza. Infine, vengono esposte le prestazioni dell'algoritmo di creazione delle Heatmap sulle WSI di test.

3.1 Validazione delle reti

Le performances delle reti nel task di classificazione del tessuto polmonare viene valutata sia sulle patch di training che su quelle di test. Per ogni architettura vengono prima presentate le curve che descrivono l'andamento dell'allenamento ed in seguito i risultati espressi secondo le metriche di validazione patch-based: confusion matrix, curve ROC, t-SNE e saliency map. Infine, vengono mostrati i risultati dell'algoritmo di ricerca dei pesi ottimali e le metriche di validazione con la combinazione dei tre modelli mediante media pesata.

3.1.1 Risultati degli allenamenti delle CNN

Allenamento VGG

In Figura 3.1 viene mostrato l'andamento dell'apprendimento in funzione del numero di epoche. L'apprendimento viene arrestato per *Earlystopping* all'epoca 22 dal momento che per 15 epoche successive non si è verificato un aumento dell'*accuratezza* sul validation set superiore a 0.01. Dalla curva in Figura 3.1a si può notare che per la VGG la loss sul training parte da un valore molto pari 0.7. Al termine dell'allenamento viene raggiunta un'*accuratezza* sul training pari

a circa l'89% ed una loss pari a 0.29. L'andamento delle curve sul validation si presenta invece più oscillante con un'accuratezza pari a circa 82.5% e una loss di 0.32 per l'epoca 22. Non si riscontra un bias elevato ma è visibile un'accuratezza inferiore sul validation set a causa del quale il training viene arrestato per evitare che si verifichi una possibile situazione di overfitting.

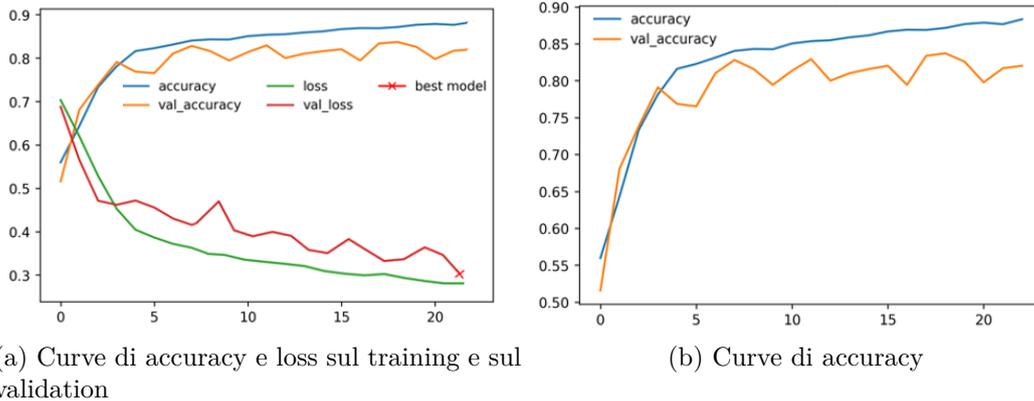


Figura 3.1: Curve di training VGG

Allenamento ResNet50

In Figura 3.2 viene mostrato l'andamento dell'apprendimento in funzione del numero di epoche. L'allenamento della ResNet50 segue la falsariga di quello della rete VGG. Anche in questo caso l'allenamento viene anticipatamente terminato per *Earlystopping* all'epoca 37. La loss per il validation e il training parte da un valore leggermente inferiore a 0.7. Al termine dell'allenamento viene raggiunta un'accuratezza sul training pari a 85.5% e una loss pari a 0.32 in corrispondenza dell'epoca in cui si verifica l'*Earlystopping*. Sul validation set le metriche sono leggermente peggiori con un'accuratezza pari a 82.5% e una loss pari a 0.355. Per la ResNet50 si presenta un piccolo drop tra le curve di loss del validation e del training set.

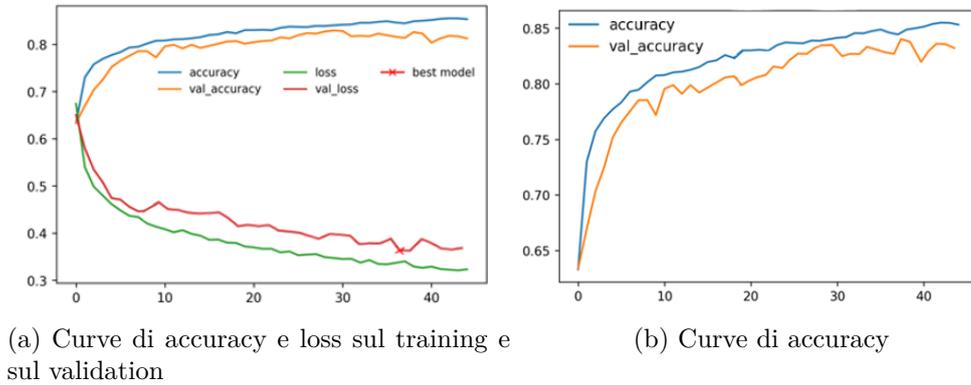


Figura 3.2: Curve di training ResNet50

Allenamento DenseNet

In Figura 3.3 viene mostrato l'andamento dell'apprendimento in funzione del numero di epoche. L'allenamento termina per *Earlystopping* all'epoca 34. Anche in questo caso la loss parte da 0.7 per entrambi i dataset. Al termine dell'allenamento viene raggiunta un'accuratezza sul training pari a 88.7% e una loss pari a 0.297 in corrispondenza dell'epoca in cui si verifica l'*Earlystopping*. Sul validation set si può osservare un andamento oscillante delle curve con accuratezza pari a circa l'85% e una loss di 0.31. Per la DenseNet si presenta un leggero drop tra le curve di accuratezza del validation e del training set.

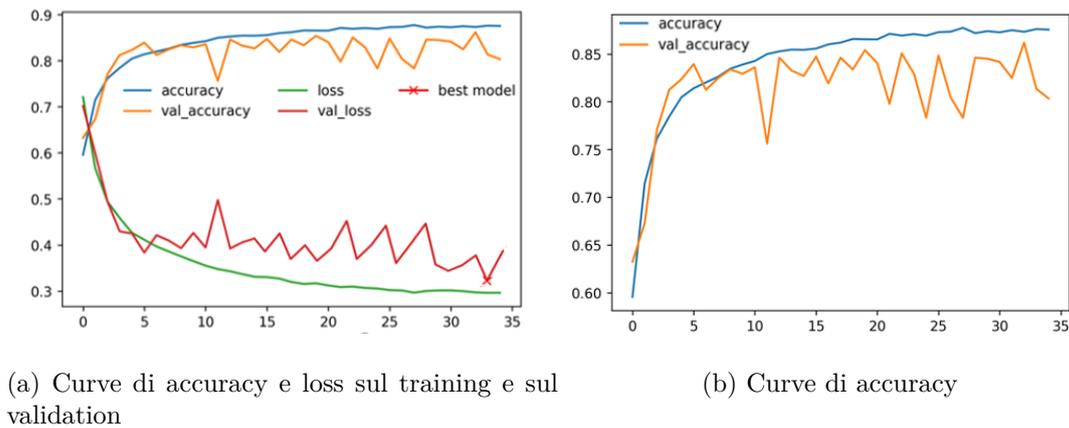


Figura 3.3: Curve di training DenseNet

3.1.2 Risultati sul Training set

In questo paragrafo vengono mostrati i risultati per ogni rete sulle patch di training espressi in funzione delle metriche descritte nel paragrafo 2.6.1.

VGG

In Figura 3.4 vengono mostrate le confusion matrix sulle patch di training e di validation considerando 0.5 come soglia per distinguere tra patch tumorale e benigna.

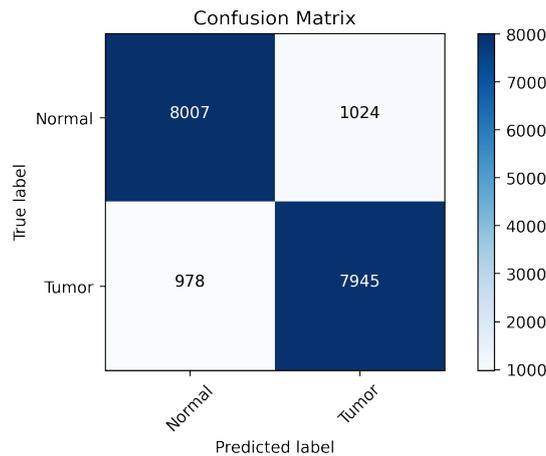


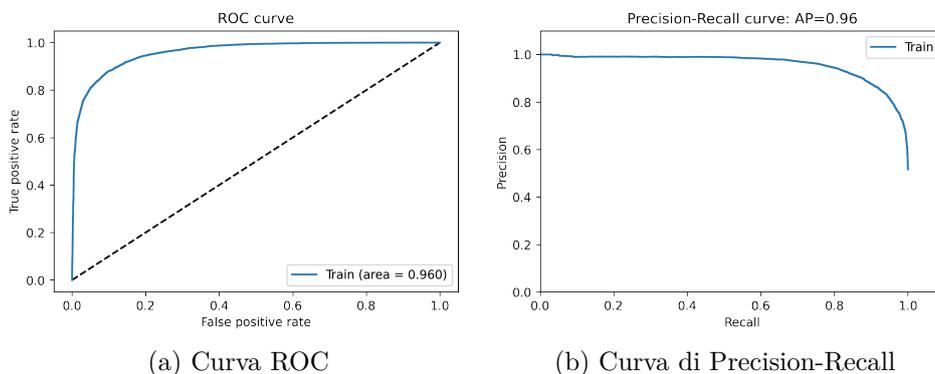
Figura 3.4: Confusion matrix sul training con rete VGG

A partire dalla tabella di contingenza sopra raffigurata vengono calcolati gli indicatori utilizzati per quantificare le prestazioni della VGG nel task di classificazione sul training set (Tabella 3.1).

Tabella 3.1: Performance VGG sul training set

Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
89%	89%	89%	88.5%	89.1%	88.7%

Osservando la tabella, è possibile apprezzare le buone performance della VGG sulle patch di training. La rete, inoltre, predice ugualmente bene le patch maligne e benigne del dataset. Unitamente alle metriche di classificazione viene riportata la curva ROC in Figura 3.5a e la curva di precision/recall in Figura 3.5b per poter valutare il potere discriminante della VGG allenata. La VGG raggiunge un AUC (*area under curve*) di 0.96 e un AP (*average precision*) di 0.96 dunque costituisce un classificatore altamente accurato nel task di classificazione.



(a) Curva ROC

(b) Curva di Precision-Recall

Figura 3.5: Curve ROC e PR della VGG

In Figura 3.6 viene inoltre mostrata la distribuzione t-SNE. Come si può osservare le patch costituiscono due cluster distinti ma non separati, sovrapponendosi per alcuni esempi del training. Ciò è dovuta all'elevata difficoltà nel riconoscimento e distinzione del tumore nelle immagini.

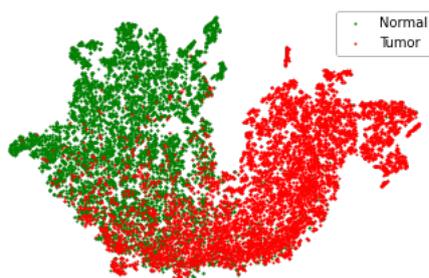


Figura 3.6: Distribuzione t-SNE del training set

Infine vengono presentate le saliency map sulle patch benigne e maligne in Figura 3.7. Si può notare come per la VGG venga sempre preso in considerazione il tessuto per la predizione della classe ed in particolare per la classe benigna la rete sembra osservare maggiormente le strutture rosa del tessuto, mentre per la classe maligna vengono maggiormente considerati i nuclei.

