POLITECNICO DI TORINO

FACOLTÀ DI INGEGNERIA

Corso di Laurea Magistrale in Ingegneria Biomedica

Sviluppo di un algoritmo automatico per l'assegnazione dello score di Gleason in immagini istopatologiche prostatiche

Relatori:

Prof. Filippo Molinari

Ing. Massimo Salvi

Candidato:

Davide Barra

Dicembre 2019

Indice

Elenco delle figure

Elenco delle tabelle

Sommario

Il cancro alla prostata rappresenta tutt'oggi il 20% di tutti i tumori diagnosticati nell'uomo. E' il secondo tumore più comune dopo il cancro al polmone ed è al terzo posto per mortalità dopo il cancro al polmone e il cancro al colon. Molto raramente si presenta prima dei 40 anni e la sua probabilità di incidenza cresce all'aumentare dell'età fino ad arrivare ad una percentuale del 75% negli uomini dagli 80 anni in su. Come tutti i tumori, il cancro prostatico è caratterizzato dalla crescita e dall'espansione incontrollata delle cellule della ghiandola prostatica.

La sua diagnosi è affidata all'ispezione della biopsia prostatica da parte di un patologo esperto, in grado di individuare il tumore ed assegnare un punteggio di malignità detto score di Gleason. La scala di Gleason, ideata dal Dr Donald Gleason rappresenta il sistema di prognosi più importante per il cancro prostatico e valuta la struttura ghiandolare assegnando una delle seguenti classi: Benigno, Gleason 3, Gleason 4 e Gleason 5. Un punteggio alto indica una grave malignità del tumore con perdita progressiva della caratteristica struttura ghiandolare.

Sebbene la scala di Gleason faccia parte di un protocollo standard ideato da patologi, la valutazione visiva della biopsia prostatica è caratterizzata da soggettività e variabilità. Siccome la terapia da seguire dipende dalla gravità e dalla malignità del tumore, è importante ridurre al minimo l'errore di assegnazione dello score di Gleason.

L'obiettivo del seguente lavoro di tesi è la realizzazione di un sistema auto-

matico in grado di assegnare un corretto score di Gleason alle zone tumorali prostatiche. E' stato sfruttato il dataset pubblico di Arvaniti et al. [1] composto da 886 immagini di TMA (tissue microarrays).
Per la realizzazione dell'algoritmo sono state utilizzate le reti neurali convoluzionali (CNN) che nell'ambito del machine learning rappresentano un approccio valido ed interessante per la classificazione di immagini istopatologiche. Sono state valutate le performance su diversi classificatori che sfruttano le CNN per distinguere 4 classi (Benigno, Gleason 3, Gleason 4 e Gleason 5). Le reti sono state allenate con le patch estratte dalle immagini tramite un algoritmo automatico di estrazione intelligente guidato dall'anatomia microscopica delle biopsie analizzate.
Il classificatore con le prestazioni migliori è stato successivamente testato su nuove immagini al fine di calcolare lo score di Gleason globale e confrontarlo con l'assegnazione di due patologi esperti.
Questo sistema, ulteriormente ottimizzabile, può ridurre i tempi di analisi bioptica e fornire un supporto ad un patologo come seconda opinione al fine di assegnare uno score di Gleason corretto e poter individuare la terapia più consona per il paziente.

Capitolo 1

Introduzione

1.1 Prostata

La prostata è una ghiandola fibro-muscolare che fa parte dell'apparato genitale maschile e svolge l'importante funzione di produrre il liquido prostatico, elemento fondamentale dello sperma. È caratterizzata da una forma a piramide di consistenza elastica con l'apice rivolto verso il basso, ma in condizioni patologiche può variare. Si colloca a 5 cm circa dal retto con la base a contatto con la vescica mentre l'apice segna il passaggio all'uretra.

In condizioni normali, presenta un diametro di circa 4 cm, una lunghezza di 3 cm e una larghezza di 2cm. Oltre ad una base e un apice, la prostata presenta una faccia anteriore, una faccia posteriore e due facce infero-laterali. In particolare, la faccia anteriore e le facce infero-laterali sono ricoperte dalla fascia endopelvica e dalla fascia prostatica laterale. Le facce infero-laterali sono in rapporto con il muscolo elevatore dell'ano. La faccia posteriore invece è separata dal retto dalla fascia del Denonvilliers o fascia retroprostatica.
Il tessuto ghiandolare prostatico è costituito da 3 zone (transizionale, centrale e periferica) e dallo stroma fibromuscolare. La zona più grande è quella periferica che occupa circa il 70% della prostata e ingloba la zona transizionale. La zona

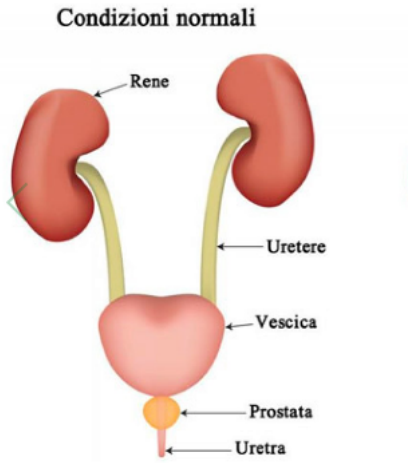


Figura 1.1: Posizione della prostata.

centrale invece occupa il 25% del volume ed è collocata posteriormente rispetto alla zona periferica. Infine la zona transizionale occupa solamente il 5% e circoa la parte distale dell'uretra.

1.1.1 Funzioni della prostata

La prostata svolge numerosi funzioni nel corpo umano [2]:

- **Produzione del liquido seminale**: insieme alle cellule spermatiche, al fluido prodotto dalla vescicola seminale e alle secrezioni rilasciate da un altre ghiandole che si trovano al di sotto della prostata (ghiandole bulbouretrali), il liquido prostatico costituisce il liquido seminale. Tutti questi fluidi sono miscelati nell'uretra.Il liquido prostatico è importante per la vitalità degli spermatozoi e di conseguenza per la fertilità. Infatti contiene enzimi come il PSA (antigene prostatico specifico) che garantiscono la corretta fluidità del seme dopo l'eiaculazione.

- **Chiusura dell'uretra e dei dotti seminali**: durante l'eiaculazione, la prostata si occupa di chiudere l'uretra fino alla vescica per impedire

l'ingresso del seme. Allo stesso modo, durante la minzione vengono chiusi i dotti prostatici per evitare l'ingresso dell'urina.

- **Metabolismo ormonale**: il testoreno, l'ormone sessuale maschile, diventa biologicamente attivo (DHT,diidrotestoreno) all'interno della prostata.

1.1.2 Anatomia microscopica

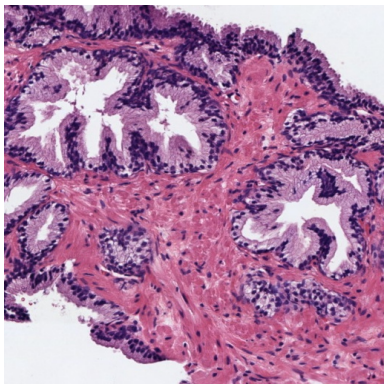


Figura 1.2: Struttura microscopica della prostata.

Il tessuto prostatico è costituito da 30-50 ghiandole tubuloalveolari ramificate immerse in uno stroma fibro-muscolare, tessuto connettivo con il compito di mantenimento e stabilità della struttura di un organo o di una ghiandola. Ogni ghiandola, situata nelle logge delimitate da setti stromali, presenta tantissimi acini (cellule secernenti) costituiti da papille che espellono il loro secreto in piccoli condotti che unendosi formano un dotto per ciascuna ghiandola. Si raccolgono in 20-30 dotti escretori che sboccano nei seni prostatici dell'uretra prostatica. I condotti escretori sono costituiti da un lume di forma e dimensione non uniforme. I piccoli dotti sono rivestiti da cellule epiteliali mentre i dotti prostatici in cui si raccolgono sono formati da un doppio strato di cellule (Figure 1.2).
Queste formazioni ghiandolari, in base alla loro posizione rispetto all'uretra

e ai dotti eiaculatori, possono essere raggruppate in 5 lobi, riconoscibili nel feto e difficilmente distinguibili in età adulta: lobo anteriore, lobo medio, lobi laterali e lobo posteriore. La maggior parte dello stroma è contenuto nella parte anteriore della prostata dove forma lo stroma fibro-muscolare, privo di ghiandole.

1.2 Patologie

1.2.1 Iperplasia prostatica benigna (IPB)

La prostata può essere soggetta a numerose patologie durante tutto il corso della vita del paziente. L'iperplasia prostatica benigna (IPB) è una patologia molto comune che è caratterizzata dall'aumento considerevole della dimensione della ghiandola prostatica. La causa dell'ingrossamento non è l'aumento delle dimensioni delle singole cellule (ipertrofia) ma l'aumento del numero delle cellule (iperplasia).

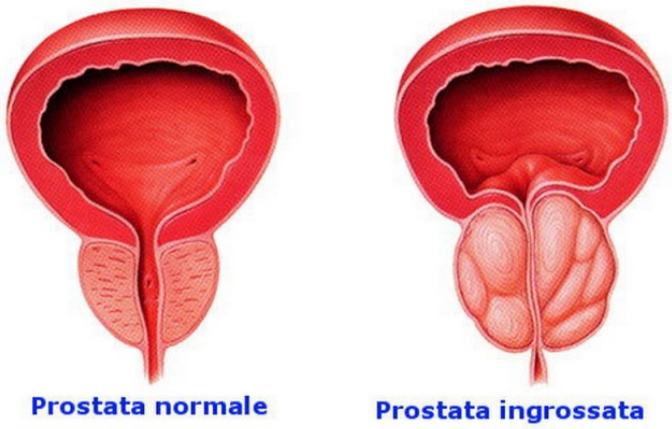


Figura 1.3: Differenza tra una prostata normale e una prostata affetta a IPB

Tra il 5% e il 10% degli uomini sopra i 40 anni ne è affetto e questa percentuale cresce in maniera esponenziale con l'invecchiamento. Le cause sono principalmente l'avanzamento dell'età e il cambiamento ormonale, ma studi

recenti hanno dimostrato anche un'importante componente genetica e di familiarità.
I sintomi sono principalmente di due tipi (ostruttivo e irritativo) e non sono correlati con le dimensioni della prostata ma con la massa della ghiandola e con il tono della muscolatura liscia. La difficoltà nella minzione e l'incompleto svuotamento della vescica fanno parte dei sintomi ostruttivi mentre l'urgenza e il bruciore nella minzione fanno parte dei sintomi irritativi.
Gli studi hanno dimostrato che non esiste nessuna correlazione tra l'IPB e il carcinoma prostatico, nonostante queste due patologie possano presentarsi contemporaneamente.

1.2.2 Carcinoma prostatico

Il carcinoma prostatico è uno dei tumori più diffusi e rappresenta circa il 20% di tutti i tumori diagnosticati. È il secondo tumore più comune nel mondo dopo il tumore al polmone ed è al terzo posto tra i tumori per mortalità dopo il tumore al polmone e il tumore al colon [3].
Le cause sono sconosciute ma possono essere collegate a diversi fattori come la genetica, la razza e lo stile di vita. L'età resta sicuramente il fattore più determinante. Infatti è raro riscontrare questo tipo di tumore negli uomini sotto i 45 anni, mentre è sempre più frequente con l'invecchiamento. Al giorno d'oggi l'età media di diagnosi si aggira intorno ai 70 anni.
Il cancro prostatico, come tutti i tumori, è caratterizzato da cellule che si riproducono in maniera incontrollata generando noduli maligni ed ingrossando la ghiandola. Successivamente, le cellule sono in grado di invadere gli organi adiacenti e ad espandersi sempre di più dando origine alle metastasi.
Una prima individuazione del carcinoma avviene tramite esame obiettivo o tramite la misurazione del PSA (antigene prostatico specifico) nel sangue. Esami più specifici per la conferma verranno poi effettuati tramite la biopsia e quindi analisi su un frammento di tessuto opportunamente asportato.

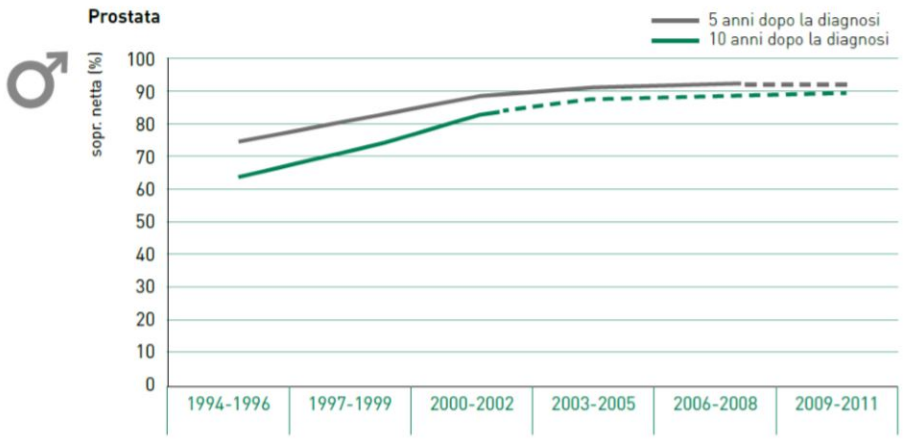


Figura 1.4: Sopravvivenza netta del carcinoma alla prostata per periodo di incidenza in Italia [4]

Diversamente rispetto ad altri tipi di tumore, il carcinoma prostatico non ha mostrato notevoli riduzioni della mortalità con la diagnosi precoce. L'incremento visibile della sopravvivenza netta nei 5 e 10 anni dopo la diagnosi (Figura 1.4) è principalmente dovuto ai continui miglioramenti delle terapie, ottimizzate per ogni paziente. Per stabilire il miglior percorso terapeutico è fondamentale riuscire ad individuare la gravità della patologia. Per questo nel campo del carcinoma prostatico, ci si affida allo score di Gleason, un indice per classificare i diversi pattern tumorali presenti e determinare la malignità del tumore.