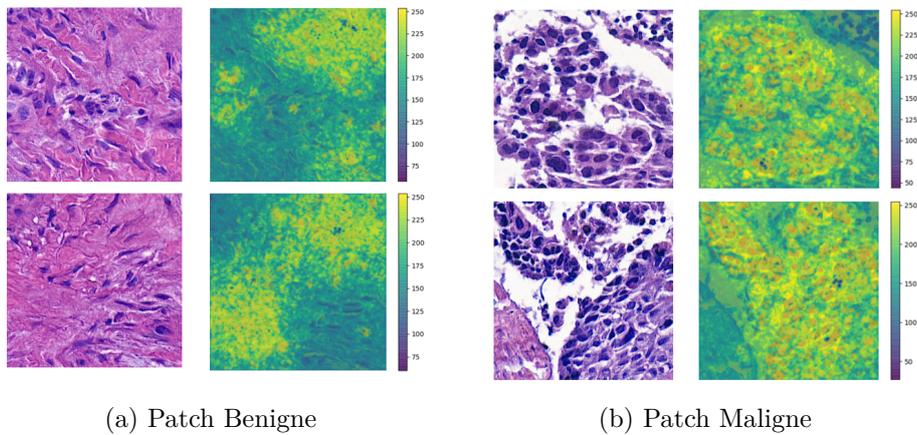


Figura 3.7: Saliency map con rete VGG

ResNet50

Seguono i risultati dell'allenamento della ResNet50. In Figura 3.8 vengono mostrate le confusion matrix sulle patch di training e di validation utilizzando per la predizione la ResNet allenata e considerando 0.5 come soglia per distinguere tra patch tumorale e benigna.

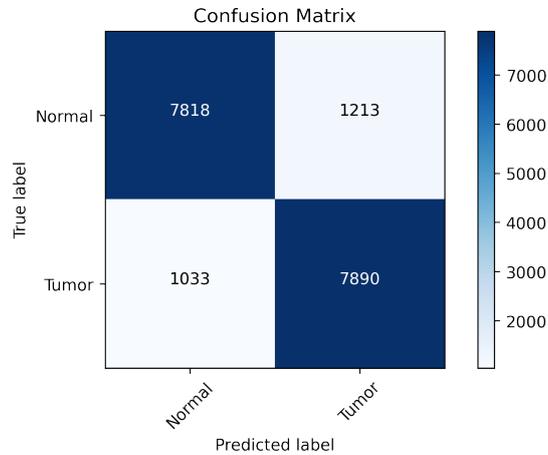


Figura 3.8: Confusion matrix sul training con rete ResNet50

A partire dalla confusion matrix sopra raffigurata vengono calcolati gli indicatori utilizzati per quantificare le prestazioni della rete nel task di classificazione sul training set (Tabella 3.2).

Osservando la tabella, è possibile apprezzare le buone performance della rete sulle patch di training. La rete, inoltre, predice ugualmente bene le patch maligne e benigne del dataset seppur leggermente peggio rispetto alla VGG. In effetti, tutte

Tabella 3.2: Performance ResNet sul training set

Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
87%	88%	87%	87%	88.3%	87.5%

le metriche mostrano valori leggermente inferiori (circa meno dell'1%) rispetto alla rete precedente. Unitamente alle metriche di classificazione viene riportata la curva ROC in Figura 3.9a e la curva di precision/recall in Figura 3.9b per poter valutare il potere discriminante della ResNet allenata.

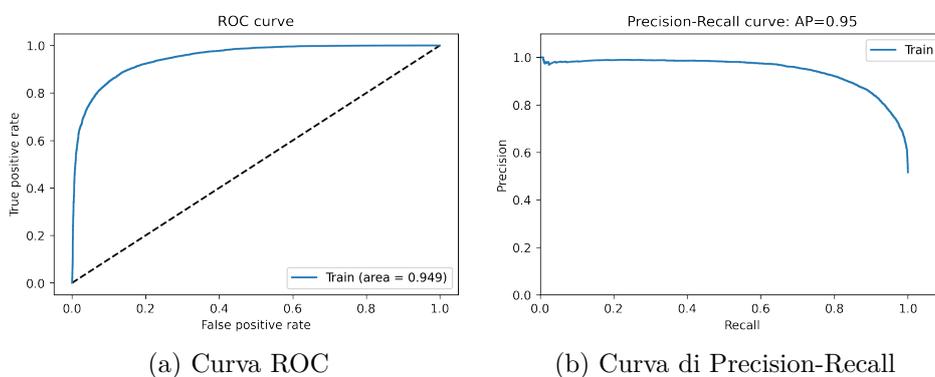


Figura 3.9: Curve ROC e PR della ResNet

Considerando anche le due curve in Figura 3.9 la ResNet sembra avere prestazioni leggermente inferiori. Tuttavia, con un AUC pari a 0.949 e una AP (*average precision*) di 0.95 la rete si dimostra comunque un classificatore molto accurato. In Figura 3.10 viene inoltre mostrata la distribuzione t-SNE. In questo caso, è possibile osservare la creazione di due cluster meno distinti rispetto alla rete precedente con un numero superiore di esempi misclassificati.

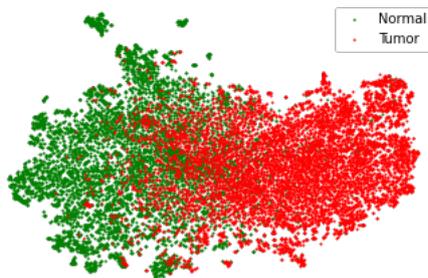


Figura 3.10: Distribuzione t-SNE del training set

Infine vengono presentate le saliency map sulle patch benigne e maligne in Figura 3.11. Si può notare come anche per la ResNet venga sempre preso in considerazione il tessuto per la predizione della classe. A differenza delle VGG in questo caso la rete non sembra prediligere delle strutture specifiche nell'immagine ma in generale viene considerata tutta la componente tissutale del vetrino.

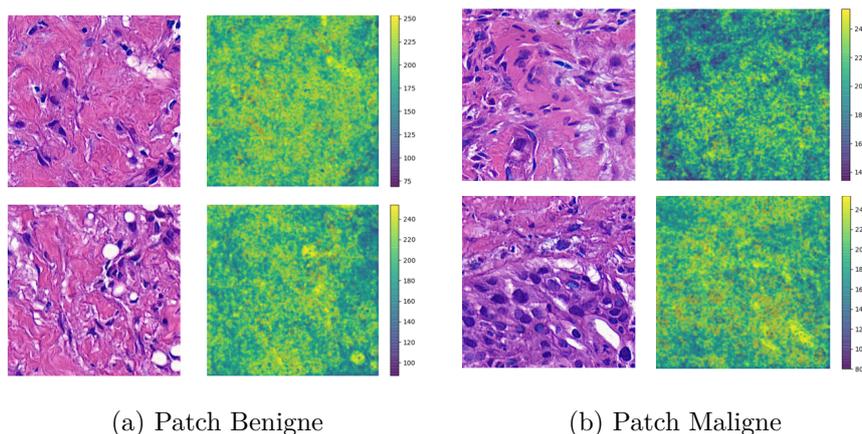


Figura 3.11: Saliency map con rete ResNet

DenseNet

In aggiunta ai precedenti, vengono presentati i risultati con la DenseNet. In Figura 3.12 vengono mostrate le confusion matrix sulle patch di training e di validation utilizzando per la predizione la DenseNet allenata e considerando 0.5 come soglia per distinguere tra patch tumorale e benigna.

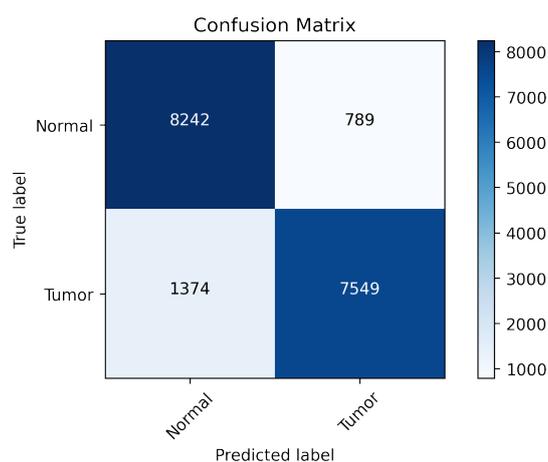


Figura 3.12: Confusion matrix sul training con rete DenseNet

A partire dalla confusion matrix sopra raffigurata vengono calcolati gli indicatori utilizzati per quantificare le prestazioni della rete nel task di classificazione sul training set (Tabella 3.3).

Tabella 3.3: Performance DenseNet sul training set

Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
88%	85%	91%	90.5%	86%	88%

Osservando la tabella, è possibile apprezzare le buone performance della rete sulle patch di training in particolare su quelle benigne. La DenseNet mostra dei valori di *Specificità* e di *Precisione* (o valore predittivo positivo) molto elevati a dimostrazione della buone performance della rete sul tessuto sano delle immagini di biopsia. Tuttavia, tra tutte è la rete meno precisa nel riconoscimento del tumore nelle patch maligne con una *Recall* minore di entrambi i modelli precedenti.

Unitamente alle metriche di classificazione viene riportata la curva ROC in Figura 3.13a e la curva di precision/recall in Figura 3.13b per poter valutare il potere discriminante della DenseNet allenata.

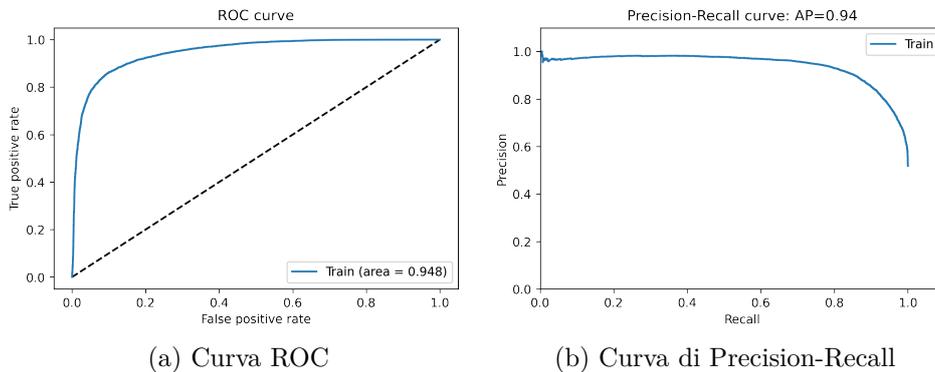


Figura 3.13: Curve ROC e PR della DenseNet

Considerando anche le due curve in Figura 3.13 la rete sembra avere prestazioni leggermente inferiori rispetto alle precedenti. Tuttavia, con un AUC pari a 0.948 e una AP (*average precision*) di 0.94, la rete rappresenta comunque un modello molto accurato.

In Figura 3.14 viene inoltre mostrata la distribuzione t-SNE. In questo caso, è possibile osservare la creazione di due cluster ben distinti rispetto alla ResNet ma con molti elementi della classe tumorale erroneamente classificati come benigni.

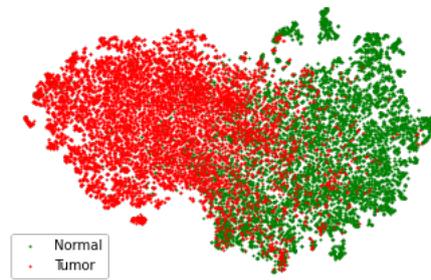


Figura 3.14: Distribuzione t-SNE del training set

Per finire, vengono presentate le saliency map sulle patch benigne e maligne in Figura 3.15. La DenseNet sembra prediligere per la classificazione corretta della classe benigna le porzioni colorate in rosa dell'immagine di input. Mentre per la classe maligna vengono maggiormente considerati i nuclei (strutture di colore blu nell'immagine), in particolare quelli di grandi dimensioni.

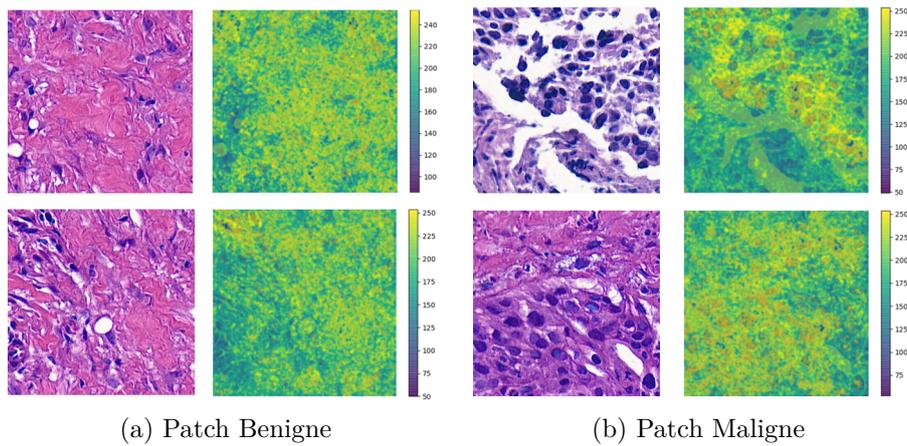


Figura 3.15: Saliency map con rete DenseNet

3.1.3 Risultati sul Test set

In questa sezione vengono presentati i risultati degli allenamenti sulle patch del test set, sconosciute alle 3 reti nella fase di training, per verificare la capacità di generalizzazione dei modelli proposti.

VGG

In Figura 3.16 vengono mostrate le confusion matrix sulle patch del test set considerando sempre 0.5 come soglia per distinguere tra patch tumorale e benigna. Le confusion matrix si ottengono confrontando l'output della rete con la classe reale di appartenenza della patch. E' necessario notare che il numero di patch maligne nel test set sono di gran lunga superiori (circa 1100 in più) rispetto a quelle benigne.

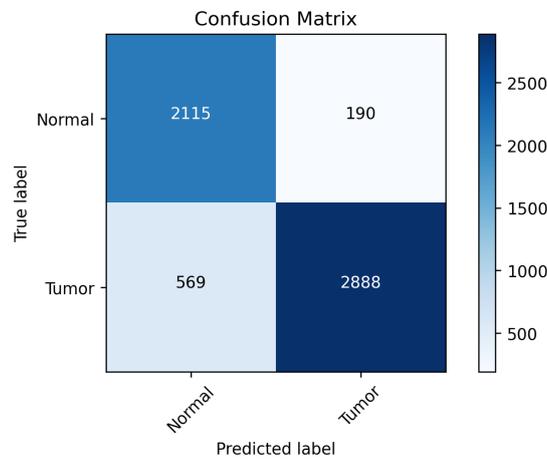


Figura 3.16: Confusion matrix sul test set con rete VGG

Nella Tabella 3.4 seguente vengono presentati gli indicatori utilizzati per la quantificazione numerica delle prestazioni della VGG nel task di classificazione sul test set.

Tabella 3.4: Performance VGG sul test set

Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
87%	84%	92%	94%	79%	89%

Nel task di classificazione delle patch di test set la VGG si rileva molto accurata e precisa. I valori di *Specificità* e *Precisione* sono molto alti. Questo dimostra una buona capacità della rete nella predizione della classe benigna o sana. Per quanto riguarda invece la predizione della classe maligna, le performance sono leggermente minori con un valore di *Recall* pari all'84%.

Unitamente alle metriche di classificazione vengono riportate la curva ROC in Figura 3.17a e la curva di precision/recall in Figura 3.17b per poter valutare il potere discriminante della VGG allenata.

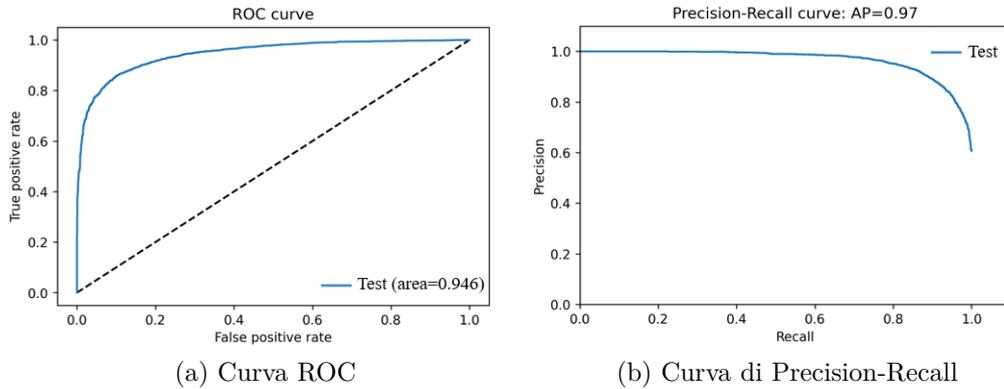


Figura 3.17: Curve ROC e PR della VGG

La VGG raggiunge un AUC (*area under curve*) di 0.946 e un AP (*average precision*) di 0.97 sul test. Il valore di AUC risulta leggermente inferiore rispetto a quello sul training mentre l'AP è superiore. La rete quindi dimostra di essere un buon classificatore sia sul training sia sul test con buone capacità di generalizzazione. In Figura 3.18 viene inoltre mostrata la distribuzione t-SNE. Come si può osservare le patch costituiscono due cluster distinti e ben separati. Sono visibili più misclassificazioni appartenenti alla classe maligna che si mischiano agli esempi della classe opposta in accordo alle metriche precedentemente presentate.

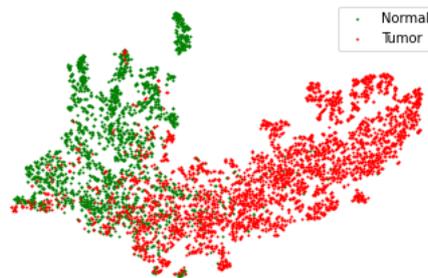


Figura 3.18: Distribuzione t-SNE del test set

ResNet50

In Figura 3.19 vengono mostrate le confusion matrix sulle patch di test set ottenute con la rete ResNet50 allenata.

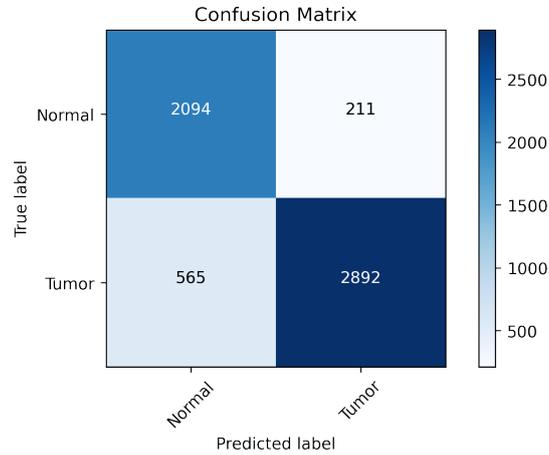


Figura 3.19: Confusion matrix sul test set con rete ResNet

A partire dalla confusion matrix, nella Tabella 3.5 seguente vengono presentati gli indicatori utilizzati per la quantificazione numerica delle prestazioni della ResNet per la classificazione del test set.

Tabella 3.5: Performance ResNet sul test set

Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
87%	84%	91%	93.2%	79%	88.4%

Nella classificazione delle patch del test set la ResNet si rileva molto accurata e precisa come la VGG precedente con valori di *Specificità* e *Precisione* leggermente più bassi. La rete è molto simile alla VGG come prestazioni. Per entrambe, osservando gli indicatori di validazione, non viene riscontrata presenza di overfitting dimostrandosi molto robuste nella generalizzazione.

Oltre alle metriche di classificazione vengono riportate la curva ROC in Figura 3.20a e la curva di precision/recall in Figura 3.20b per il test set.

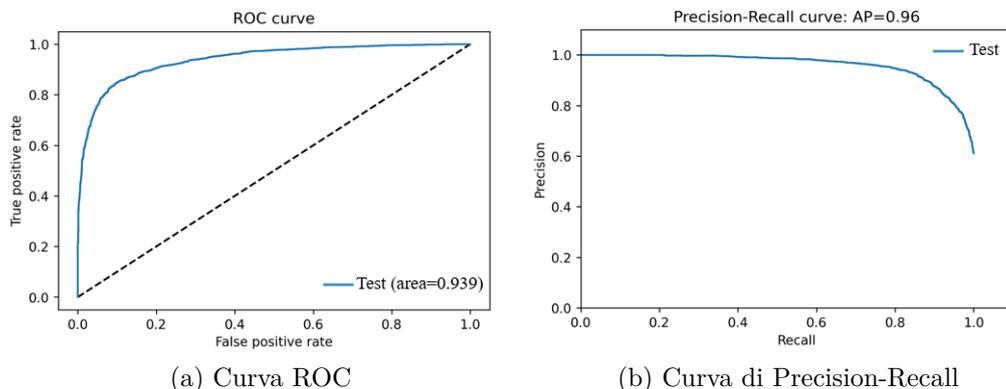


Figura 3.20: Curve ROC e PR della VGG

La ResNet raggiunge un AUC (*area under curve*) di 0.939 e un AP (*average precision*) di 0.96 sul test. Come avveniva per il training anche la sul test set, la rete mostra valori leggermente inferiori rispetto alla VGG. Nonostante ciò anche questa rete è un buon classificatore in termini di generalizzazione.

In Figura 3.21 viene inoltre mostrata la distribuzione t-SNE. La rete crea due nuvole di punti distinguibili con una leggera sovrapposizione nella porzione centrale. I punti verdi (classe normale) sembrano occupare un minore spazio rispetto a quelli rossi, segno di una maggiore capacità di discriminazione degli esempi benigni che, a parte pochi outliers, sono tutti concentrati e vicini nel cluster di appartenenza.

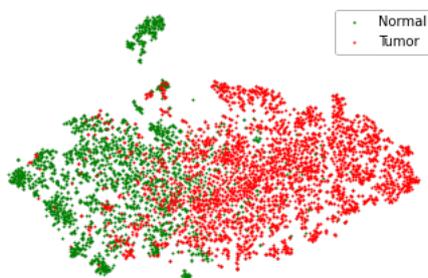


Figura 3.21: Distribuzione t-SNE del test set

DenseNet

In Figura 3.22 vengono mostrate le confusion matrix ottenute confrontando l'output della rete con la classe reale di appartenenza della patch del test set.

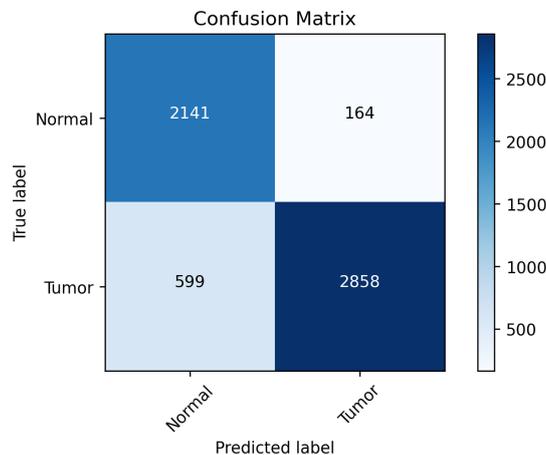


Figura 3.22: Confusion matrix sul test set con rete DenseNet

Nella Tabella 3.6 seguente sono riportati i valori degli indicatori utilizzati per la validazione della DenseNet nella classificazione del test set .