1.3 Score di Gleason

Lo score di Gleason viene introdotto per la prima volta nel 1964 dal Dr. Donald Gleason al termine di uno studio durato 5 anni nel quale aveva notato differenti pattern istologici all'interno del tumore prostatico [5]. Per distinguerli creò questa scala di malignità che rimane tuttora il sistema di prognosi più importante per il carcinoma prostatico.
Negli anni è stato sottoposto a continue revisioni e cambiamenti, l'ultimo nel 2014 durante la conferenza promossa dall'International Society of Urological

Pathology (ISUP) [6].
Questo score inizialmente prevedeva un range di valori da 2 a 10 che si otteneva sommando lo score dei due pattern tumorali più presenti, a cui veniva assegnato un valore compreso tra 1 e 5. Quindi se il pattern più frequente ha score 3 e il secondo pattern più frequente ha score 4, lo score di Gleason dell'intera biopsia sarà $3 + 4 = 7$. Se è presente un pattern tumorale di un solo tipo, lo score di Gleason si otterrà semplicemente raddoppiando lo score di quel pattern (es. 3+3=6). I pattern con Gleason 1 e 2 (quindi Gleason score totale 2-5) non venivano più assegnati perchè nella maggior parte dei casi non erano adenocarcinomi ma adenosi quindi conformazioni benigne. Nella conferenza del 2014 è stato introdotto un nuovo sistema di classificazione che prevede 5 classi chiamate Grade Group (GG) che raggruppano le classi della vecchia classificazione.
In particolare il GG1 comprende tutti gli elementi che nella classificazione precedente avevano uno score di Gleason ≤ 6. Il GG2 comprende i vecchi score di Gleason 7 ma solamente i $3 + 4$. Il GG3 anche comprende i vecchi score di Gleason 7 ma solo i $4 + 3$. Il GG4 raggruppa tutti gli score di Gleason 8, quindi i $4 + 4$, $5 + 3$ e $3 + 5$. Infine il GG5 raggruppa gli score di Gleason 9 e 10 quindi i $4 + 5$, $5 + 4$ e i $5 + 5$. Questo nuovo metodo di assegnazione dello score di Gleason permette una classificazione più semplice ed immediata.

Il sistema su cui si basa lo score di Gleason discrimina i diversi pattern istologici in base all'architettura ghiandolare [7]. In particolare:

- **Gleason 1 (GP1)**: pattern caratterizzato da ghiandole nettamente separate e molto vicine fra loro. Sono circa tutte della stessa grandezza intermedia e mantengono la loro forma naturale arrotondata. È un pattern molto vicino al tessuto prostatico sano, per questo è molto raro e non più assegnato.

- **Gleason 2 (GP2)**: le ghiandole sono ancora di forma normale legger-

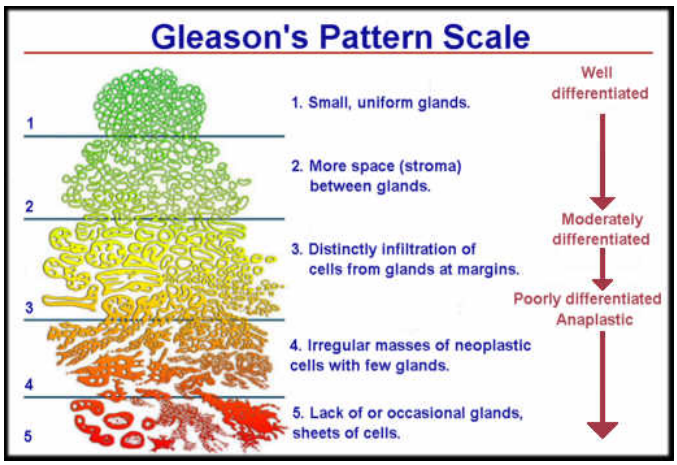


Figura 1.5: Pattern con Gleason differenti [7]

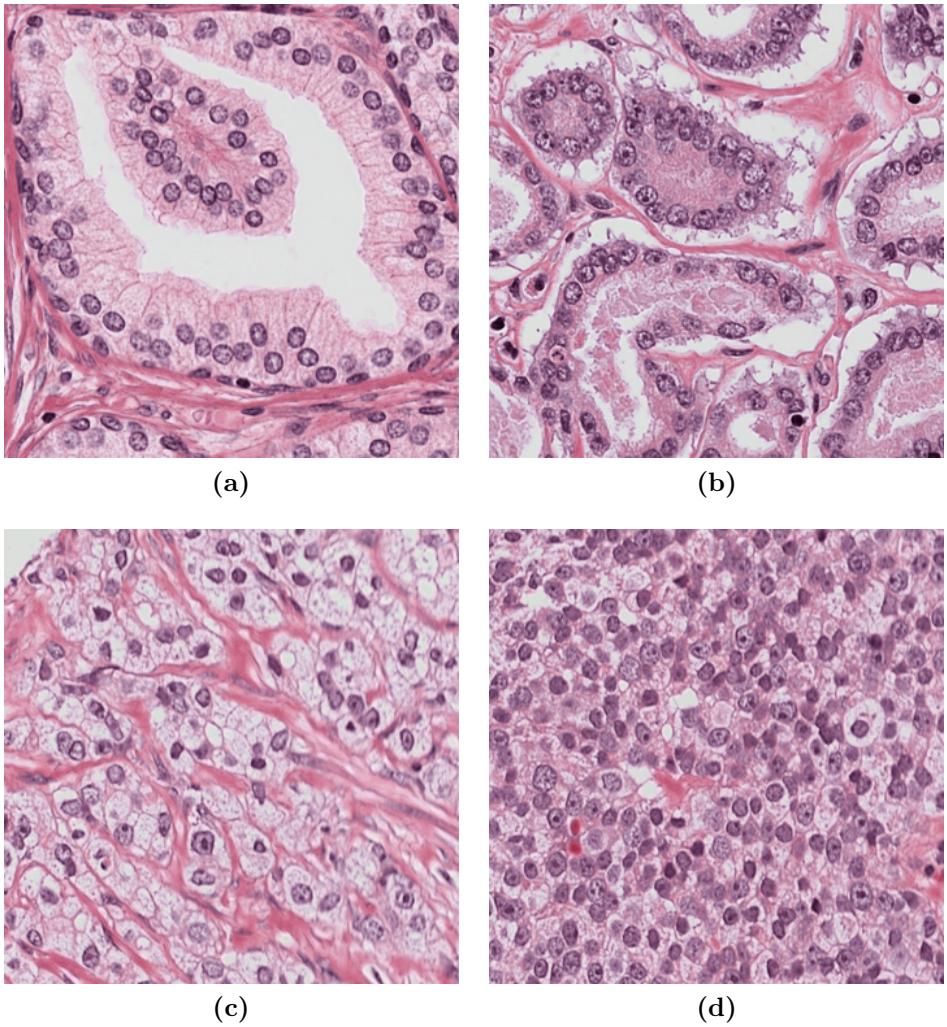
mente ovale ma le forme e le dimensioni non sono più così uniformi come nel pattern di tipo 1. Sono molto più distanti tra loro per l'aumento della crescita dello stroma. È un pattern più comune rispetto all'1 ma poco considerato nell'assegnazione del Gleason poichè il tumore è ancora alquanto differenziato quindi simile al tessuto normale.

- **Gleason 3 (GP3)**: È decisamente il pattern più comune. Le ghiandole sono ancora riconoscibili ma sono più distanti e alcune cellule maligne si sono allontanate e hanno invaso i tessuti non neoplastici vicini. Il tumore è ancora differenziato ma molto meno rispetto ai pattern 1-2.

- **Gleason 4 (GP4)**: Le ghiandole iniziano ad essere difficilmente riconoscibili perchè si sono fuse tra di loro e hanno perso la loro forma naturale. Il lume è quasi del tutto sparito e non è più completamente circondato da epitelio. Il carcinoma è scarsamente differenziato e di conseguenza aggressivo.

- **Gleason 5 (GP5)**: le ghiandole hanno perso totalmente la loro struttura caratteristica, il loro lume è scomparso e di conseguenza non sono più riconoscibili. Numerosi gruppi di cellule hanno ormai invaso lo stroma

e i tessuti circostanti. È un carcinoma non differenziato quindi molto aggressivo.



(a) **(b)**

(c) **(d)**

Figura 1.6: Esempi di pattern differenti: Benigno(a), Gleason 3 (b), Gleason 4 (c) e Gleason (5).

1.3.1 Problematiche

Assegnare un corretto ed accurato score di Gleason sulla base di una valutazione microscopica è fondamentale per poter scegliere ed intraprendere il percorso

terapeutico più adatto per il paziente. Proprio per la sua criticità, è una scelta che richiede elevata esperienza e tanto tempo, ma rimane comunque affetta da limitata riproducibilità.
Infatti diversi patologi spesso non si trovano in accordo sulla valutazione, soprattutto nel distinguere il pattern 3 dal pattern 4 dove la percentuale di riproducibilità varia tra il 25% e il 47% [8].
In un altro studio [9] in cui l'obiettivo era proprio valutare la scarsa riproducibilità dello score di Gleason, 20 casi diversi di adenocarcinoma prostatico sono stati posti all'attenzione di 21 patologi che ne hanno assegnato uno score soggettivo . I differenti risultati sono stati confrontati con lo score di riferimento di ogni slide che è stato ottenuto calcolando la mediana delle osservazioni di tutti i patologi.

Observers	Gleason scores				Total no. of readings	Percent (%) agreement with consensus
	2-4	5-6	7	8-10		
1	0	5	4	11	20	85.7
2	2	4	1	13	20	66.7
3	0	2	8	10	20	71.4
4	0	3	4	13	20	80.9
5	0	0	9	11	20	57.1
6	0	5	5	10	20	66.7
7	0	4	5	11	20	61.9
8	0	1	6	13	20	61.9
9	0	6	4	10	20	66.7
10	0	8	3	9	20	57.1
11	0	1	7	12	20	76.2
12	0	0	9	11	20	42.9
13	0	0	7	13	20	71.4
14	0	1	9	10	20	57.1
15	0	1	8	11	20	66.7
16	0	2	10	8	20	52.4
17	0	0	7	13	20	57.1
18	0	0	9	11	20	61.9
19	0	5	5	10	20	76.2
20	0	3	10	7	20	61.9
21	0	1	7	12	20	61.9
Total	2	52	137	229	420	68.0

Figura 1.7: riproducibilità inter-osservatore [9]

Come si può notare dalla 1.7 la percentuale media di consenso tra i patologi è del 68% e oscilla tra il 43% e il 92%.
Questo è solamente uno dei tantissimi studi che mettono in luce i problemi della soggettività del Gleason Score e testimoniano la necessità di sviluppare un

metodo robusto e consistente che sia di supporto al patologo al fine di ovviare alle problematiche che derivano da una assegnazione errata. Lo sviluppo di un sistema automatico in grado di assegnare lo score di Gleason di una biopsia con una certa accuratezza ridurrebbe di molto i tempi pratici di analisi della biopsia e potrebbe fornire un'importante seconda opinione per un patologo. Inoltre potrebbe risolvere il grave problema di riproducibilità dei risultati.

1.4 Stato dell'arte

L'obiettivo del seguente lavoro di tesi è lo sviluppo di un algoritmo automatico in grado di elaborare e processare un'immagine di biopsia della prostata al fine di assegnare lo Score di Gleason. Negli ultimi anni si sono moltiplicate le pubblicazioni riguardanti la ricerca di un metodo automatico per lo score di Gleason, vista anche l'espansione notevole nei campi del Machine Learnig e dell'Intelligenza artificiale.
Di seguito vengono presentati alcuni lavori recenti che daranno un'idea delle direzioni verso le quali si stanno concentrando i maggiori sforzi.

Gleason grading with convolutional neural networks

Nel 2017 *Anna Gummeson* ha ottenuto l'assegnazione automatica del Gleason Score utilizzando le reti neurali convoluzionali come classificatore[10]. Questo metodo di deep learning non è un classico metodo di classificazione con features costruite manualmente, ma si crea le proprie features allenandosi su un grossa quantità di dati.
In particolare in questo lavoro non viene usata una rete già esistente ma ne viene costruita una dal nulla al fine di ottimizzare i risultati.
L'input è costituito da immagini istologiche di tessuto prostatico suddivise in benigne, gleason 3, gleason 4 e gleason 5.La risoluzione è 40x, poi ridotta tramite un'operazione di filtro passabasso e ricampionamento. Dalle immagi-

ni vengono poi estratte le patch al fine di essere date in pasto alla rete per l'allenamento.
L'output della rete è quindi la classificazione della diverse patch dell'immagine, dalle quali si risale alla clasificazione finale dell'intera immagine.
I risultati sono positivi e l'accuratezza della classificazione dell'intera immagine raggiunge il 92.7%.

Gleason grading using feature extraction

Questo lavoro di *Hongming Xu* del 2018 invece raggiunge l'obiettivo dello score di Gleason automatico estraendo le features fondamentali e classificando con il metodo Support Vector Machine (SVM) la biopsia del paziente in Gleason 6, Gleason 7 o Gleason >8.
In particolare l'immagine della biopsia viene divisa in parti e vengono considerate sono quelle dove è presente il tumore. In seguito le features vengono estratte con un metodo innovativo che gli autori hanno chiamato CSLBP (completed and statistical local binary pattern) [11]. Prevede di caratterizzare i diversi pattern del Gleason decodificando le variazioni di intensità dei pixel rispetto al vicinato circolare. Infine il SVM multi-classe assegna ad ogni immagine il proprio score di Gleason
Questo metodo, eseguito su 312 differenti pazienti, ha raggiunto l'accuratezza del 79% che è comunque un risultato migliore rispetto ai precedenti lavori simili.

High grade cancer detection using feature selection

Nel 2019 *W. Han* ha sviluppato e validato un sistema automatico in grado di individuare un tessuto canceroso con alto score di Gleason e distinguerlo da tessuti con basso score di Gleason [12]. Questo lavoro è stato combinato con il precedente sistema di individuazione delle zone tumorali nelle immagini istopatologiche.

Dopo aver distinto nelle immagini le cellule, le ghiandole e lo stroma, sono state estratte e selezionate 22 features utilizzando leave-one-patient-out cross validation (CV) e un classificatore lineare di Fisher.
In seguito, 3 classificatori differenti (un classificatore discriminante quadrato di Fisher, un classificatore lineare logistico e il support vector machine) sono stati applicati in modo che classificassero ogni ROI (Region of interest) come tumorale o non tumorale e poi le ROI tumorali come Gleason alto o Gleason basso.
Le migliori performance sono state ottenute con il classificatore di Fisher che ha raggiunto un Area Under Curve (AUC) di 0.87.