Tabella 3.6: Performance DenseNet sul test set

Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
87%	83%	93%	95%	78%	88.6%

Nel task di classificazione delle patch banigne appartenenti al test set la DenseNet risulta essere la rete più precisa tra tutte con un valore di *Specificità* e di *VP⁺* rispettivamente del 93% e del 95%. La predizione della classe tumorale si rileva sempre essere il task più complicato per tutte le reti ed in particolare per la DenseNet che presenta il maggior numero di falsi negativi. Ciò nonostante, la rete mantiene comunque dei valori di *accuratezza* e *F₁* elevati, di poco inferiori rispetto al training set manifestando una buona capacità di predizione di esempi nuovi e sconosciuti.

Vengono poi riportate la curva ROC in Figura 3.23a e la curva di precision/recall in Figura 3.23b per poter valutare il potere discriminante della DenseNet allenata.

La DenseNet raggiunge un AUC (*area under curve*) di 0.949 e un AP (*average precision*) di 0.97 sul test. Il valore di AUC è superiore a tutte le altre reti mentre di AP è superiore a quello della ResNet e uguale a quella della VGG. Inoltre, la DenseNet è l'unica tra le reti a presentare i valori di AUC e di AP sul test superiori

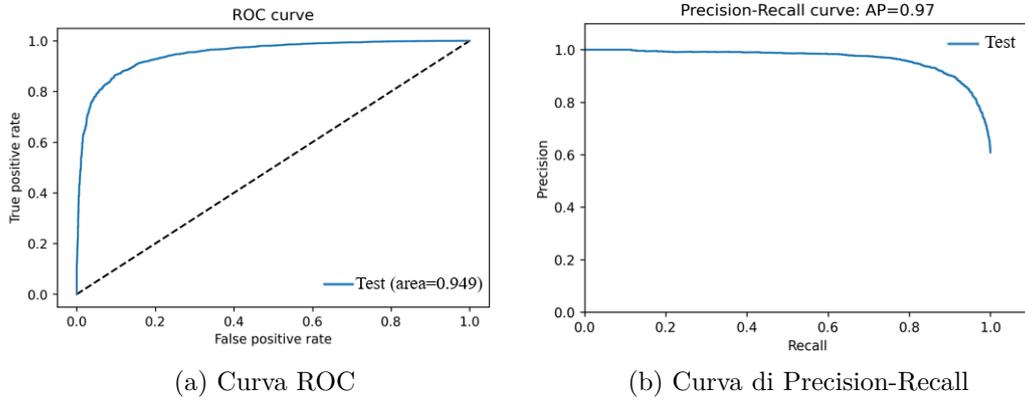


Figura 3.23: Curve ROC e PR della ResNet

a quelli sul training set. La rete è quindi assolutamente in grado di riconoscere nuove immagini senza perdere in prestazioni.

In Figura 3.24 viene inoltre mostrata la distribuzione t-SNE. Anche per la Dense net sono presenti più misclassificazioni appartenenti alla classe maligna che rientrano nel cluster verde o benigno e pochissimi esempi benigni misclassificati. Il cluster della classe benigna si mostra anche per questa rete molto più compatto nello spazio.

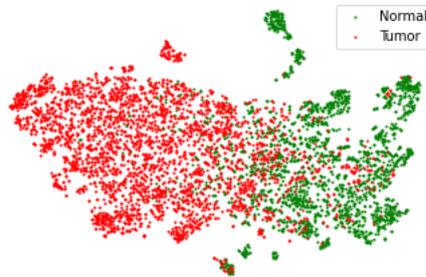


Figura 3.24: Distribuzione t-SNE del test set

3.1.4 Risultati con fusione mediante media pesata

In ultimo vengono presentati i risultati sul training set e sul test set utilizzando il metodo di fusione mediante media pesata sugli output delle reti. In primo luogo, come descritto nel paragrafo 2.4.4, vengono ricavate le predizioni singolarmente per ogni rete sulle patch di training e poi vengono ricercati i pesi che consentono di massimizzare l'*accuratezza* totale. Dalla ricerca a griglia i pesi ottenuti per le varie reti sono:

Tabella 3.7: Pesi delle reti sull'output finale

ResNet	VGG	DenseNet
0.324	0.380	0.296

Come si può osservare dalla Tabella 3.7, la rete con il peso maggiore per l'*accuratezza* finale è la VGG. Questa CNN, infatti, risultava essere la migliore sul training set e per questa ragione costituisce il modello più affidabile tra i tre proposti. Segue la ResNet50 che come mostrato precedentemente, presentava dei risultati leggermente inferiori alla VGG. In ultimo, in termini di contributo all'output finale, la DenseNet che costituiva il modello migliore per la predizione delle classe benigna ma il peggiore per la tumorale. Le tre reti di fatto non partecipano all'output finale in egual misura e questo contribuisce all'incremento dell'*accuratezza* sul modello di fusione finale.

In Figura 3.25 vengono quindi mostrate le confusion matrix ottenute confrontando l'output della rete con la classe reale di appartenenza della patch del training (Figura 3.25a) e di test (Figura 3.25b).

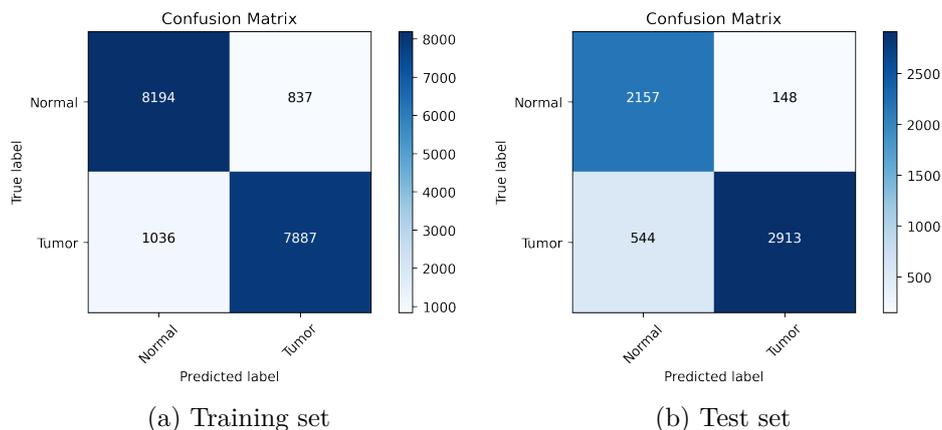


Figura 3.25: Confusion matrix con modello di insieme

A prima vista, è già evidente come i risultati siano migliorati rispetto ai modelli singoli. In particolare questo è ancora più evidente sul test set dove si hanno dei miglioramenti netti rispetto a tutti i modelli. Per una migliore comprensione dei risultati, di seguito viene riportata la Tabella 3.8 con gli indicatori utilizzati per la validazione dei modelli.

Tabella 3.8: Performance modello di insieme sul training set e sul test set

	Accuratezza	Recall	Specificità	VP ⁺	VP ⁻	F ₁
Training set	90%	88.4%	91%	90.4%	89%	89.4%
Test set	88%	84.3%	93.6%	95.2%	80%	89.4%

E' evidente come i risultati su i due dataset siano molto positivi. In particolare, considerando gli indicatori sul training set, si può monitorare un incremento rispetto a tutti i modelli singoli dell'*accuratezza* e del F_1 score. Questi valori (prossimi al 90%) mettono in luce le capacità del modello di insieme di predire ugualmente bene le due classi del training set. Per quanto concerne invece il valore di *recall* e di *VP* solo la VGG consegue dei valori lievemente superiori. In termini di *Specificità* il modello di insieme eguaglia il migliore dei tre ovvero la DenseNet e presenta un valore di VP^+ di 0.1% minore di quest'ultima. Tutti gli indicatori presenti superano i risultati conseguiti dalla ResNet50.

Analizzando, invece, le performance sul test set è possibile osservare un incremento di tutte le metriche di validazione rispetto ai modelli singoli. Nello specifico, il valore di *accuratezza* acquista un 1% rispetto a tutte le tre reti. La *recall* viene incrementata di 0.3% rispetto alle due reti migliori ovvero la VGG e la ResNet. La *Specificità* e il VP^+ rispetto ai valori migliori delle tre (quelli della DenseNet) aumentano rispettivamente di 0.6% e 0.2%. Il VP^- aumenta dell'1% rispetto a quello della ResNet e VGG (valori migliori di VP^-). Infine anche l' F_1 score aumenta di 0.4% rispetto al migliore delle tre (quello della VGG).

In Figura 3.26 vengono invece riportate le curve di precision-recall (PR) per il training set (Figura 3.26a) e il test set (Figura 3.26b). Osservando le curve di PR e confrontandole tra loro, si può notare come il valore di *average precision* del test set (0.97) superi quello del training set (0.96). Inoltre, confrontandoli con i valori ottenuti per i singoli modelli, gli indicatori eguagliano i modelli migliori (per il training la VGG e per il test la VGG e la DenseNet).

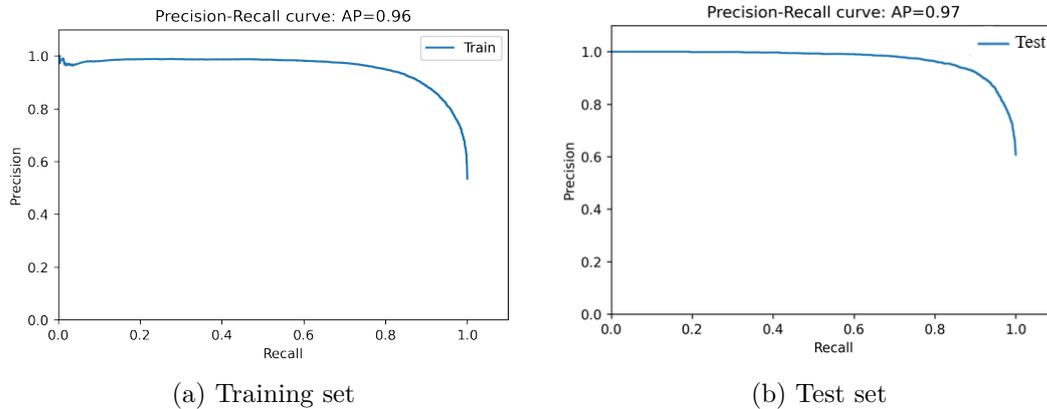


Figura 3.26: Curve PR per il modello di insieme

Infine, in Figura 3.27 vengono mostrate le curve ROC sul training (Figura 3.27a) e sul test set (Figura 3.27b).

In questo caso il valore AUC sul training supera di poco quello del test set (differenza di circa 0.004). Questo comportamento non si manifesta per i singoli modelli, per i quali si riscontra un maggior distacco tra i valori di AUC nei due dataset. Nello specifico, analizzando le differenze con i singoli modelli si nota un aumento lieve dell' AUC sul training set rispetto al miglior modello (VGG). Tuttavia sul test set, in cui il miglior modello in termini di AUC era la DenseNet (con 0.949), si ha un netto aumento delle prestazioni (con AUC pari a 0.957). Il modello finale di fusione si presenta quindi come un ottimo classificatore nei due dataset e con capacità di generalizzazione molto superiori ai tre modelli presi singolarmente.

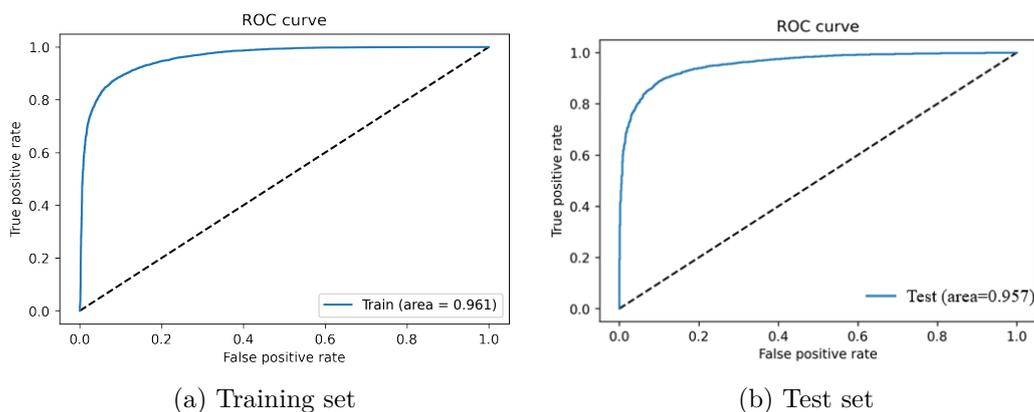


Figura 3.27: Curve ROC per il modello di insieme

3.2 Validazione delle Heatmap

In questa sezione vengono presentati i risultati sulle 30 WSI che costituiscono il test. Più precisamente, dopo aver ottenuto le tre reti neurali per la classificazione, si procede con la creazione delle heatmap come espresso nel paragrafo 2.4.3. Una volta ottenute le softmax con l'output delle tre reti segue la segmentazione con global thresholding (soglia=0.5) al fine di ottenere una maschera. Le maschere binarie delle heatmap, infine, vengono confrontate con le annotazioni (o *ground truth*) realizzate da un patologo esperto. La validazione delle heatmap avviene a livello del singolo pixel dell'immagine con le metriche già descritte nel paragrafo 2.6.2. Nello specifico verranno mostrati prima i risultati considerando i modelli singoli: VGG, ResNet50 e DenseNet; e poi quelli ottenuti con il modello combinato. Più nel dettaglio, per ogni architettura gli indicatori verranno calcolati sia considerando unicamente il tessuto sia conteggiando il bianco del vetrino nel computo dei veri negativi e falsi negativi. Inoltre, per ciascuno dei due casi verranno esposti i risultati nelle seguenti condizioni:

1. *Only thresholding*: in cui viene unicamente eseguita l'operazione di segmentazione con soglia globale.
2. *Label refine*: in cui oltre all'operazione di global thresholding vengono smussati i contorni dell'annotazione manuale e di quella automatica moltiplicando entrambe per la maschera contenente unicamente i pixel del tessuto (complementare della maschera dei bianchi).

L'operazione di label refine viene eseguita anche sulla *ground truth* dal momento che le annotazioni appaiono molto imprecise comprendendo grandi porzioni del vetrino in bianco. Ciò chiaramente comporta una riduzione delle metriche di validazione qualora l'algoritmo si rilevi più preciso nei contorni rispetto all'annotazione manuale.

VGG

In primo luogo vengono analizzati i risultati considerando unicamente i pixel del tessuto senza conteggiare i bianchi del vetrino. In Figura 3.28 vengono mostrati i boxplot delle metriche di validazione pixel-based con la VGG. In particolare, vengono presentate le metriche nelle due condizioni di only thresholding (Figura 3.28a) e con label refine (ridefinizione della maschera manuale e automatica) in Figura 3.28b. Analizzando i boxplot è evidente una distribuzione per tutte le metriche che tende a dei valori elevati sia con la sola operazione di global thresholding sia considerando la label refine.

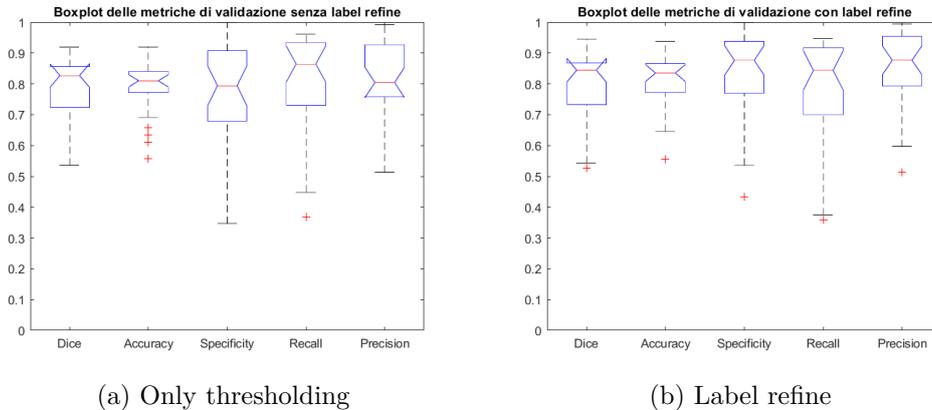


Figura 3.28: Metriche di validazione Heatmap con VGG considerando solo il tessuto

Con la sola segmentazione il valore migliore risulta essere quello della *Recall* a causa di un numero relativamente basso di falsi negativi. Mentre il valore peggiore è quello di *Specificità*, probabilmente dovuto al basso numero di veri negativi quindi di pixel neri correttamente classificati. Includendo l'operazione di refining sulle annotazioni, si nota un incremento di tutti gli indicatori eccetto la *recall* a causa di una riduzione dei veri positivi. Infatti, la moltiplicazione delle annotazioni con la maschera del tessuto elimina un gran numero di pixel appartenenti al vetrino e inclusi erroneamente nell'annotazione tumorale. In compenso si ha un netto miglioramento della metrica di *Precisione* per la riduzione del numero dei falsi positivi.

Seguono i grafici ottenuti prendendo in considerazione anche i bianchi del vetrino e non solo il tessuto nel conteggio dei veri e falsi negativi (Figura 3.29).

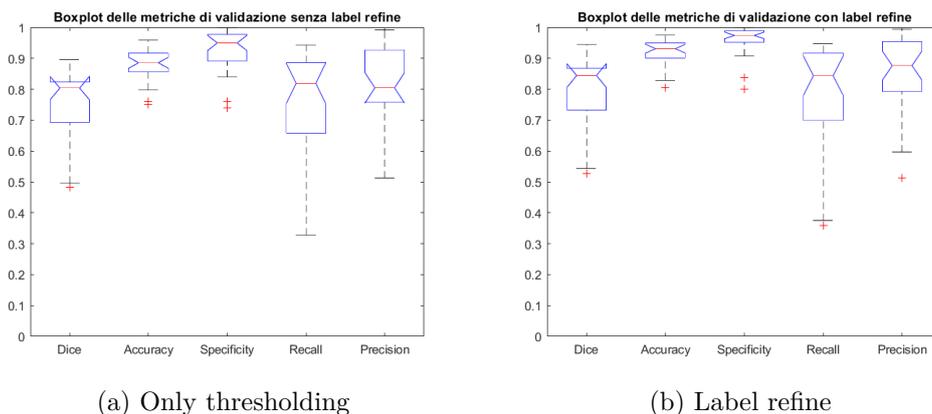


Figura 3.29: Metriche di validazione Heatmap con VGG con i pixel del vetrino

Dai grafici precedenti si nota un deciso incremento delle metriche di *accuratezza* e *specificità* per entrambe le condizioni in Figura 3.29a e 3.29b dovuto ad un netto aumento dei veri negativi appartenenti al vetrino.

In sintesi in Tabella 3.9 sono elencati i valori medi e le deviazioni standard per ogni metrica in tutte le situazioni descritte.

Tabella 3.9: Metriche Heatmap VGG

	Dice	Accuratezza	Specificità	Recall	Precisione
Only thresholding (solo tessuto)	0.789 \pm 0.098	0.789 \pm 0.086	0.76 \pm 0.174	0.8112 \pm 0.16	0.808 \pm 0.133
Label refine (solo tessuto)	0.794 \pm 0.111	0.801 \pm 0.093	0.840 \pm 0.140	0.783 \pm 0.174	0.852 \pm 0.125
Only thresholding (con vetrino)	0.754 \pm 0.109	0.88 \pm 0.053	0.93 \pm 0.07	0.752 \pm 0.176	0.809 \pm 0.133
Label refine (con vetrino)	0.794 \pm 0.111	0.923 \pm 0.041	0.960 \pm 0.046	0.783 \pm 0.174	0.852 \pm 0.125

Come è possibile constatare dalla Tabella 3.9 il *Dice* non mostra miglioramenti rispetto al caso in cui viene analizzato il solo tessuto; anzi con la sola operazione di thresholding peggiora ulteriormente così come la *recall*. L'*accuratezza* e la *Specificità* invece migliorano significativamente. L'operazione di label refine sembra dare un riscontro positivo in ogni caso eccetto per la *recall* con il solo tessuto.

ResNet50

Analogamente alla VGG, in figure 3.30 vengono riportati i boxplot delle metriche di validazione considerando la sola operazione di thresholding sulla maschera binaria senza postprocessing e con postprocessing (o label refine). Anche per questa rete di classificazione tutte le metriche mostrano dei risultati molto soddisfacenti in entrambe le condizioni. Il *dice score* risente molto poco dell'effetto di post processing a differenza di tutti gli altri indicatori. Nello specifico l'operazione di label refining sembra migliorare di molto l'*accuratezza*, la *specificità* e la *precisione* della rete nella segmentazione del tumore. Questi incrementi sono da addebitare ad una riduzione dei falsi positivi. In effetti mediante la moltiplicazione delle annotazioni con la maschera del tessuto molti pixel del contorno che includevano porzioni del vetrino non di interesse vengono eliminati e classificati come negativi, incrementando il numero dei veri negativi e riducendo i falsi positivi. A risentire dell'operazione di refining è invece la metrica di *recall* per la quale risulta ridotto il numero di veri positivi rispetto alla sola segmentazione. In questo caso, eliminando per entrambe

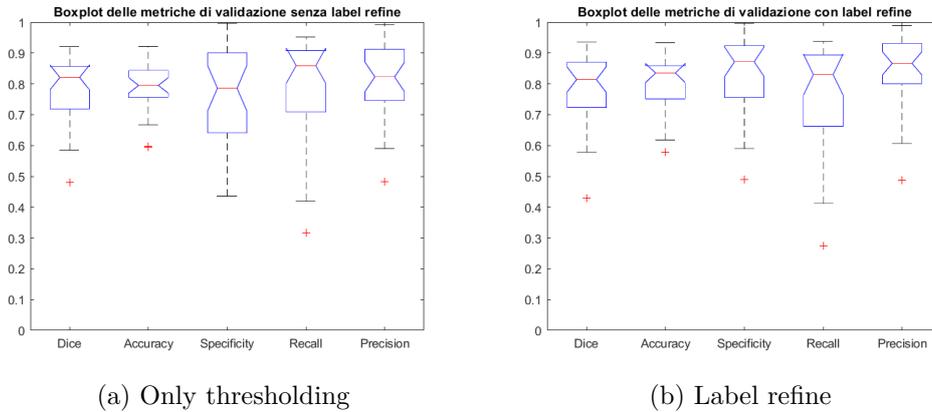


Figura 3.30: Metriche di validazione Heatmap con ResNet50 considerando solo il tessuto

le maschere (reale e automatica) dei pixel del vetrino classificati come tumorali sia dal patologo che dall'algoritmo, il numero dei veri positivi si riduce di molto. Seguono i boxplot delle metriche prendendo in considerazione anche i pixel bianchi del vetrino e non solo il tessuto nel calcolo dei veri e falsi negativi (Figura 3.31).