Capitolo 2

Metodi

2.1 Dataset

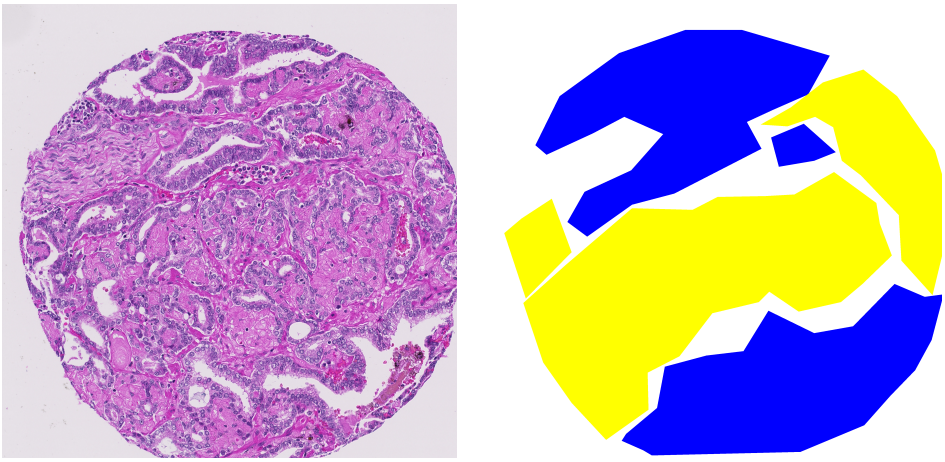
Per l'elaborazione e lo svolgimento della tesi è stato utilizzato il dataset pubblico di Arvaniti E. [1]. Il dataset comprende 5 tissue microarrays (TMAs) ovvero cilindretti di tessuto di diversi pazienti contenuti in blocchi di paraffina. Ogni TMA contiene circa 200-300 campioni di tessuto, dopo l'esclusione di quelli con artefatti o non contenenti tessuto prostatico. Alla fine sono 846 i campioni che sono stati sottoposti alla colorazione con ematossilina ed eosina (H/E).

Questo tipo di colorazione è la più utilizzata nello studio microscopico di tessuto animale. In particolare si basa sul diverso pH che presentano vari tessuti e cellule. Il nucleo e le strutture acide vengono colorate dall'ematossilina in viola mentre strutture basiche come il tessuto muscolare, il tessuto connettivo e il tessuto osseo vengono colorati dall'eosina in rosa.

Questi campioni sono stati digitalizzati con risoluzione a 40x (0.23 µm per pixel) utilizzando lo scanner digitale NanoZoomer-XR. L'immagine risultante si presenta in formato jpg e dimensione di 3100x3100x3 pixel.

Tutti i campioni sono stati accuratamente annotati da un primo patologo che

ha assegnato uno score di Gleason di 3 (blu), 4 (giallo) o 5 (rosso) ad ogni zona tumorale individuata. Tutti i campioni privi di regioni tumorali, sono stati annotati come benigni (verde).



(a) **(b)**

Figura 2.1: Esempio dell'immagine jpg del campione ZT76-39-A-1-12 (a) e delle annotazioni del patologo (b) [1]

Il dataset è poi stato diviso in training set e test set. Dato che TMA 80 è il gruppo più numeroso (245 pazienti), è stato scelto come test set ed è stato annotato da un secondo patologo per valutare quantitativamente la variabilità inter-patologo. Tutti gli altri TMAs sono stati utilizzati come training set (641 pazienti). In questo modo sia il training set che il test set risultano ben bilanciati tra le diverse classi annotate dai patologi.
Per l'intero svolgimento del lavoro è stato utilizzato il software MATLAB© 2019a.

2.2 Struttura dell'algoritmo

L'algoritmo per la realizzazione di un sistema automatico in grado di riconoscere il tumore prostatico e assegnare lo score di Gleason è strutturato come in 2.2.

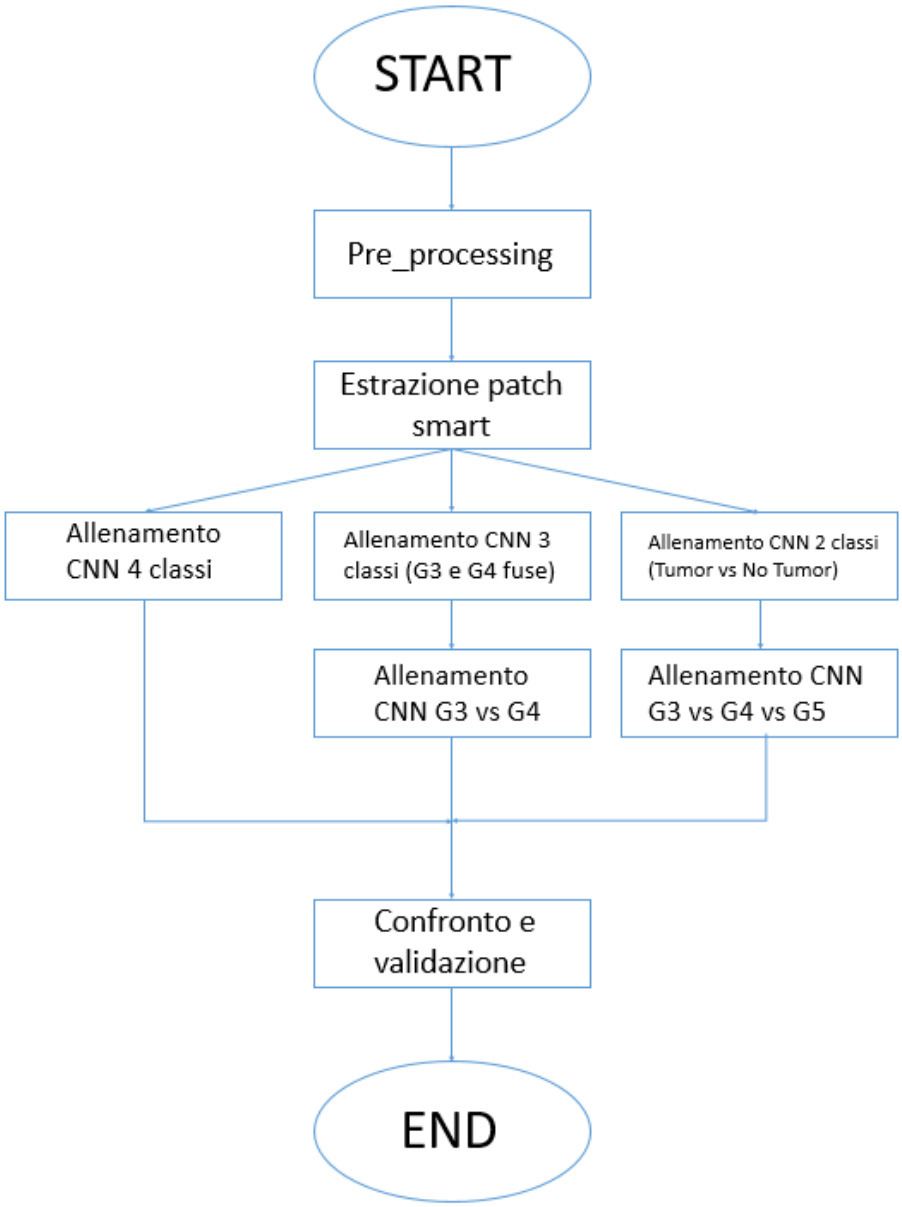


Figura 2.2: Flowchart dell'algoritmo.

L'immagine originale con dimensione 3100x3100x3 pixel viene sottoposta ad un elaborato preprocessing che prevede diversi passaggi di preparazione allo step successivo. Sono stati utilizzati script precedentemente implementati che lavorano sull'immagine per estrarre le maschere delle zone bianche, dello stro-

ma, dei nuclei e del tessuto. Inoltre l'immagine originale viene normalizzata al fine di uniformare la colorazione.
L'estrazione delle patch è stata pensata in maniera intelligente inserendo numerosi requisiti al fine di ottimizzare il risultato finale. Sono state sfruttate le caratteristiche e proprietà anatomiche microscopiche dell'immagine per poter estrarre le patch più utili in relazione al fine da raggiungere.
In seguito queste patch sono state utilizzate per portare avanti tre approcci di classificazione differenti, che prevedono tutti l'addestramento di diverse reti neurali convoluzionali (Cnn, Convolutional neural networks). Una Cnn è un tipo di rete neurale particolarmente adatta allo studio e all'analisi delle immagini. Tutti e tre questi approcci che sono stati sviluppati hanno lo stesso scopo finale ovvero discriminare i pattern di tessuto prostatico in 4 classi differenti: Benigno, Gleason 3, Gleason 4 e Gleason 5.
Una volta addestrate le reti, i risultati sono stati confrontati al fine di stabilire quale dei tre approcci fornisse un risultato più promettente e robusto.
Infine tutti i campioni designati per il test set sono stati sottoposti alle reti che classificano le regioni tumorali individuate e di seguito è stato calcolato e assegnato uno score di Gleason globale dell'intera immagine, per poter poi confrontare la classificazione finale con le annotazioni dei due patologi.

2.3 Preprocessing

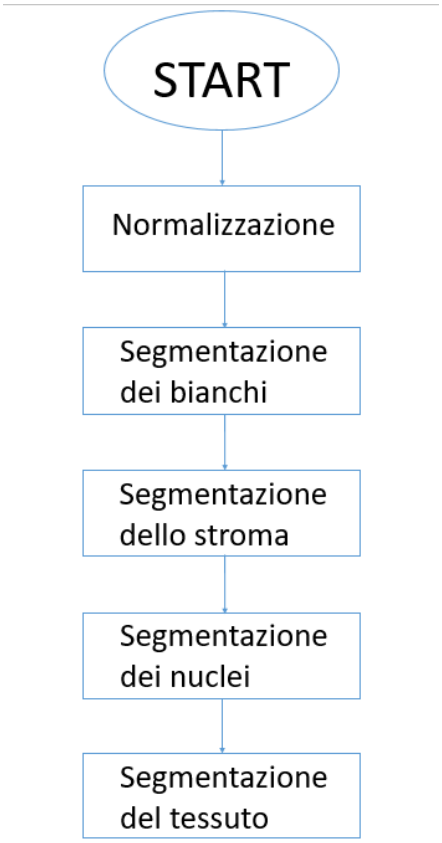


Figura 2.3: Flowchart del preprocessing

Lo scopo del preprocessing è di preparare le immagini del dataset all'estrazione delle patch. Per far rendere al meglio le reti neurali convoluzionali, è necessario allenarle su immagini più omogenee possibili per avere tutti gli input in un range di valori simile e abbattere al massimo la variabilità. Inoltre, è opportuno ricercare gli elementi nell'immagine che risultano inutili al fine del lavoro. Non considerarli nell'estrazione delle patch, permette alla rete di focalizzarsi durante l'allenamento sulle caratteristiche realmente funzionali al riconoscimento dei diversi pattern.

2.3.1 Normalizzazione dello stain delle immagini

Le immagini sono campioni di TMA trattati con la colorazione Ematossilina/Eosina che si lega a diverse sostanze biologiche. I coloranti assorbono la luce e per questo i campioni vengono analizzati al microscopio con una luce che li illumina dal basso. Nelle zone in cui non c'è tessuto, la luce riesce a passare completamente e il risultato sarà una zona di colore molto chiaro. Le zone dove la colorazione ha aderito ad una sostanza del tessuto, assorbiranno una determinata quantità di luce.
L'ammontare della luce assorbita dipende da molti fattori. La proporzione di ogni lunghezza d'onda assorbita forma lo stain vector (vettore colorazione). La colorazione può variare molto tra le diverse immagini e i fattori più importanti sono la quantità di colorazione utilizzata nel trattamento dei vetrini e la loro successiva conservazione. Per questo è fondamentale normalizzare la colorazione delle immagini del dataset per uniformarlo. Per realizzare la normalizzazione delle immagini è stato seguito il metodo proposto da Macenko et al.[13].
L'obiettivo è la separazione degli stain quindi dei diversi colori dell'immagine per poi passare alla normalizzazione. La separazione degli stain rappresenta la stima della mappa delle densità di ogni colorazione. Per la legge di Lambert-Beer, detta I_0 la luce incidente sul campione:

$$I = I_0 10^{-WH} \tag{2.1}$$

dove $I \in R^{m,n}$ è la matrice RGB delle intensità con m numero di canali RGB e n il numero di ogni pixel su ogni canale. $W \in R^{m,r}$ è la matrice dei vettori di colorazione con r numero delle colorazioni principali, quindi 2 nel caso della colorazione H/E. Infine $H \in R^{r,n}$ è la matrice di concentrazione di ogni colorante, dove le righe rappresentano i valori di concentrazione per ogni pixel del colorante r-esimo.

Dalla matrice delle intensità I si può ricavare la matrice di densità ottica dell'immagine V attraverso l'equazione

$$V = \log \frac{I}{I_0} \tag{2.2}$$

Inoltre la densità ottica V può essere definita come

$$V = WH \tag{2.3}$$

L'obiettivo è ricavare la matrice W in maniera empirica per poi calcolare la matrice H e passare alla normalizzazione vera e propria.
Per la stima della matrice W, si parte dall'immagine RGB che viene convertita in matrice OD (Optical Density). La matrice V di densità ottica, a differenza della matrice delle intensità I è linearmente separabile permettendo la stima delle colorazioni dell'immagine.

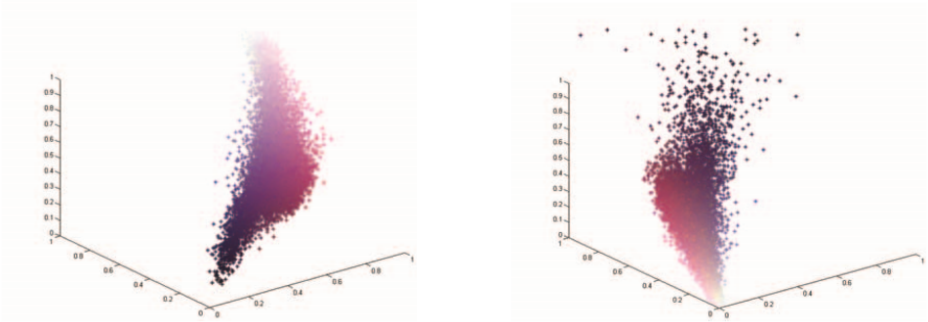


Figura 2.4: a) Si può notare come la colorazione blu e quella rosa siano separabili ma tramite una curva. b) Convertendo in densità ottica, le due colorazioni sono nuovamente separabili, ma adesso in maniera lineare.

Dalla matrice in OD vengono rimossi i pixel al di sotto di una certa soglia stabilita per evitare di considerare zone che non hanno assorbito luce, quindi dove il colorante non ha aderito.
A questo punto si passa alla scomposizione in valori singolari (SVD) della

matrice OD. Il primo passo del processo è calcolare il piano corrispondente ai due vettori con i valori singolari maggiori. Successivamente tutti i pixel vengono proiettati su questo piano e normalizzati rispetto alla lunghezza unitaria. Ora per ogni punto viene calcolato l'angolo rispetto alla prima direzione SVD cercando degli estremi robusti e considerando come massimo e minimo il $(100-\alpha)^{th}$ percentile e il α^{th} percentile. Empiricamente, α=1 fornisce risultati robusti. Una volta convertiti gli estremi in OD, si ottiene la matrice dei vettori di colorazione W che moltiplicata per l'inverso della matrice V permette di ricavare anche la matrice delle concentrazioni H.
A questo punto si passa alla normalizzazione vera e propria dell'immagine [14]. Si parte scegliendo un'immagine con una corretta ed omogenea colorazione che un patologo esperto indica come immagine di target (t) e l'obiettivo è di normalizzare un'immagine (s) uniformandola alla colorazione del target.
Partendo dalle matrici di OD V_s e V_t è possibile stimare le matrici W_s e W_t e poi calcolare H_s e H_t. A questo punto una versione scalata della matrice delle concentrazioni H_s viene combinata con la matrice W_t invece che con la matrice W_s per generare l'immagine originale normalizzata. Il procedimento è il seguente:

$$H_s^{norm}(j,:) = \frac{H_s(j,:)}{H_s^{RM}(j,:)} H_t^{RM}(j,:), j = 1, ..., r \tag{2.4}$$

$$V_s^{norm} = W_t H_s^{norm} \tag{2.5}$$

$$I_s^{norm} = I_0 10^{-V_s^{norm}} \tag{2.6}$$

dove r sono i coloranti utilizzati e H_s^{RM} è la matrice ottenuta compiendo lo pseudo massimo per ogni riga al 99^{th} percentile. Questo procedimento conserva la struttura della colorazione dell'immagine in termini di densità della colorazione H, andando a modificare solamente W avvicinandola al target.

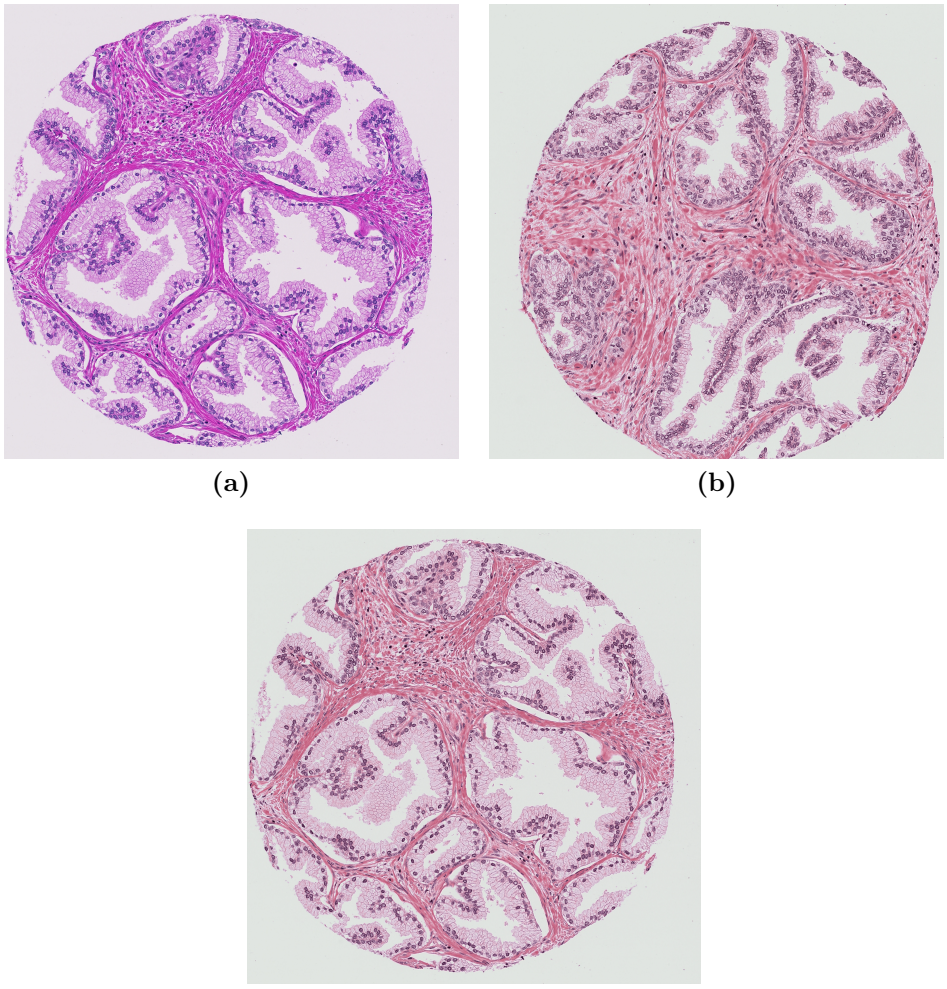
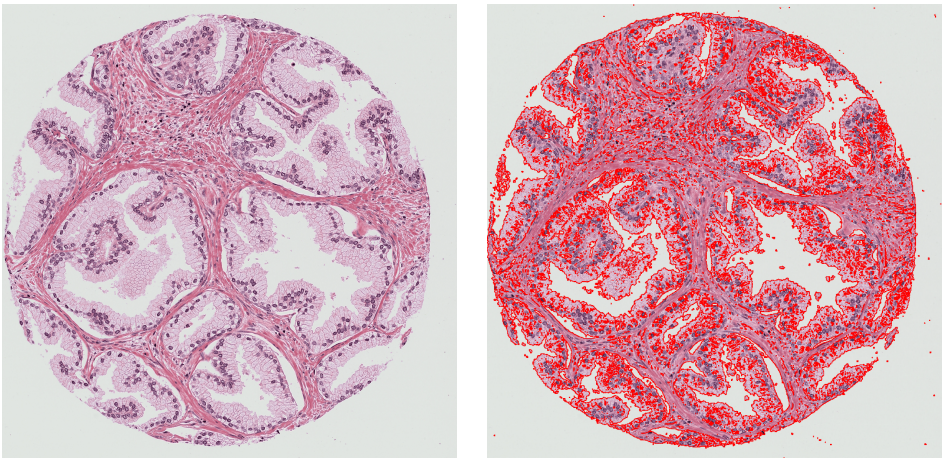


Figura 2.5: Esempio di di normalizzazione di un immagine. (a) immagine originale, (b) immagine target e (c) immagine normalizzata.

2.3.2 Segmentazione zone bianche, stroma, nuclei e tessuto

La differenza sostanziale tra il pattern sano e il pattern tumorale e anche tra i diversi pattern tumorali è da ricercare nella struttura e forma ghiandolare. Per questo, prima di estrarre le patch da dare in pasto alle reti sono state eliminate le zone dell'immagine non di interesse.

Per prima cosa, a partire dall'immagine RGB con valori compresi tra 0 e 1, tramite un thresholding con soglia impostata a 0.9 sono state individuate le zone bianche ovvero le zone in cui non è presente tessuto e la luce, non essendo stata assorbita, è riuscita a passare in maniera completa.



(a) **(b)**

Figura 2.6: Immagine originale e maschera delle zone bianche.

Anche le zone dell'immagine che contengono troppo stroma non sono da tenere in considerazione ai fini della classificazione tumorale poichè sono caratterizzate da assenza di ghiandole. Inoltre è importante la segmentazione dei nuclei poichè le cellule epiteliali si raggruppano a ridosso dei dotti prostatici fornendo un'importante informazione sulla posizione delle ghiandole. Per raggiungere l'obiettivo è stato utilizzato un metodo con thresholding adattivo [15]. Sapendo che l'Ematossilina aderisce alle strutture acide come i nuclei e l'Eosina aderisce alle strutture basiche come lo stroma, dalla maschera ottenuta nella separazione degli stain si può ottenere la maschera dei nuclei e dello stroma. In particolare consideriamo l'immagine a toni di grigio con le intensità dei pixel espresse da numeri interi tra 0 e N. L'istogramma è la distribuzione con$N + 1$ classi e rappresenta la frequenza con cui è presente ogni tono di grigio nell'immagine. Considerando una generica classe P dell'istogramma

$(0 \leq P \leq N)$ si può calcolare la curva media pesata progressiva (PWM_{CURVE}:

$$PWM_{CURVE} \frac{\sum_{i=0}^{P} w_i x_i}{\sum_{i=0}^{P} w_i} \tag{2.7}$$

dove w_i è il conto di ogni classe sull'istogramma e x_i è la rispettiva posizione. La PWM_{curve} è considerata per ogni classe come la media ponderata di tutti i valori in scala di grigio dell'istogramma fino a quella classe. Quindi le caratteristiche della distribuzione dei colori possono essere estratte da questa funzione. In particolare se ci sono variazioni importanti di colore da un certo punto dell'istogramma in poi, mi aspetto un cambio di concavità nella PWM_{curve}. Questi punti rappresentano delle potenziali soglie per la segmentazione dei nuclei e dello stroma.

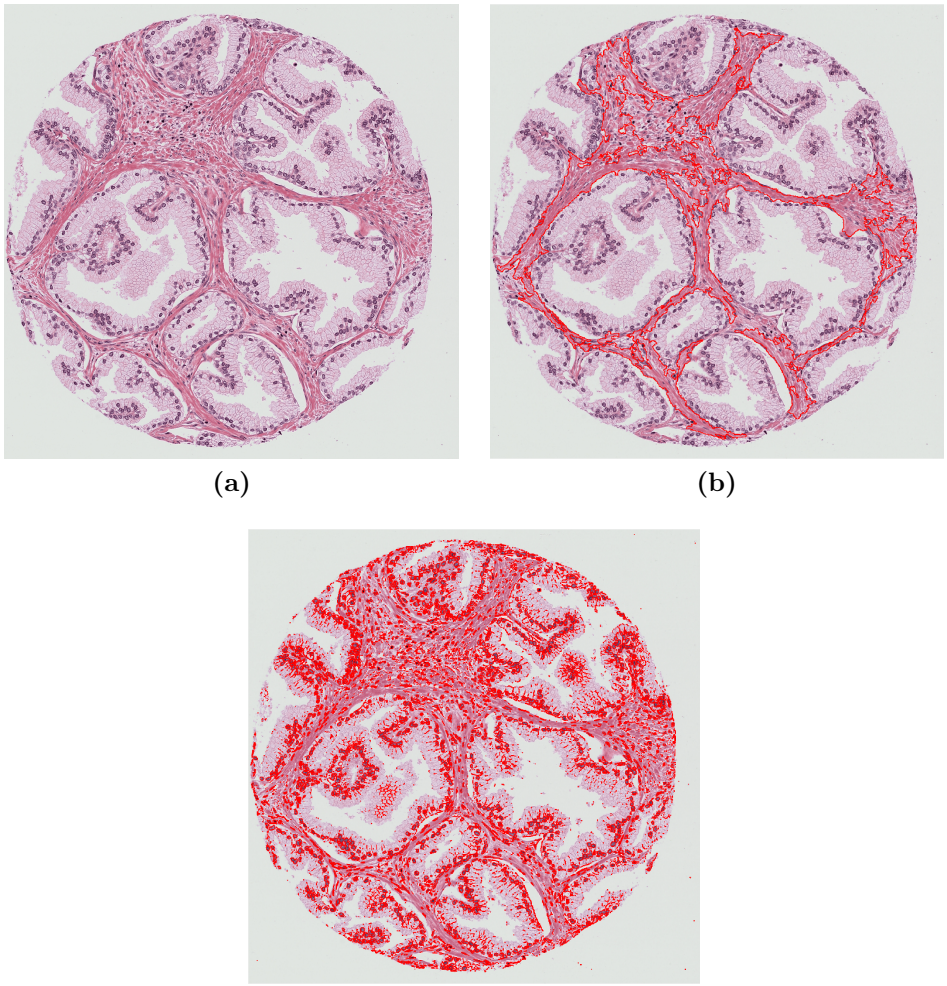
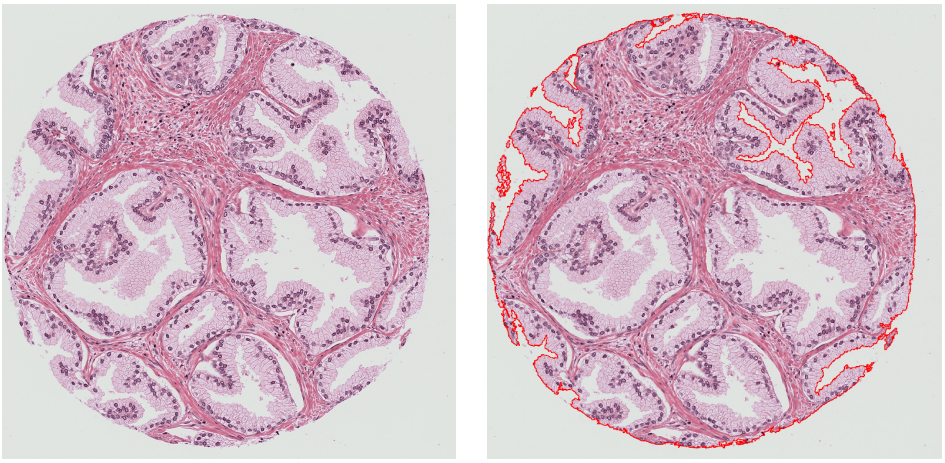


Figura 2.7: a) Immagine originale, b) maschera dello stroma e c) maschera dei nuclei

Per ultimo è stata individuata la porzione di immagine che contiene il tessuto al fine di indirizzare la ricerca delle patch solamente dove è presente la biopsia. Per fare questo, a partire dalla maschera dei bianchi sono stati applicati diversi operatori morfologici di chiusura e rimozione degli oggetti piccoli.



(a) (b)

Figura 2.8: Immagine originale e maschera del tessuto.