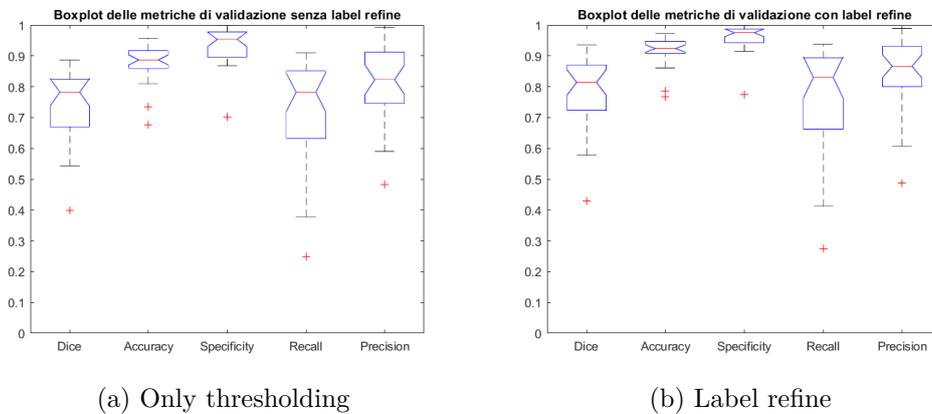


Figura 3.31: Metriche di validazione Heatmap con ResNet50 con i pixel del vetrino

La considerazione dei pixel chiari del vetrino come veri negativi risulta di fatto in un aumento delle metriche di *accuratezza* e di *specificità*. Nel caso di *only thresholding*, si nota invece una riduzione della *recall* e conseguentemente del *dice*. Per il *dice*, la *recall* e la *precisione* con operazione di post processing le metriche risultano invece invariate pur considerando anche il vetrino.

Per sintetizzare in Tabella 3.10 vengono riportati i valori medi e le deviazioni standard per ogni metrica in tutte le situazioni descritte.

Tabella 3.10: Metriche Heatmap ResNet50

	Dice	Accuratezza	Specificità	Recall	Precisione
Only thresholding (solo tessuto)	0.781 ± 0.108	0.788 ± 0.08	0.763 ± 0.16	0.789 ± 0.161	0.811 ± 0.127
Label refine (solo tessuto)	0.782 ± 0.117	0.803 ± 0.086	0.833 ± 0.128	0.762 ± 0.168	0.846 ± 0.121
Only thresholding (con vetrino)	0.743 ± 0.116	0.877 ± 0.06	0.934 ± 0.06	0.723 ± 0.168	0.811 ± 0.121
Label refine (con vetrino)	0.782 ± 0.117	0.919 ± 0.05	0.960 ± 0.044	0.762 ± 0.168	0.846 ± 0.125

DenseNet

In Figura 3.32 sono quindi riportati i risultati con la rete DenseNet per la validazione dell’algoritmo di creazione delle heatmap.

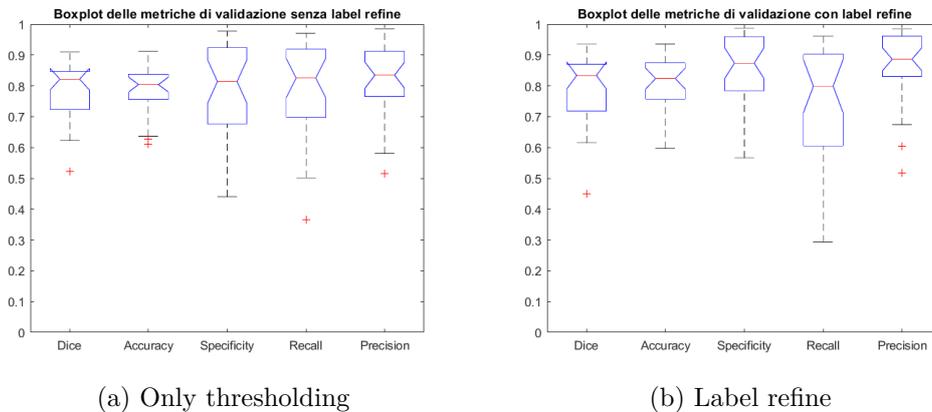


Figura 3.32: Metriche di validazione Heatmap con DenseNet considerando solo il tessuto

Anche per la DenseNet valgono le stesse considerazioni, con valori delle metriche molto buoni. La rete si dimostra essere la migliore in termini di *accuratezza*, *specificità* e *precisione* a causa del suo numero ridotto di falsi positivi. Mentre come *Dice* e *recall* presenta dei risultati intermedi alle due reti precedenti. Per la DenseNet così come per le altre reti l’operazione di post processing migliora di molto le metriche di *specificità* e *precisione*, riducendo però la *recall*. Si può dunque concludere che il refining eserciti di fatto un contributo positivo sulla segmentazione finale.

Seguono i boxplot delle metriche considerando anche i pixel bianchi del vetrino del computo dei veri negativi (Figura 3.33).

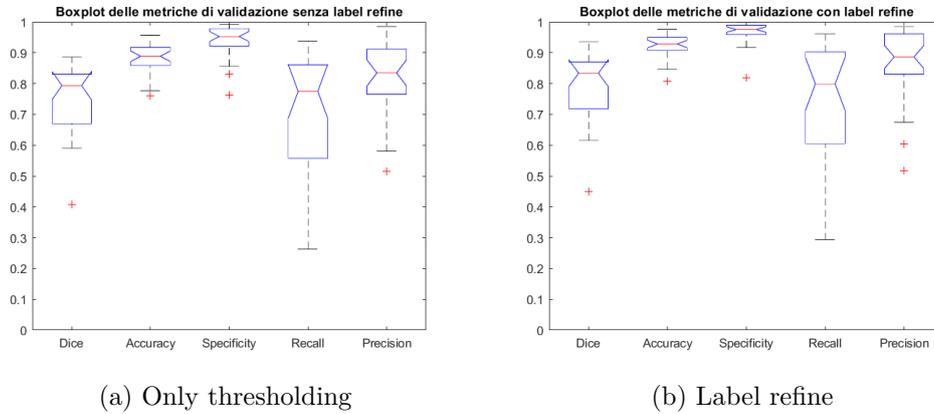


Figura 3.33: Metriche di validazione Heatmap con DenseNet con i pixel del vetrino

Dai grafici sopra è facilmente osservabile l'aumento degli indicatori di *accuratezza* e *specificità*. Per quest'ultima, addirittura la DenseNet con label refining rappresenta la rete migliore. Come per le altre, la considerazione dei veri negativi del vetrino porta ad una riduzione netta della *recall* con solo thresholding.

Le metriche con i loro valori medi e deviazioni standard sono riassunte nella Tabella 3.11.

Tabella 3.11: Metriche Heatmap DenseNet

	Dice	Accuratezza	Specificità	Recall	Precisione
Only thresholding (solo tessuto)	0.789 ± 0.093	0.793 ± 0.076	0.79 ± 0.146	0.792 ± 0.154	0.822 ± 0.12
Label refine (solo tessuto)	0.789 ± 0.11	0.808 ± 0.089	0.862 ± 0.11	0.762 ± 0.17	0.862 ± 0.114
Only thresholding (con vetrino)	0.751 ± 0.106	0.882 ± 0.05	0.94 ± 0.054	0.73 ± 0.17	0.822 ± 0.12
Label refine (con vetrino)	0.789 ± 0.11	0.923 ± 0.037	0.966 ± 0.037	0.762 ± 0.17	0.862 ± 0.114

Fusione dei modelli con media pesata

In conclusione vengono esposti i risultati del modello finale ottenuto dalla media pesata delle reti precedenti. Grazie alla ricerca dei pesi migliori, il modello finale è in grado di beneficiare degli aspetti positivi di ciascuna rete.

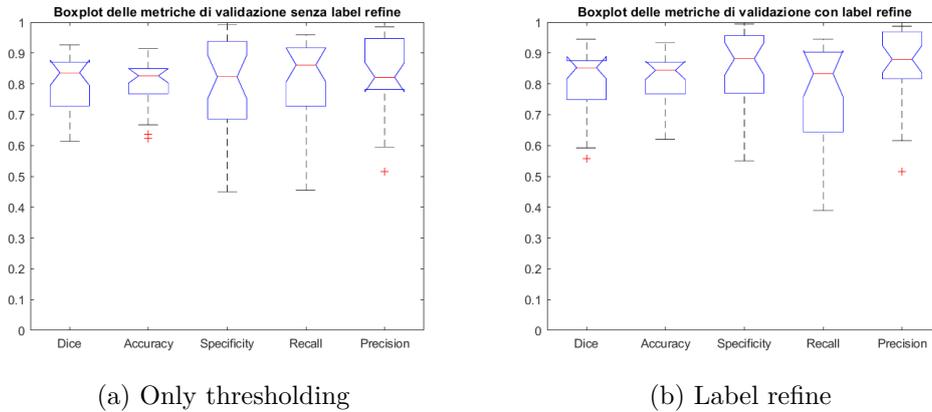


Figura 3.34: Metriche di validazione Heatmap con fusione dei modelli considerando solo il tessuto

Paragonando le performances delle reti singole a quelle del modello finale si può monitorare un incremento di tutte le metriche. Nello specifico, considerando il caso della sola operazione di sogliatura sulla softmax, il modello di insieme è il migliore in termini di *Dice*, *Accuratezza*, *Specificità* e *Precisione*; per la *Recall* è secondo solo alla VGG. Analoghi risultati si hanno includendo il post processing delle maschere; per le quali la *Recall* questa volta è quasi uguale a quella raggiunta dalla VGG (0.78 contro 0.783). Tuttavia, la *Specificità* della DenseNet supera leggermente quella del modello finale.

Vengono poi riportati i boxplot delle metriche precedenti includendo i pixel bianchi del vetrino nei veri negativi (Figura 3.35).

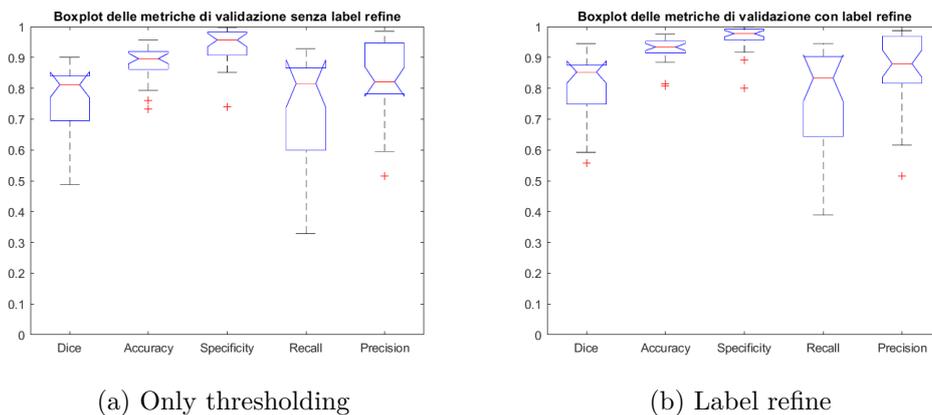


Figura 3.35: Metriche di validazione Heatmap con fusione dei modelli e con i pixel del vetrino

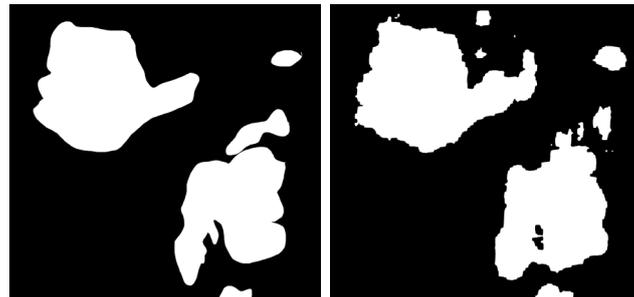
Considerando anche il vetrino bianco, si riscontrano gli stessi risultati della condizione precedente. In questo caso però, come avveniva per i modelli singoli si assiste

ad un aumento considerevole dell'*Accuratezza* e della *Specificità*. In conclusione vengono riassunte nella Tabella 3.12 le metriche del modello finale.