2.4 Estrazione smart delle patch

La scelta delle patch da utilizzare per l'addestramento delle reti neurali è un passaggio chiave ai fini dell'obiettivo che si vuole raggiungere. L'apprendimento della rete si basa esclusivamente su ciò che vede nelle immagini fornite e sulla loro classificazione originale quindi risulta fondamentale scartare le patch che potrebbero confondere il classificatore e introdurre ulteriore errore nell'allenamento. Per questo numerosi parametri di selezione sono stati ottimizzati tramite tuning al fine di ottenere le migliori immagini possibili.
L'immagine originale e normalizzata è stata ridimensionata per ottenere un'immagine 1500x1500x3, quindi risoluzione 20x, e successivamente sono state estratte le patch di dimensione 350x350x3. Poichè la scelta del Gleason da assegnare si basa principalmente sull'analisi delle ghiandole, questo compromesso dimensionale tra immagine e patch permette di inquadrare regioni ghiandolari sufficientemente ampie e di conseguenza avere una finestra adatta al riconoscimento del pattern. La patch viene fatta scorrere sull'immagine con un overlap del 25% per evitare di aumentare la ripetibilità e quindi generalizzare il più

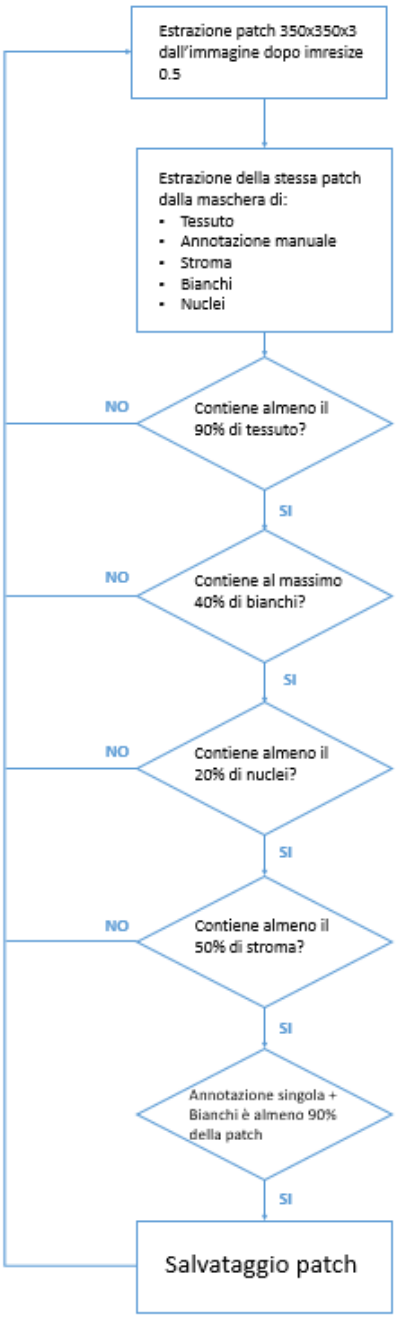


Figura 2.9: Flowchart dell'algoritmo di estrazione intelligente delle patch.

possibile l'apprendimento. Allo stesso tempo, un overlap dello 0% provocherebbe la possibile perdita di informazioni utili ed una riduzione del numero di

patch.
A questo punto la stessa Roi (Region of interest) è viene analizzata nelle maschere estratte dal preprocessing. In particolare vengono considerate solamente le patch nelle quali la maschera del tessuto presenta un'area di almeno il 90% dell'area dell'intera patch. Inoltre l'area delle zone bianche all'interno della patch non deve essere maggiore del 40% dell'area totale. La percentuale è così elevata per assicurarsi di tenere in considerazione le patch sul bordo della ghiandola e quindi con all'interno una discreta porzione di lume. La maschera corrispondente dei nuclei deve avere un'area di almeno il 20% dell'area totale poichè i nuclei nella prostata si raccolgono vicino alle ghiandole e ai dotti prostatici. Il controllo sulla presenza di stroma è stato effettuato utilizzando una percentuale limite del 50%. La scelta è dovuta al fatto che un requisito troppo restrittivo sullo stroma avrebbe scartato molte patch con pattern di tipo 5, molto simili appunto allo stroma fibro-muscolare.
Superati questi requisiti, la patch viene definitivamente salvata se all'interno è presente un'annotazione singola, quindi zona tumorale di un solo tipo oppure benigna. Inoltre la maschera della singola annotazione presente deve occupare almeno il 90% dell'intera patch. Vengono tenute in considerazione anche le patch nelle quali la somma della maschera della singola annotazione e della maschera delle zone bianche raggiunga la percentuale sopra indicata. Questo controllo è stato aggiunto poichè alcune annotazioni in ghiandole particolarmente grosse non comprendono il lume, portando ad un eccessivo scarto di patch potenzialmente molto utili.
Per riconoscere le diverse annotazioni, una volta importata su matlab, la maschera delle annotazioni presenta pixel con valori diversi a seconda dell'annotazione:

- Valore 0: Benigno (colore verde)
- Valore 1: Gleason 3 (colore blu)
- Valore 2: Gleason 4 (colore giallo)

- Valore 3: Gleason 5 (colore rosso)
- Valore 4: Nessuna annotazione (nessun colore)

Inoltre, al fine di rendere il più eterogeneo possibile il training set, è stato sviluppato un algoritmo che controlla la provenienza delle patch e ne seleziona un numero uguale per ogni immagine. Questo controllo viene eseguito distintamente per tutte le classi Benigno, Gleason 3, Gleason 4 e Gleason 5. In questo modo la rete riesce ad osservare pattern appartenenti alla stessa classe ma molto diversi tra loro poichè appartenenti a immagini e pazienti diversi. Questo dovrebbe aiutare ulteriormente la generalizzazione dell'apprendimento.

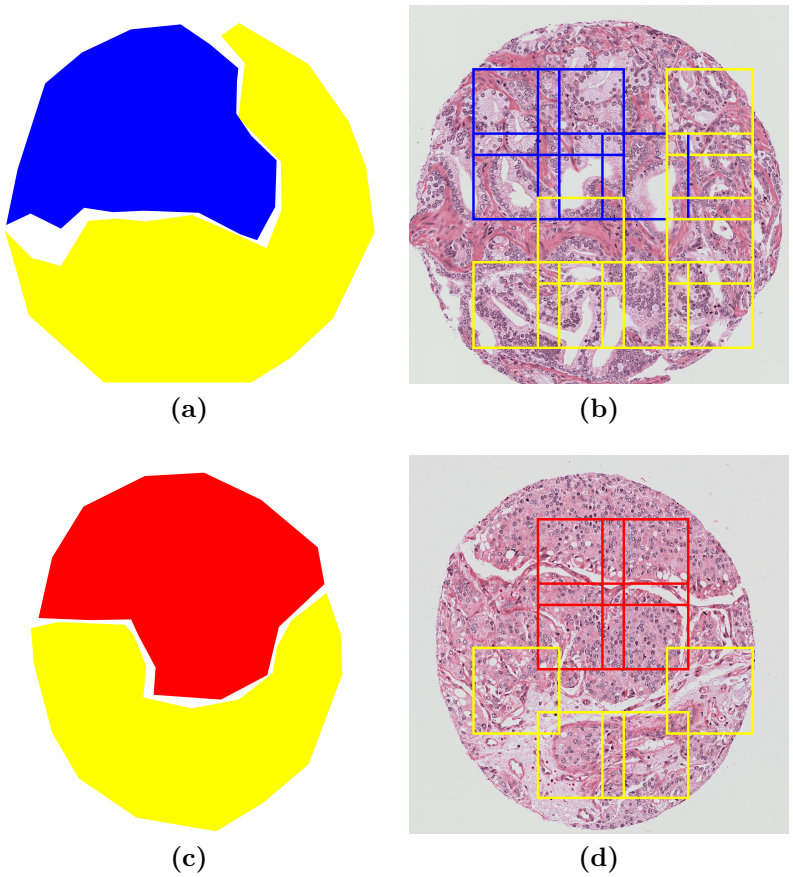


Figura 2.10: Esempi di estrazione intelligente delle patch dalle immagini ZT76-39-B-2-3 (a) e ZT111-4-A-8-9 (c) confrontate con le relative maschere delle annotazioni, rispettivamente (b) e (d).

2.5 Reti Neurali Convoluzionali

Una rete neurale convoluzionale (CNN) è una particolare rete neurale feed-forward, cioè dove le connessioni non formano cicli ma si muovono solo in una direzione. Si ispirano ai processi biologici dei neuroni della corteccia visiva e sono spesso utilizzate nel campo del riconoscimento di immagini e video.
Per tutti gli anni dal 2012, le reti neurali convoluzionali hanno vinto il prestigioso contest internazionale chiamato Imagenet Large Scale Visual Classification Challenge (ILSVRC), gara alla quale partecipano anche giganti come Google e Microsoft [16]. Il loro studio e il loro sviluppo è uno dei temi caldi degli ultimi decenni, periodo nel quale le CNN hanno intrapreso un percorso di crescita importante all'interno del mondo del Machine Learning. Infatti questo tipo di apprendimento automatico è attualmente utilizzato in classificazione di immagini, riconoscimento di modelli (pattern), individuazione di oggetti, segmentazione semantica e numerosi altri compiti.

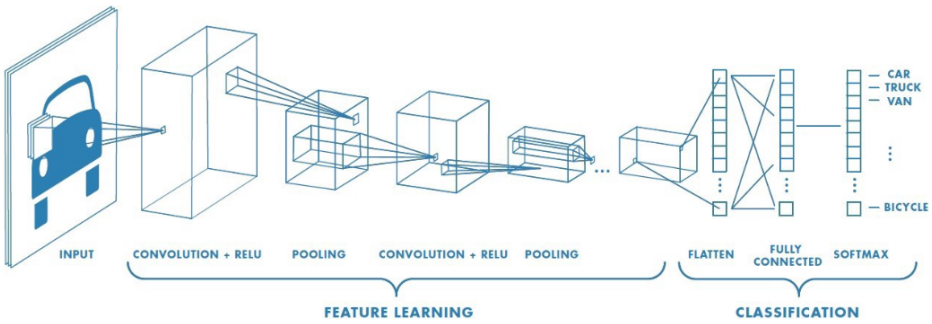


Figura 2.11: Struttura base di una semplice rete neurale convoluzionale [17]

A differenza delle classiche reti neurali, le Cnn lavorano sulle immagini quindi dispongono i neuroni in 3 dimensioni. I neuroni inoltre non sono collegati a tutti i neuroni dello strato precedentemente, ma solamente ad una regione specifica. Il termine convoluzionale indica che la rete svolge ripetute operazioni di convoluzione, ovvero fa scorrere sull'immagine una serie di filtri.

La struttura generale prevede un layer di input, un layer di output e numerosi layer nascosti. Una tipica rete neurale convoluzionale è composta da [18]:

- **Layer convoluzionale**: costituisce la parte principale della rete ed è caratterizzato da un'operazione di convoluzione con un kernel che trasla su tutta la dimensione dell'immagine restituendo un valore di output. Questi kernel sono dei veri e propri filtri, la cui terza dimensione deve ovviamente combaciare con la terza dimensione dell'immagine di input. Ogni filtro restituisce un output 2D e la profondità di questo output dipende dal numero di filtri scelti. L'altezza e la larghezza dipendono dalla scelta del parametro di stride, ovvero il passo del filtro, e dal parametro di padding, ovvero l'eventuale aggiunta di zeri dopo il bordo dell'immagine per controllare la dimensione dell'output.

- **Layer di pooling**: successivo a quello convoluzionale, è uno strato di sottocampionamento. L'obiettivo di questo blocco è ridurre le dimensioni dell'input e le operazioni più comuni sono il max pooling (operazione di ricerca massimo all'interno di una regione) e l'average pooling (operazione di media all'interno di una regione).

- **Layer completamente connesso**: è il blocco di classificazione ed è a tutti gli effetti un vettore di neuroni. In questo caso l'output di ciascun neurone viene mandato a tutti i neuroni del layer successivo.

Come tutte le reti neurali, anche le Cnn sono composte da neuroni che restituiscono un valore di output ottenuto applicando la funzione di attivazione al valore in input. La funzione di attivazione più utilizzata è la RELU (Rectified Linear Unit): $f(x) = \max(0, x)$. Tutti i neuroni sono caratterizzati da pesi(W) e bias(b), parametri da modificare nell'apprendimento. Ogni vettore di pesi e bias rappresenta caratteristiche particolari dell'input, fondamentali per raggiungere l'obiettivo di riconoscimento degli oggetti, delle forme, dei pattern ecc. L'apprendimento consiste nella propagazione all'indietro dell'errore tra

il valore atteso e il valore reale e l'iterativa correzione dei pesi. L'obiettivo è minimizzare la cosiddetta funzione obiettivo (loss function):

$$J(W, b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}_i, y_i) \tag{2.8}$$

dove L è la funzione di perdita tra l'output $\hat{y}_i$ ottenuto dall'i-esimo input e l'output atteso y_i. Nella maggior parte dei casi la funzione obiettivo viene modificata per ridurre il fenomeno dell'overfitting.
Si parla di overfitting quando un modello si adatta solamente ai dati osservati (training set) ma non riesce a generalizzare la sua predizione. Per fare ciò, viene utilizzato il processo di regolarizzazione e in particolare la regolarizzazione di tipo L2 che rappresenta il metodo largamente più utilizzato.

$$J(W, b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}_i, y_i) + \lambda R(f) \tag{2.9}$$

Questo processo comporta l'aggiunta alla funzione di perdita di un parametro di regolarizzazione R controllato da λ che a seconda del suo valore ne conferirà più o meno importanza. Valori bassi di λ tendono ad azzerare il termine di regolarizzazione e a tornare alla funzione di perdita originale. Valori alti di λ invece penalizzano molto la funzione di perdita, portando i pesi vicini allo zero. In questo modo oltre alla riduzione dell'overfitting, viene notevolmente semplificato il modello.
Con lo stesso scopo di semplificare il modello e ridurre l'overfitting, la rete GoogleNet si affida al metodo di Dropout. Questo processo consiste nell'ignorare casualmente alcuni neuroni o alcune connessioni durante l'apprendimento per ridurre la complessità e per evitare che vengano interpellati sempre gli stessi neuroni durante l'apprendimento, contrastando la possibilità di overfitting.

2.5.1 Transfer Learning

Per raggiungere gli obiettivi preposti e in seguito a ripetuti tentativi volti a individuare la rete neurale migliore, è stata utilizzata la rete neurale convoluzionale chiamata GoogLeNet. In particolare si è deciso di usare il metodo del Transfer Learning [19] che consiste nell'utilizzare una Cnn pre-allenata, sfruttando le conoscenze apprese in precedenza, per allenare un nuovo modello. Questo tipo di approccio rende il processo di apprendimento più veloce, più accurato e necessita di un numero decisamente ridotto di dati per l'allenamento.

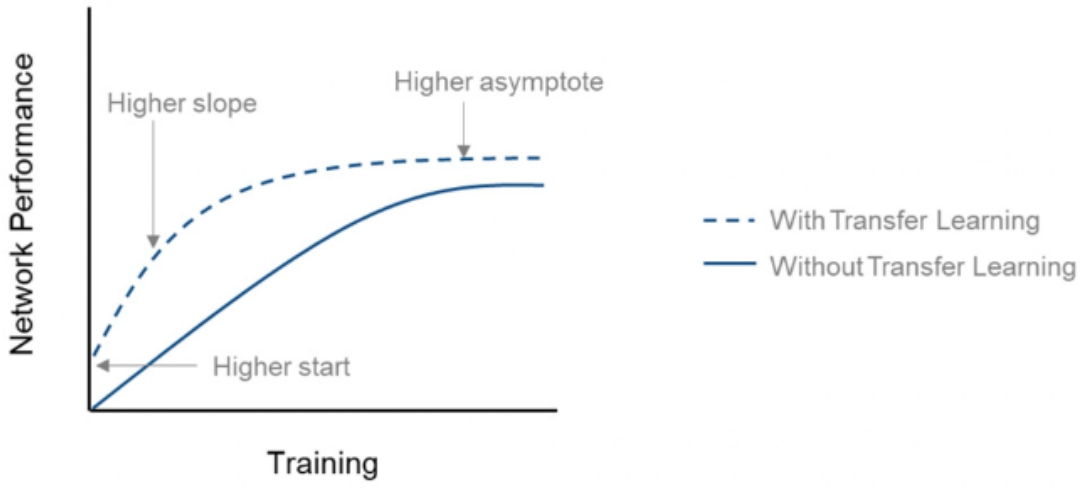


Figura 2.12: Prestazione della rete allenandola da zero e utilizzando il transfer learning [20]

L'allenamento da zero funziona bene per compiti molto specifici e per i quali allenamenti precedenti non risultano utili. D'altra parte, per avere risultati soddisfacenti e accurati, ci vorrebbe una mole di dati significativa.
L'approccio del transfer learning consiste nell'utilizzare gli stessi strati della rete pre-allenata, indipendenti dal problema di classificazione, sostituendo gli ultimi layer specifici per il problema in questione. In particolare bisogna so-

stituire il classificatore con un nuovo layer di classificazione in cui indicare il numero di classi che devono essere riconosciute.

2.5.2 GoogLeNet

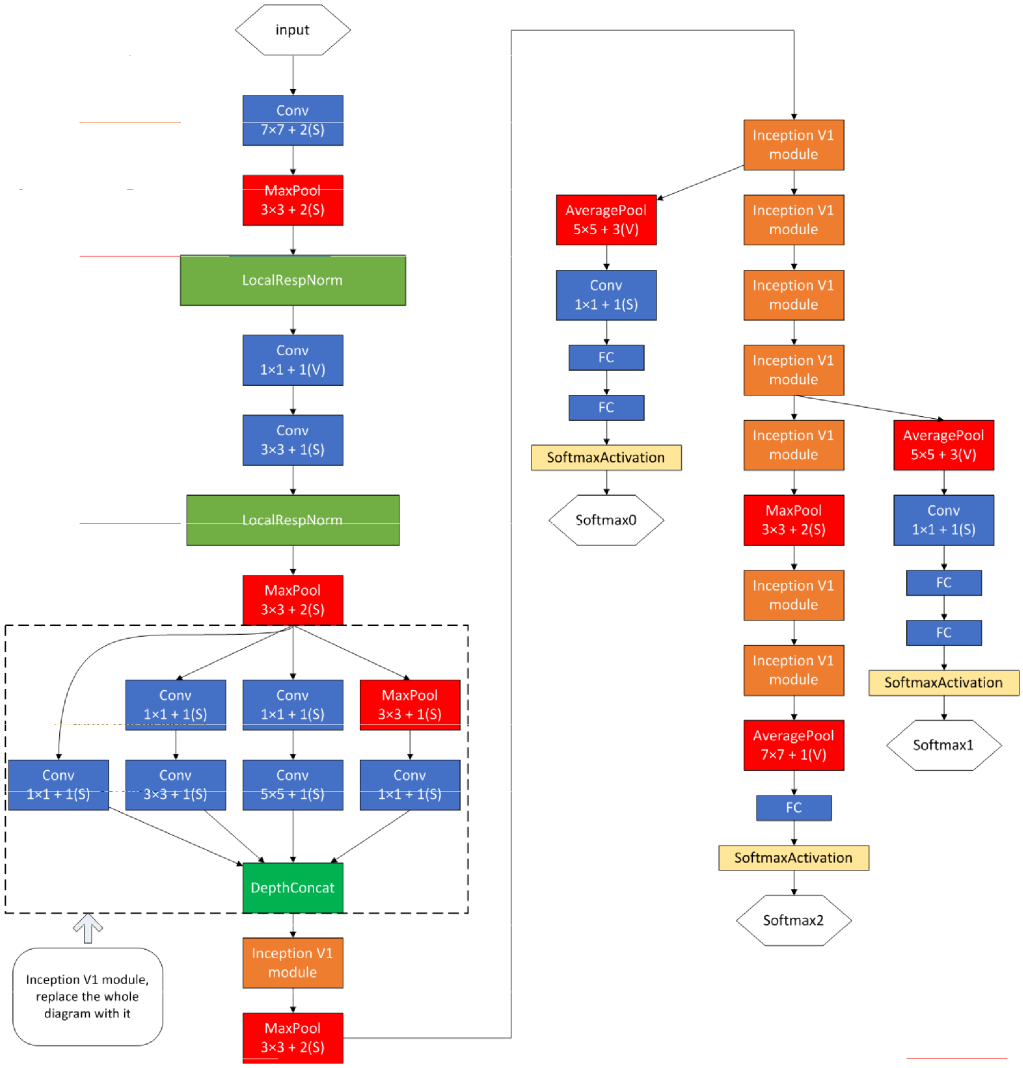


Figura 2.13: Struttura della rete GoogLeNet [21]

La GoogLeNet è una rete che nel 2014 ha vinto la competizione ILSVRC sulla classificazione di immagini e si ispira al modello Network in Network [22]. E' una rete preallenata su più di un milione di immagini, in grado di ricono-

scere 1000 classi ed è caratterizzata da 22 layer e circa 7 milioni di parametri. Introduce il modulo Inception v1 [23] che permette al modello di descrivere meglio il contenuto dei dati di input aumentando ulteriormente la profondità e la larghezza del modello di rete [21].
Una delle funzioni del layer convoluzionale è di ridurre o aumentare le dimensioni utilizzando il numero di filtri. Il modulo Inception v1 gestisce la dimensione introducendo e utilizzando principalmente il kernel convoluzionale 1x1, che permette di ridurre il numero di features senza modificare la dimensione dell'immagine. Il modello utilizza anche kernel di tipo 3x3 e 5x5. Questo permette quindi di ridurre notevolmente il numero di parametri e la dimensione del modello, aiutando a prevenire il problema dell'overfitting.
Un numero eccessivo di parametri in relazione ai dati osservati è sicuramente una causa dell'overfitting.
Questa rete prevede una funzione finale di softmax che consente di trasformare i valori di output del livello precedente in probabilità delle classi
E' stato utilizzato il modello preallenato fornito da MATLAB© 2019a e sono stati mantenuti tutti i layer. E' stato modificato il layer completamente connesso indicando il numero di classi che devono essere riconosciute durante lo specifico allenamento. Inoltre è stato modificato il parametro chiamato InputSize del primo layer della rete, ovvero quello di input, per poter cambiare le dimensioni delle patch di ingresso. Infatti la GoogleNet, per svolgere l'allenamento, di default accetta patch di dimensioni 224x224x3 ma per il raggiungimento dell'obiettivo sono state estratte patch di dimensione 350x350x3.

2.6 Allenamento delle CNN

Una volta estratte le patch con il metodo visto in precedenza, il problema di classificazione delle diverse classi è stato affrontato sviluppando tre metodi di apprendimento differenti. Tutti e tre sfruttano la GoogLeNet, che viene alle-

nata con dati diversi a seconda dell'obiettivo che si vuole raggiungere. Infatti, a partire dall'insieme di tutte le patch, ne sono state considerate 800 per ogni classe (Benigna, Gleason 3, Gleason 4 e Gleason 5) divise in maniera equa con almeno una patch per ognuna delle 641 immagini di training. Da questo macrogruppo di 3200 patch utilizzato nel primo approccio, sono stati poi creati i diversi sottogruppi necessari agli altri approcci.
Le scelte della dimensione del mini-batch e del learning rate rappresentano più di tutti le criticità dell'allenamento di una rete neurale convoluzionale. Per poter allenare la rete con un subset del training set robusto, la dimensione del mini-batch è stata scelta circa il 5% del totale.
Il learning rate influenza la convergenza dell'ottimizzazione e governa i pesi nella direzione opposta ai gradienti [24]. Un learning rate troppo basso rende l'allenamento più affidabile ma molto più lento, con la possibilità concreta di finire in un minimo locale della funzione di perdita. Un learning rate troppo alto potrebbe far divergere l'allenamento. Per questo solitamente si inizia con un learning rate alto per poi diminuirlo periodicamente.
In tutte le reti allenate, a causa della numerosità ridotta del training set, è stato applicato un aumento dei dati tramite diverse operazioni. In particolare, alle patch sono state applicate:

- Riflessione rispetto all'asse x
- Riflessione rispetto all'asse y
- Traslazione lungo l'asse x in un range di pixel compreso tra -25 e +25
- Traslazione lungo l'asse y in un range di pixel compreso tra -25 e +25
- Rotazione casuale in un range compreso tra -180° e +180 °

2.6.1 Primo Approccio: 4 classi distinte

Poichè lo scopo del lavoro è la realizzazione di un sistema automatico per l'assegnazione dello score di Gleason, si è partiti realizzando una rete con l'obiettivo di distinguere le 4 classi separatamente. Per questo, 800 patch per classe sono state utilizzate per l'allenamento della GoogLeNet. In particolare il 95% sono state selezionate come training set e il 5% come validation set. Il numero massimo di epoche di allenamento della rete è stato impostato a 50 e una dimensione del mini-batch di 128.
Al fine di ridurre il più possibile la variabilità, assicurandosi che il modello rimanga più generale possibile riducendo dunque la possibilità di overfitting, i dati di training vengono rimescolati prima di ogni epoca e i dati di validazione con frequenza data dal parametro di validation frequency.
Dopo un tuning sui parametri, è stato scelto un learning rate iniziale di 0.01 che ogni 10 epoche viene ridotto di un fattore moltiplicativo di 0.2. La validation frequency, ovvero l'intervallo con il quale vengono calcolati accuratezza e funzione obiettivo sul validation set, è stata impostata a 20 iterazioni. La funzione obiettivo (loss function) è una funzione della differenza tra valore atteso e valore reale e deve essere minimizzata. Il validation patience è stato impostato a 20. Questo parametro rappresenta il numero di volte in cui la funzione obiettivo nel validation set può essere maggiore o uguale al precedente valore minimo della funzione stessa prima che si arresti l'allenamento. Come algoritmo di ottimizzazione è stato utilizzato la discesa stocastica del gradiente con momento (sgdm) che aiuta ed accelerare i gradienti verso la corretta direzione. Il parametro di regolarizzazione è stato lasciato al valore di default di 0.0001.

PARAMETRI DELLA RETE	
Algoritmo di ottimizzazione	SGDM
Momento	0,9
Learning rate iniziale	0,01
Fattore di drop learning rate	0,2
Periodo di drop learning rate	10
L2 regolarizzazione	0,0001
Max epoche	50
Dimensione mini-batch	128
Frequenza validazione	20
Validation patience	20
Rimescolamento dati	ogni epoca

Tabella 2.1: Parametri della rete a 4 classi.

2.6.2 Secondo Approccio: 3 classi (Gleason 3 e Gleason 4 fuse)

Dal primo approccio si evince come i maggiori problemi siano nella distinzione tra il pattern 3 e il pattern 4, mentre i pattern di tipo benigno e 5 vengono ben riconosciuti dal classificatore. In generale il problema di riproducibilità e di variabilità tra gli operatori è molto più comune tra il gleason 3 e il gleason 4 [25].

La differenza sostanziale tra i due pattern dovrebbe essere nella forma e dimensione delle ghiandole. Infatti nel pattern di tipo 3 le ghiandole iniziano a perdere la loro forma arrotondata e le loro dimensioni standard, ma ancora sono differenziate e si possono distinguere tra loro. Nel pattern di tipo 4 invece le ghiandole non sono più riconoscibili e separabili ed iniziano a fondersi tra loro.

I problemi si presentano quando vengono analizzati pattern dubbi, al limite tra il pattern 3 e il pattern 4. In questi casi molti operatori si trovano in disaccordo tra loro.

Per cercare di risolvere questa situazione, si è pensato di passare ad un approccio a 3 classi con le classi gleason 3 e gleason 4 fuse insieme al fine di massimizzare il riconoscimento tra il pattern benigno, la classe fusa e il pattern di tipo 5. In seguito è stata allenata una rete per riconoscere solamente i pattern 3 e 4 ed è stata legata in cascata alla prima per avere nuovamente le 4 classi distinte.