Tabella 3.12: Metriche Heatmap con fusione delle reti

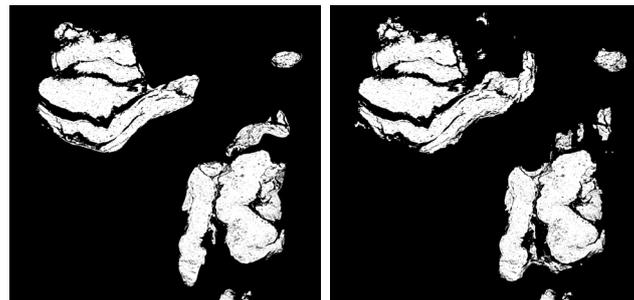
	Dice	Accuratezza	Specificità	Recall	Precisione
Only thresholding (solo tessuto)	0.8 ± 0.092	0.802 ± 0.076	0.79 ± 0.156	0.807 ± 0.153	0.83 ± 0.125
Label refine (solo tessuto)	0.8 ± 0.105	0.816 ± 0.087	0.857 ± 0.123	0.78 ± 0.167	0.864 ± 0.12
Only thresholding (con vetrino)	0.762 ± 0.105	0.886 ± 0.053	0.94 ± 0.058	0.744 ± 0.17	0.83 ± 0.125
Label refine (con vetrino)	0.8 ± 0.105	0.926 ± 0.041	0.964 ± 0.042	0.78 ± 0.166	0.864 ± 0.12

Per il modello finale viene quindi scelto di utilizzare il post processing nella definizione delle maschere binarie del tessuto tumorale (Figura 3.36). Il modello con label refine presenta quindi valori elevati di tutte le metriche eccetto della *Recall* che andrebbe migliorata aumentando il numero di veri positivi della maschera automatica.



(a) Maschera manuale

(b) Maschera automatica



(c) Maschera manuale con label refine

(d) Maschera automatica con label refine

Figura 3.36: Maschere del modello finale di insieme

Capitolo 4

Conclusioni e sviluppi futuri

4.1 Conclusioni

In questo progetto di tesi è stato proposto e sviluppato un algoritmo automatico per il riconoscimento del carcinoma polmonare in immagini istopatologiche. Il sistema elaborato si pone come supporto alla diagnosi definitiva del patologo o come strumento da utilizzare per una seconda opinione sulla biopsia polmonare in esame. L'output rapido dell'algoritmo consente di ridurre i tempi di analisi incanalando il medico verso la valutazione finale anziché partire completamente da zero. Il progetto è incentrato sul deep learning e nello specifico sulle reti convoluzionali o CNN. Le performances del modello finale, ottenuto come combinazione mediante media pesata di tre architetture diverse di rete, sono molto promettenti. Complessivamente il sistema realizzato ha raggiunto un' *AUC* di 96.1% sul training set e di 95.7% sul test set, rivelandosi un classificatore molto accurato e con buona capacità di generalizzazione. Le prestazioni dell'algoritmo nella creazione delle Heatmaps sulle intere WSI sono anch'esse molto soddisfacenti con un *Dice score* medio del 80%. Il tempo computazionale impiegato dall'algoritmo durante la fase di inference è molto variabile e dipende dalle dimensioni della WSI in esame. Si stima che per WSI di medie dimensioni il tempo per l'elaborazione sia pari a 15.17 ± 10.54 min, con GPU NVIDIA TESLA P100, disponibile sulla piattaforma online Kaggle. Per WSI contenenti porzioni tissutali di piccole dimensioni il tempo per la generazione delle heatmap si riduce anche a meno di 2 min. Dunque le potenzialità del lavoro di tesi in termini di affidabilità e di velocità, potrebbero garantire un'ottima base di partenza, al netto di miglioramenti futuri, nell'ottica di progettazione di un *tool* che possa assistere il patologo nell'analisi istopatologica del polmone.

4.2 Sviluppi futuri

Il sistema realizzato, pur manifestando un buon livello di accuratezza nel riconoscimento del tessuto tumorale, presenta alcuni limiti da superare per considerarlo un valido strumento di assistenza al patologo. In primo luogo, si potrebbe pensare ad un nuovo algoritmo di estrazione smart delle patch da applicare all'intera WSI e non solo alle Tiles 3000x3000. Tale passaggio porterebbe ad un aumento significativo del numero di patch per l'allenamento delle reti con la possibilità di un ulteriore miglioramento delle prestazioni del modello finale. Tuttavia, il nuovo algoritmo richiederebbe un ulteriore step di selection delle patch in modo da estrarle da zone tissutali della WSI distanti spazialmente e non concentrate in un'unica regione. In aggiunta, nella versione finale del modello si potrebbe includere lo step di stain normalization nella fase di preprocessing con lo scopo di ridurre le variazioni di colore e di intensità delle immagini al fine di aumentare ulteriormente l'accuratezza finale con un training meno rumoroso. Infine, nell'ambito della creazione delle Heatmap si potrebbe invece proporre un algoritmo che richiami l'estrazione smart delle patch. In sostanza si attribuirebbe un peso elevato alla predizione della rete qualora la patch soddisfacesse i criteri di estrazione e un peso minore nel caso opposto. Ciò permetterebbe di ridurre la predizione del modello finale in zone in cui sono presenti artefatti di tissue folds e/o di sfocatura o in regioni caratterizzate da una ridotta densità dei nuclei cellulari.

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.

Un sentito ringraziamento al Prof. Filippo Molinari per avermi dato l'opportunità di realizzare questo progetto di tesi e per aver suscitato in me, con le sue lezioni, grande interesse e passione per gli argomenti. Un ringraziamento particolare all'Ing. Massimo Salvi per la sua infinita disponibilità e il suo costante supporto durante tutto il lavoro svolto. Grazie per tutti gli utili consigli, le spiegazioni e gli insegnamenti di questi mesi.

Ringrazio tutti gli amici e colleghi universitari che hanno condiviso con me questi anni duri e difficili ma indimenticabili.

Ringrazio gli amici di una vita della mia compagnia, gli unici capaci di donarmi quella spensieratezza e leggerezza che il percorso universitario molto spesso sottrae. Spero che il nostro rapporto rimanga sempre così. Vi voglio bene.

Ringrazio la mia grande famiglia italiana e spagnola per aver sempre creduto nelle mie potenzialità e avermi dato tanta forza in alcuni momenti della mia vita. Tra tutti un ringraziamento speciale a Nonna Assunta per essere da sempre il collante di tutti noi. Vi voglio bene.

Ringrazio Antonio, Lucia e Felice, la mia seconda famiglia. Grazie per avermi accolta con tanta dolcezza e tanto amore nella vostra casa. Mi sento davvero molto fortunata ad avervi incontrati. Siete una famiglia meravigliosa e io sono immensamente lieta di farne parte, spero per sempre.

Ringrazio la mia splendida famiglia.

Grazie alla mia sorellina acquisita Tabea, che con la sua infinita purezza e delicatezza è entrata senza troppa difficoltà nel mio cuore, conquistandomi totalmente. Non cambiare mai, rimani così come sei semplice e dolce come solo tu sai essere. Sei stupenda, ti voglio bene.

Mamma, Papà, Christian e Vincenzo. Per me risulta davvero difficile nominare uno di voi prima dell'altro perché voi tutti insieme siete il mio cuore e la mia vita. Vi ringrazio davvero per tutto. Grazie per essere la mia forza ogni giorno, per essere l'esempio a cui mi ispiro oggi e anche domani, per essere il mio punto di riferimento

in ogni momento. Voi tutti insieme siete la parte migliore di me e io posso solo ringraziarvi per esistere nella mia vita. Ogni mio traguardo non avrebbe senso se non ci foste voi con cui dividerlo.

Infine ringrazio te, angelo mio, che da lassù mi hai sempre protetta quando ne avevo bisogno. In ogni momento di ansia e preoccupazione sapevo di averti sempre lì con me e questo mi ha dato la forza e la sicurezza di affrontare qualunque cosa. Non mi sono mai sentita sola in questo percorso. Non sai quanto mi faccia male non averti qui presente in uno dei giorni più importanti della mia vita, ma ciò che mi conforta è pensare che da lassù mi guardi ogni giorno, spero orgoglioso della persona che sono diventata. Grazie José, sei il nostro angelo.

Bibliografia

- [1] C. Stanfield, *Fisiologia*, EdiSES, 2012.
- [2] francycara04, «tes» [Online]. Available: <https://www.tes.com/lessons/FCLbG9aBmsvrtA/apparatorespiratorio>. [Consultato il giorno 25 10 2020].
- [3] M. P. Trainer, «Apparato Respiratorio,» 17 03 2020. [Online]. Available: <https://www.my-personaltrainer.it/salute-benessere/apparato-respiratorio.html>. [Consultato il giorno 25 10 2020].
- [4] Wikipedia, «Wikipedia,» [Online]. Available: <https://it.wikipedia.org/wiki/Alveolo-polmonare>. [Consultato il giorno 25 10 2020].
- [5] Wikipedia, «Wikipedia,» [Online]. Available: <https://it.wikipedia.org/wiki/Polmone>. [Consultato il giorno 25 10 2020].
- [6] Wikipedia, «Wikipedia,» [Online]. Available: <https://it.wikipedia.org/wiki/Pleura>. [Consultato il giorno 25 10 2020].
- [7] J. Tomashefski e C. Farver, «Anatomy and Histology of the Lung,» Springer, pp. 20-48, 2008.
- [8] C. Iannuzzi, «DocPlayer,» 2016-2017. [Online]. Available: <http://docplayer.it/44068389-Apparato-respiratorio.html>. [Consultato il giorno 25 10 2020].
- [9] Unife, «Apparato respiratorio».
- [10] Y. Khan e D. T. Lynch, «Histology, Lung,» StatPearls Publishing,, 2020.
- [11] T. H. Guide, «Functions of the Respiratory Portion,» [Online]. Available: <https://www.histology.leeds.ac.uk/respiratory/respiratory.php>. [Consultato il giorno 25 10 2020].
- [12] E. w. book, «ERS white book,» [Online]. Available: <https://www.erswhitebook.org/files/public/Italian%20PDFs/19.%20White%20book%20lung%20cancer.pdf>. [Consultato il giorno 25 10 2020].
- [13] ISSalute, «ISSalute,» 28 02 2018. [Online]. Available: <https://www.issalute.it/index.php/la-salute-dalla-a-alla-z-menu/t/tumore-al-polmone>: :text