GoogleNet: Benigno vs Gleason 3-4 vs Gleason 5

Per l'allenamento della seguente rete, sono state create 3 classi contenenti 800 patch ciascuna. In particolare, la classe Benigno e la classe Gleason 5 contengono le stesse 800 patch utilizzate nella rete a 4 classi del primo approccio. La classe fusa Gleason 3-4 è composta da 400 patch di pattern 3 e 400 patch di pattern 4 estratte dalle classi della rete del primo approccio, cercando sempre di avere almeno una patch per ogni immagine diversa del training set. Dalla totalità di 2400, il 90% sono state utilizzate come training set il 10% come validation set. Il numero massimo di epoche è stato impostato a 50 e la dimensione del mini-batch a 64. Anche in questo caso il rimescolamento dei dati avviene ad ogni epoca. Come per la rete a 4 classi, il learning rate iniziale è stato impostato a 0.1 con un decadimento di un fattore 0.2 ogni 10 epoche. Il parametro di validation frequency è stato impostato a 15 e il validation patience a 10. Anche qui il parametro di regolarizzazione è stato lasciato al valore di default di 0.0004.

PARAMETRI DELLA RETE	
Algoritmo di ottimizzazione	SGDM
Momento	0,9
Learning rate iniziale	0,01
Fattore di drop learning rate	0,2
Periodo di drop learning rate	10
L2 regolarizzazione	0,0001
Max epoche	50
Dimensione mini-batch	64
Frequenza validazione	15
Validation patience	10
Rimescolamento dati	ogni epoca

Tabella 2.2: Parametri della rete a 3 classi.

GoogleNet: Gleason 3 vs Gleason 4

Dopo avere concluso la rete che distingue le classi benigno, pattern 3 e 4 fusi e pattern 5, in cascata viene allenata una rete per distinguere il pattern 3 dal pattern 4. Per fare questo, le stesse patch di tipo 3 e 4 utilizzate nella rete precedente vengono divise nuovamente nella loro classe originale per essere riutilizzate come training set. Quindi da queste due classi composte da 400 elementi ciascuna, il 90% viene usato come training set e il 10% come validation set. Anche qui il numero massimo di epoche è stato impostato a 50 e la dimensione del mini-batch a 32 a causa della ridotta numerosità del training set. Il rimescolamento viene eseguito ogni epoca. Il learning rate iniziale è stato impostato a 0.005 correndo il rischio di finire in un minimo locale, ma rendendo più sicuro l'allenamento, vista la difficoltà della distinzione delle due classi in questione. La validation frequency è stata impostata a 10 e il validation patience ulteriormente ridotto a 8 per evitare l'overfitting. Il parametro di regolarizzazione nuovamente mantenuto al valore di default.

PARAMETRI DELLA RETE	
Algoritmo di ottimizzazione	SGDM
Momento	0,9
Learning rate iniziale	0,005
Fattore di drop learning rate	0,2
Periodo di drop learning rate	10
L2 regolarizzazione	0,0001
Max epoche	50
Dimensione mini-batch	32
Frequenza validazione	10
Validation patience	8
Rimescolamento dati	ogni epoca

Tabella 2.3: Parametri della rete G3 vs G4.

2.6.3 Terzo Approccio: 2 classi (Gleason 3, Gleason 4 e Gleason 5 fuse)

I pattern tumorali differiscono tra loro principalmente per la struttura ghiandolare che tende a perdere la propria forma e dimensione standard fino ad arrivare alla fusione e alla completa scomparsa delle ghiandole. Spesso risulta difficile assegnare una classe ad una zona tumorale perchè presenta caratteristiche vicine a due classi. Molti tumori considerati di tipo 5 sono molto vicini al pattern 4 e lo stesso avviene tra il pattern 3 e il pattern 4. Al contrario, il pattern benigno ha caratteristiche che si discostano nettamente dai pattern tumorali, tranne in casi di Gleason 3 al limite quindi molto vicini al pattern sano.

Si è pensato quindi di sviluppare un diverso approccio, in grado di distinguere le 4 diverse classi, che si basa nuovamente su due reti neurali convoluzionali in cascata. Adesso però la prima rete si occupa di distinguere il pattern sano da quello tumorale. La seconda rete, in cascata alla prima, è stata allenata con patch di Gleason 3, Gleason 4 e Gleason 5 per massimizzare il riconoscimento

tra i pattern tumorali.

GoogleNet: Tumore vs Non Tumore

Per l'allenamento di questa rete, sono state considerate tutte le patch di tipo benigno estratte dalle immagini di training estraendone 830, sempre mantenendo le proporzioni tra le diverse immagini. Queste patch sono andate a comporre la classe non tumorale. Per la creazione della classe tumorale, 277 campioni dal Gleason 3, Gleason 4 e Gleason 5 sono stati estratti dalle rispettivi classi. Il 90% delle patch è stato selezionato come training set mentre il restante 10% come validation set. In questo caso il learning rate iniziale è stato ristabilito a 0.01 con un decadimento di un fattore 0.2 ogni 10 epoche e la dimensione del mini-batch a 64. Il numero massimo di epoche di allenamento rimane 50 con una validation frequency di 20 e un validation patience di 10. Come sempre il rimescolamento dei dati avviene ogni epoca e il parametro di default è stato lasciato al suo valore di default.

PARAMETRI DELLA RETE	
Algoritmo di ottimizzazione	SGDM
Momento	0,9
Learning rate iniziale	0,01
Fattore di drop learning rate	0,2
Periodo di drop learning rate	10
L2 regolarizzazione	0,0001
Max epoche	50
Dimensione mini-batch	64
Frequenza validazione	20
Validation patience	10
Rimescolamento dati	ogni epoca

Tabella 2.4: Parametri della rete Tumore vs Non Tumore.

GoogleNet: Gleason 3 vs Gleason 4 vs Gleason 5

La seguente rete sarà legata in cascata alla rete che distingue i pattern tumorali da quelli non tumorali. Per questo sono state utilizzate le stesse patch tumorali della rete precedente che sono state opportunamente redistribuite nelle loro classi originali. A causa del limitato numero di patch, il 75% della totalità delle patch è stato impostato come training set e il 25% come validation set. La dimensione del mini-batch è stata impostata a 32 con un learning rate iniziale di 0.005 e un fattore di decadimento di 0.5 ogni 10 epoche. I parametri di validation frequency e di validation patience non sono stati modificati. Anche in questo allenamento il rimescolamento dei dati avviene ogni epoca.

PARAMETRI DELLA RETE	
Algoritmo di ottimizzazione	SGDM
Momento	0,9
Learning rate iniziale	0,005
Fattore di drop learning rate	0,5
Periodo di drop learning rate	10
L2 regolarizzazione	0,0001
Max epoche	50
Dimensione mini-batch	32
Frequenza validazione	20
Validation patience	10
Rimescolamento dati	ogni epoca

Tabella 2.5: Parametri della rete G3 vs G4 vs G5.

2.7 Metodi di confronto e validazione

Una volta completati i tre approcci descritti per la realizzazione di un classificatore in grado di distinguere lo score di Gleason delle diverse zone tumorali, è stato operato un confronto sulle patch del training set al fine di individuare il metodo più robusto. Per confrontare i risultati dei 3 diversi classificatori,

sono state valutate le confusion matrix [26].
La confusion matrix è la misura ideale delle performance per un problema di classificazione con l'utilizzo del machine learning. L'output può essere rappresentato da 2 classi (classificatore binario) o più come nel nostro caso. La classificazione finale dell'algoritmo viene messa a confronto con il ground truth (riferimento), che per questo lavoro è rappresentato dalle annotazioni di un patologo esperto.

OUTPUT	Benigno	CORRECT			
	Gleason 3		CORRECT		
	Gleason 4			CORRECT	
	Gleason 5				CORRECT
		Benigno	Gleason 3	Gleason 4	Gleason 5
		TARGET			

Tabella 2.6: Confusion matrix a 4 classi.

Sulle colonne della confusion matrix si trova le classi reali (target) mentre sulle righe le classi predette dal classificatore (output). Nel nostro caso troviamo 3 tipi di elementi:

- **Elementi corretti classificati**: sono gli elementi che sia il classificatore che il ground truth hanno riconosciuto nella stessa classe. Si trovano sulla diagonale principale della matrice.
- **Elementi sottostimati**: sono gli elementi che il classificatore individua in una classe tumorale inferiore rispetto al ground truth. Si trovano a destra rispetto alla diagonale principale (Rosso).
- **Elementi sovrastimati**: sono gli elementi che il classificatore individua in una classe tumorale superiore rispetto al ground truth. Si trovano a sinistra rispetto alla diagonale principale (Verde).

Le confusion matrix sulle patch del training set sono state valutate utilizzando il parametro di accuratezza:

$$Acc = \frac{\sum corretti\, classificati}{Totale\, elementi} \tag{2.10}$$

ovvero la somma degli elementi sulla diagonale principale diviso il totale degli elementi all'interno della matrice. Fornisce una misura diretta della bontà del classificatore ed è un parametro facilmente confrontabile tra le diverse confusion matrix.
Sono state valutate inoltre le percentuali relative di elementi sottostimati ed elementi sovrastimati. Trattandosi di un sistema di assegnazione dello score di Gleason e di conseguenza un sistema di prognosi, si preferisce un classificatore che sovrastimi piuttosto che sottostimi, per evitare di indicare come benigni pattern potenzialmente tumorali.

$$\%sottostimati = \frac{\sum Elementi\, sottostimati}{Totale\, elementi} \tag{2.11}$$

$$\%sovrastimati = \frac{\sum Elementi\, sovrastimati}{Totale\, elementi} \tag{2.12}$$

Una volta scelto l'approccio che fornisce un classificatore più robusto e coerente con l'obiettivo del lavoro, è stato testato sulle 245 immagini del TMA80, designato come test set. E' stato impostato un algoritmo che data l'immagine di test, restituisce le mappe di probabilità per ogni classe Benigno, Gleason 3, Gleason 4 e Gleason 5. Per ogni immagine vengono classificate le patch 350x350x3 estratte con i criteri scelti nell'algoritmo di estrazione automatica. Successivamente, per ogni pixel sono state contate le patch classificate in cui è contenuto e gli è stato assegnato un valore per ogni classe. Questi 4 valori rappresentano le medie delle probabilità di appartenenza alle 4 classi di tutte le patch contenenti il pixel.
Da queste mappe di probabilità è stato calcolato lo score di Gleason globale

dell'intera immagine individuando il pattern principale, ovvero il più rappresentato, e il pattern secondario, ovvero il pattern che, oltre ad essere il secondo più rappresentato, raggiunge anche una determinata soglia minima. Lo score di Gleason totale è poi ottenuto sommando gli score dei due pattern trovati o raddoppiando lo score nel caso di pattern tumorale singolo. Se il classificatore non individua pattern tumorali, l'immagine viene classificata come benigna.
Il risultato del classificatore è stato confrontato con la classificazione fornita da entrambi i patologi che hanno annotato le immagini di test. Sono state messe a confronto le confusion matrix:

- Classificatore vs Patologo 1
- Classificatore vs Patologo 2
- Patologo 1 vs Patologo 2

sia sul calcolo dello score di Gleason totale sia sulla ricerca del pattern principale e del pattern secondario e sono state valutate in termini di accuratezza e percentuale di sottostimati e sovrastimati.

Capitolo 3

Risultati

3.1 Validazione sul training set per la scelta del miglior classificatore

Per la scelta del miglior classificatore sono state valutate le confusion matrix finali sul training set di tutti e 3 gli approcci sviluppati. Oltre a ricercare l'accuratezza più elevata, il classificatore migliore è stato scelto anche in base alla percentuale di elementi sottostimati. E' molto importante che il sistema automatico non sottovaluti un problema, come il cancro prostatico in questo caso, ma che piuttosto sopravvaluti ponendo comunque all'attenzione dei patologi la biopsia dubbia.
Dunque una volta allenate le reti, sono state applicate alle patch del training set per valutare il comportamento del classificatore rispetto alla loro classificazione vera data dall'annotazione del patologo.

La confusion matrix del primo classificatore allenato con le patch di 4 classi distinte mostra un'accuratezza totale del 83.5%. Come previsto le patch di classe benigna e di classe gleason 5 vengono facilmente individuate dal classificatore (rispettivamente 98.8% e 93.1% di corretti classificati) mentre i problemi si presentano nella distinzione tra il gleason 3 e il gleason 4, in

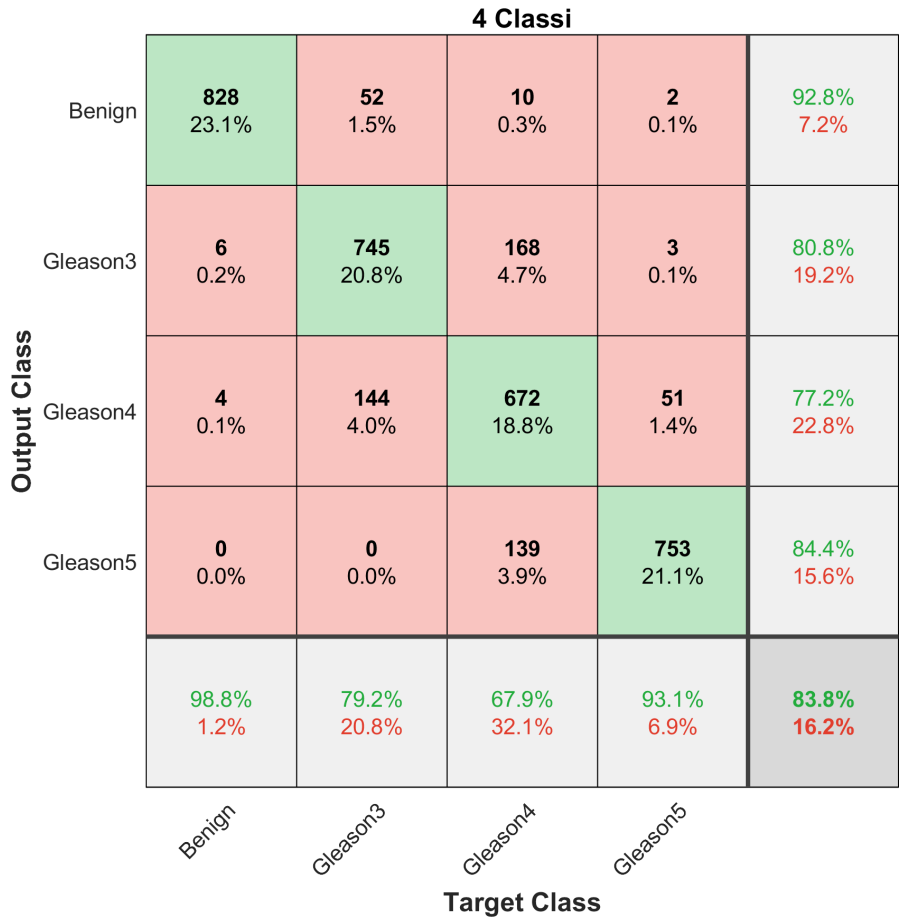


Tabella 3.1: Confusion matrix 1° approccio: 4 classi distinte.

maniera sbilanciata verso quest'ultima classe dove la percentuale di corretti classificati si ferma al 67.9%. La percentuale delle patch sottostimate è 8.1%, la maggior parte di classe gleason 4 che sono state classificate come gleason 3. L'andamento globale della confusion matrix rimane comunque positivo poichè a partire dalla classe di riferimento gli errori diminuiscono progressivamente andando via via verso le classi più distanti.