- =Sebbene%20anche%20persone%20che%20non,che%20si%20fuma%20una%20sigaretta..
[Consultato il giorno 25 10 2020].
- [14] AIRC, «AIRC,» 29 05 2020. [Online]. Available: <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-al-polmone>. [Consultato il giorno 25 10 2020].
- [15] «Lungevity,» [Online]. Available: <https://lungevity.org/for-patients-caregivers/lung-cancer-101/types-of-lung-cancer>. [Consultato il giorno 25 10 2020].
- [16] M. P. Trainer, «My Personal Trainer,» 25 02 2020. [Online]. Available: <https://www.my-personaltrainer.it/salute/cancro-polmone-non-a-piccole-cellule.html>. [Consultato il giorno 25 10 2020].
- [17] Wikipedia, «Carcinoma del polmone,» [Online]. Available: <https://it.wikipedia.org/wiki/Carcinomadelpolmone>. [Consultato il giorno 25 10 2020].
- [18] Lungevity, «Lungevity,» [Online]. Available: <https://lungevity.org/sites/default/files/request-materials/LUNGevity-adenocarcinoma-booklet-020516.pdf>. [Consultato il giorno 25 10 2020].
- [19] Infomedics, «Carcinoma Polmonare: Stadiazione clinica,» [Online]. Available: <https://www.infomedics.it/therapeutic-areas/carcinoma-polmonare/iter-diagnostico/stadiazione-clinica.html>. [Consultato il giorno 25 10 2020].
- [20] Lungevity, «Diagnosing Lung Cancer,» [Online]. Available: <https://lungevity.org/for-patients-caregivers/lung-cancer-101/diagnosing-lung-cancer>. [Consultato il giorno 25 10 2020].
- [21] sMQ, «Ventajas de la cirugía laparoscópica,» [Online]. Available: <https://www.especialistasencirugia.com/blog/ventajas-cirugia-laparoscopica>. [Consultato il giorno 25 10 2020].
- [22] L. Spaggiari, «Lorenzo Spaggiari,» [Online]. Available: <https://www.lorenzospaggiari.com/2019/04/05/focus-aprile-2019-la-chirurgia-maggiore/>. [Consultato il giorno 25 10 2020].
- [23] T. O. Program, «Non-Small Cell Lung Cancer,» [Online]. Available: <https://top.ucsf.edu/conditions-procedures/non-small-cell-lung-cancer.aspx>. [Consultato il giorno 25 10 2020].
- [24] Wikipedia, «Toracocentesi,» [Online]. Available: <https://it.wikipedia.org/wiki/Toracocentesi>. [Consultato il giorno 25 10 2020].

- [25] Lungevity, «Biopsies,» [Online]. Available: <https://lungevity.org/for-patients-caregivers/lung-cancer-101/diagnosing-lung-cancer/biopsies>. [Consultato il giorno 25 10 2020].
- [26] Wikipedia, «Istologia,» [Online]. Available: <https://it.wikipedia.org/wiki/Istologia>. [Consultato il giorno 25 10 2020].
- [27] M. p. tutti, «Fasi di allestimento di un campione (prelievo, fissazione, inclusione, colorazione),» 08 03 2010. [Online]. Available: <https://www.medicinapertutti.it/argomento/fasi-di-allestimento-di-un-campione-prelievo-fissazione-inclusione-colorazione/>. [Consultato il giorno 25 10 2020].
- [28] M. Minniti, «microbiologiaitalia,» 11 05 2020. [Online]. Available: <https://www.microbiologiaitalia.it/microscopia/preparazione-dei-campioni-per-la-microscopia-elettronica/>. [Consultato il giorno 25 10 2020].
- [29] A. d. Istologia, «Colorazioni istologiche,» [Online]. Available: <http://www.istologia.unige.it/page1/page29/page29.html>. [Consultato il giorno 25 10 2020].
- [30] S. Nam, Y. Chong, C. Jung, T. Kwak, J. Lee, J. Park, M. Rho e H. Go, «Introduction to digital pathology and computer-aided pathology,» *Journal of pathology and translational medicine*, vol. 54, n. 2, pp. 125-134, 2019.
- [31] Leica Biosystems, «Leica Biosystems,» [Online]. Available: <https://www.leicabiosystems.com/digital-pathology/scan/aperio-cs2/>. [Consultato il giorno 23 10 2020].
- [32] S. Heffner, O. Colgan e C. Doolan, «Leica biosystems,» [Online]. Available: <https://www.leicabiosystems.com/knowledge-pathway/digital-pathology/WhataretheBenefitsofDigitalPathology>. [Consultato il giorno 10 22 2020].
- [33] S. Taqi, S. Sami, L. Sami e S. Zaki, «. A review of artifacts in histopathology,» *J Oral Maxillofac Pathol*, vol. 22, n. 2, p. 279, 2018.
- [34] G. O. Rolls, N. J. Farmer e J. B. Hall, «Artificats in Histological and Cytological Preparations,» *Scientia Leica Microsystems' Education Series*, vol. 21, 2008.
- [35] *Journal of Pathology Informatics*, «Journal of Pathology Informatics,» [Online]. Available: <https://www.jpathinformatics.org/viewimage.asp?img=JPatholInform20167126181770f4.jpg>. [Consultato il giorno 23 10 2020].
- [36] Z. L. e. al., «ACDC Lung HP,» [Online]. Available: <https://acdc-lunghp.grand-challenge.org/>. [Consultato il giorno 25 10 2020].
- [37] Z. Li, J. Zhang, T. Tan, X. Teng, X. Sun, Y. Li, L. Liu, Y. Xiao, B. Lee, Q. Zhang, S. Sun, Y. Zheng, J. Yan, N. Li, Y. Hong, J. Ko, H. Jung, Y. Liu, Y. Chen, C. Wang, V. Yurovskiy, P. Maevskikh, V. Khanagha, Y. Jiang, X. Feng e Li, «Deep Learning Methods for Lung Cancer Segmentation in Whole-slide Histopathology Images - the ACDC@LungHP Challenge 2019,» *ArXiv*, 2020.

- [38] M. Salvi, N. Michielli e F. Molinari, «Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology,» Elsevier, 2020.
- [39] M. Salvi e F. Molinari, «Multi-tissue and multi-scale approach for nuclei segmentation in HE stained images,» *BioMed Eng OnLine*, vol. 17, n. 89, 2018.
- [40] M. Saha, C. Chakraborty e D. Racoceanu, «Efficient Deep Learning Model for Mitosis Detection,» Elsevier, vol. 64, pp. 29-40, 2018.
- [41] Kothari S, Phan JH, Wang MD. Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. *J Pathol Inform* 2013;4:22.
- [42] PyImageSearch, A. Rosebrock, (2015 September 7) Blur detection with OpenCV [Online]. Available: <https://www.pyimagesearch.com/2015/09/07/blur-detection-with-opencv/>
- [43] Moles Lopez X, D'Andrea E, Barbot P, Bridoux A-S, Rorive S, et al. (2013) An Automated Blur Detection Method for Histological Whole Slide Imaging. *PLoS ONE* 8(12): e82710. doi:10.1371/journal.pone.0082710
- [44] AIconcept, «Cos'è una rete neurale artificiale,» AIconcept.
- [45] S. Saha, «A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,» towards data science, 15 12 2018. [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. [Consultato il giorno 27 10 2020].
- [46] L. Govoni, «Semplice architettura di rete neurale convoluzionale,» [Online]. Available: <https://lorenzogovoni.com/architettura-di-rete-neurale-convoluzionale/>. [Consultato il giorno 27 10 2020].
- [47] P. Marcelino, «towards data science- Transfer learning from pre-trained models,» 23 10 2018. [Online]. Available: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>. [Consultato il giorno 27 10 2020].
- [48] S. Singh, «Understanding the Bias-Variance Tradeoff,» towards data science, 21 05 2018. [Online]. Available: <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>. [Consultato il giorno 27 10 2020].
- [49] J. Wei, «VGG Neural Networks: The Next Step After AlexNet,» 03 07 2019. [Online]. Available: <https://towardsdatascience.com/vgg-neural-networks-the-next-step-after-alexnet-3f91fa9ffe2c>. [Consultato il giorno 29 10 2020].
- [50] A. Sachan, «CV-TRICKS.com- Detailed Guide to Understand and Implement ResNets,» [Online]. Available: <https://cv-tricks.com/keras/understand-implement-resnets/>. [Consultato il giorno 28 10 2020].

- [51] V. Kurama, «PaperspaceBlog- A Review of Popular Deep Learning Architectures: ResNet, InceptionV3, and SqueezeNet,» 05 06 2020. [Online]. Available: <https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/>. [Consultato il giorno 29 10 2020].
- [52] A. Arora, «DenseNet Architecture Explained with PyTorch Implementation from TorchVision,» 02 08 2020. [Online]. Available: <https://amaarora.github.io/2020/08/02/densenets.html>. [Consultato il giorno 30 10 2020].
- [53] J. Brownlee, "How to Develop a Weighted Average Ensemble for Deep Learning Neural Networks", Machine Learning Mastery, [Online]. Available: <https://machinelearningmastery.com/weighted-average-ensemble-for-deep-learning-neural-networks/>
- [54] L. Govoni, «Lorenzo Govoni,» [Online]. Available: <https://lorenzogovoni.com/matrici-di-confusione/>. [Consultato il giorno 21 10 2020].
- [55] MC.AI, «Confusion Matrix No More Confusing,» 30 09 2018. [Online]. Available: <https://mc.ai/confusion-matrix-no-more-confusing/>. [Consultato il giorno 21 10 2020].
- [56] A. Mishra, «towards data science,» 24 02 2018. [Online]. Available: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>: :
text=Area%20Under%20Curve(AUC)%20is,a%20randomly%20chosen%20negative%20example.. [Consultato il giorno 21 10 2020].
- [57] E. Bino, «Tesi di Laurea: Curve ROC per la valutazione di test diagnostici per eventi el tempo,» Università degli studi di Padova, 2009-2010.
- [58] V. F. M. A. e PhD, «Commentary: Should we rely on receiver operating characteristic curves? From submarines to medical tests, the answer is a definite maybe!,» The Journal of Thoracic and Cardiovascular Surgery, vol. 157, pp. 2354-2355, 2018.
- [59] Wikipedia, «Receiver operating Characteristic,» [Online]. Available: https://en.wikipedia.org/wiki/Receiver_operating_characteristic. [Consultato il giorno 21 10 2020].
- [60] G. D'Arrigo, F. Provenzano, C. Torino, C. Zoccali e G. Tripepi, «I TEST DIAGNOSTICI E L'ANALISI DELLA CURVA ROC,» G Ital Nefrol, vol. 28, n. 6, pp. 642-647, 2011.
- [61] I. Mayachita, «towards data science,» 09 08 2020. [Online]. Available: <https://towardsdatascience.com/graph-convolutional-networks-on-node-classification-2b6bbec1d042>. [Consultato il giorno 22 10 2020].
- [62] Wikipedia, «t-distributed stochastic neighbor embedding,» [Online]. Available: https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding. [Consultato il giorno 21 10 2020].

- [63] H. Gupta, «buZZrobot,» 04 12 2017. [Online]. Available: <https://buzzrobot.com/using-t-sne-to-visualise-how-your-deep-model-thinks-4ba6da0c63a0>. [Consultato il giorno 22 10 20].
- [64] Kotikalapudi, Raghavendra e contributors, «GitHub,» 2017. [Online]. Available: <https://raghakot.github.io/keras-vis/visualizations/saliency/>. [Consultato il giorno 22 10 2020].
- [65] Matlab, «Understand Network Predictions Using Occlusion,» [Online]. Available: <https://ch.mathworks.com/help/deeplearning/ug/understand-network-predictions-using-occlusion.html>. [Consultato il giorno 22 10 2020].
- [66] d_believer, «GeeksforGeeks,» 10 05 2020. [Online]. Available: <https://www.geeksforgeeks.org/what-is-saliency-map/>. [Consultato il giorno 22 10 2020].
- [67] Wikipedia, «Sørensen–Dice coefficient,» [Online]. Available: https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient. [Consultato il giorno 22 10 2020].
- [68] Wikipedia, «Indice di Jaccard,» [Online]. Available: <https://it.wikipedia.org/wiki/IndicediJaccard> [Consultato il giorno 22 10 2020].