A questo punto, visti i problemi nella classificazione delle patch di gleason 3 e gleason 4 è stato sviluppato un classificatore che inizialmente concentrasse i suoi sforzi nel distinguere la classe benigna e la classe gleason 5 dalle classi gleason 3 e 4. In seguito un altro classificatore in cascata che si occupasse di

discriminare le classi gleason 3 e 4.

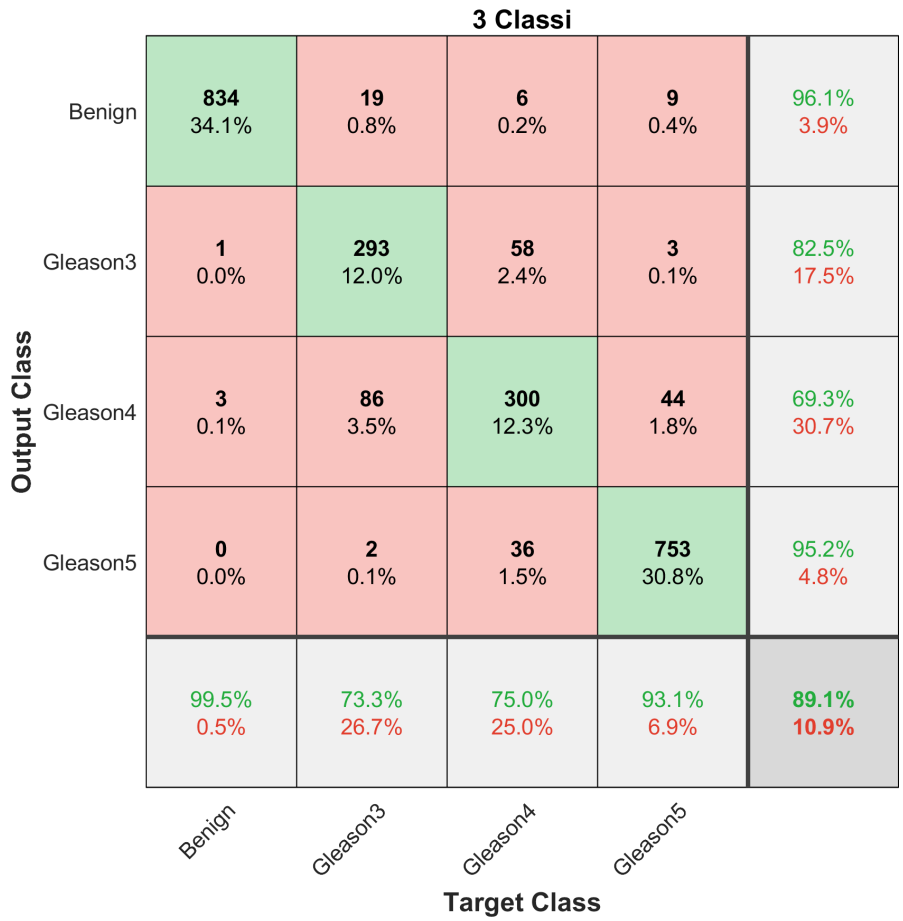


Tabella 3.2: Confusion matrix 2° approccio: 3 classi

L'accuratezza globale raggiunge l'89.1%. Nuovamente le percentuali di corretti classificati per la classe benigna e la classe gleason 5 raggiungono valori elevati. Rimangono le difficoltà nelle classi gleason 3 e gleason 4 ma le percentuali di corretti classificati sono più bilanciate per le due classi (73.3% per gleason 3 e 75.0% per il gleason 4). Gran parte dei degli errori di classificazione appartengono a queste due classi e la percentuale di sottostimati è scesa a 5.7%.

Per cercare di minimizzare ancora di più gli errori tra le classi tumorali gleason 3, gleason 4 e gleason 5 è stato sviluppato il terzo classificatore che distingue

inizialmente le patch tumorali da quelle non tumorali e successivamente è stato allenato per distinguere il gleason 3, il gleason 4 e il gleason 5 con una rete neurale in cascata alla prima.

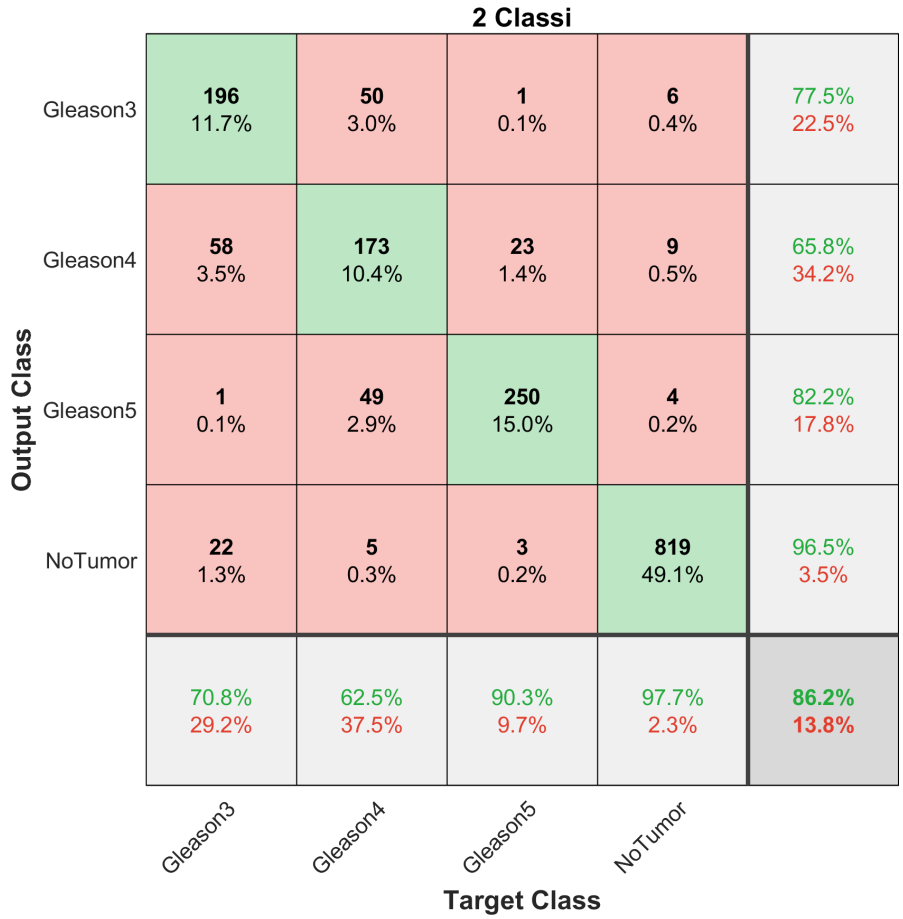


Tabella 3.3: Confusion matrix 3° approccio: 2 classi

Con questo approccio non si apprezzano miglioramenti rispetto ai due classificatori precedenti. Nonostante l'accuratezza complessiva sia 86.2%, le classi gleason 3 e gleason 4 hanno una percentuale di corretti classificati (70.8% e 62.5%) minore rispetto ad entrambi gli altri 2 approcci.
Tutte e tre i classificatori mostrano andamenti simili. Gli errori si distribuiscono soprattutto nelle due classi gleason 3 e gleason 4 e in generale gli elementi con classificazione errata rimangono nelle classi immediatamente adiacenti alla

classe reale.
Il sistema sviluppato tramite il 2° approccio, caratterizzato da due reti in cascata per ottimizzare il riconoscimento tra le classi Gleason 3 e Gleason 4 è stato scelto come classificatore più robusto e quindi da utilizzare per la verifica sulle immagini di test. Infatti oltre a presentare un'accuratezza migliore, mostra anche la minor percentuale di elementi sottostimati.

3.2 Validazione sulle immagini di test

Scelto il miglior classificatore, è stato testato sulle 245 immagini del TMA80 scelte come test set per la numerosità e l'eterogeneità.
Sono state valutate le mappe di probabilità risultanti dall'applicazione del classificatore sulle immagini. Questo tipo di visualizzazione permette un primo rapido confronto tra l'annotazione dell'algoritmo automatico e le annotazioni dei due patologi. Infatti le immagini utilizzate come training set sono state annotate da un solo patologo, mentre le immagini di test sono state annotate anche da un secondo patologo per poter valutare la variabilità inter-operatore e la soggettività che contraddistingue lo score di Gleason.

La figura 3.1 mostra le mappe di probabilità di 4 immagini scelte con lo scopo di fornire un esempio rappresentativo di tutto il dataset. Ogni sottofigura è composta dalle annotazioni dei due patologi nella colonna più a sinistra (blu: Gleason 3, giallo: Gleason 4, rosso: Gleason 5) e dalle mappe di probabilità di ogni Gleason estratte dalla classificazione dell'algoritmo sviluppato.
Su 3 immagini l'algoritmo non riesce ad estrarre sufficienti patch per riuscire ad ottenere mappe di probabilità indicative e per questo sono state escluse dai successivi calcoli dello score di Gleason finale dell'immagine. Queste immagini sono caratterizzate da una quantità di stroma o di zone senza tessuto troppo elevata a tal punto da non soddisfare mai i requisiti sull'estrazione delle patch. Di conseguenza le immagini di test sulle quali viene calcolato lo score di Glea-

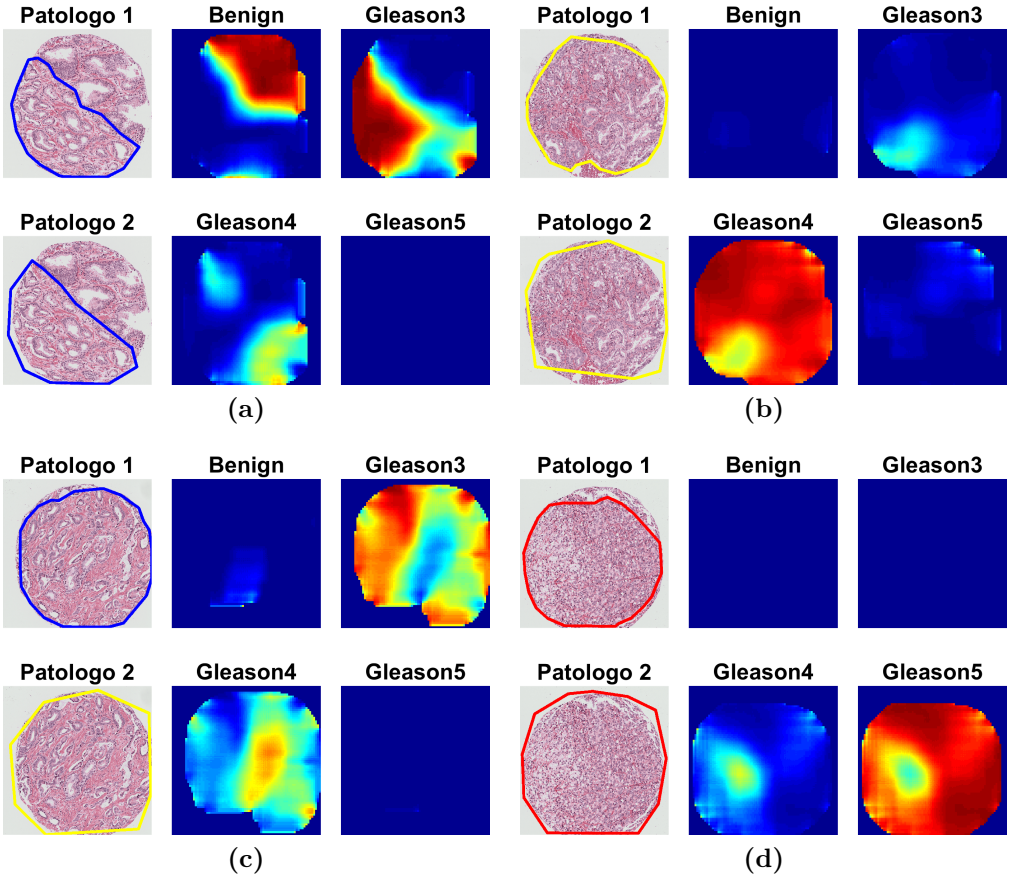


Figura 3.1: Esempi di mappe di probabilità comparate con le annotazioni dei due patologi.

son globale sono 242.

Questi 4 esempi di mappe di probabilità sulle immagini di test, che hanno lo scopo di fornire una prima valutazione visiva immediata, confermano l'andamento valutato nelle confusion matrix sul training set. Il TMA 80 nell'esempio a) presenta una zona benigna che viene ben riconosciuta dal modello e il pattern 3 viene quasi totalmente individuato a meno di una zona con una probabilità leggermente più alta verso il pattern 4. Nel TMA 80 nell'esempio b) entrambi i patologi hanno annotato un Gleason 4 e il modello trova bene questo pattern con una probabilità vicina all'1 a meno di una zona dove la probabilità è leggermente più bassa ma sempre verso il Gleason

4. Stesso discorso si può fare per il TMA 80 nell'esempio d) dove il modello individua molto bene il pattern 5 tranne in una zona dove riconosce il pattern 4. Il TMA 80 nell'esempio c) presenta un'annotazione diversa tra i due patologi e questa variabilità viene trasmessa anche al classificatore che individua alcune zone come Gleason 3 e altre come Gleason 4.
Osservando le mappe di probabilità, il classificatore implementato sembra rispondere bene e soprattutto rispecchia il comportamento visto sulla confusion matrix del training set. Questo significa che il sistema è riuscito a generalizzare il problema di classificazione in maniera discreta.

3.2.1 Calcolo dello score di Gleason globale

A partire dalle mappe di probabilità, per avere una valutazione quantitativa della bontà del classificatore, è stato calcolato lo score di Gleason sull'intera immagine, sommando i due pattern più presenti, ed è stato confrontato con le annotazioni dei due patologi.

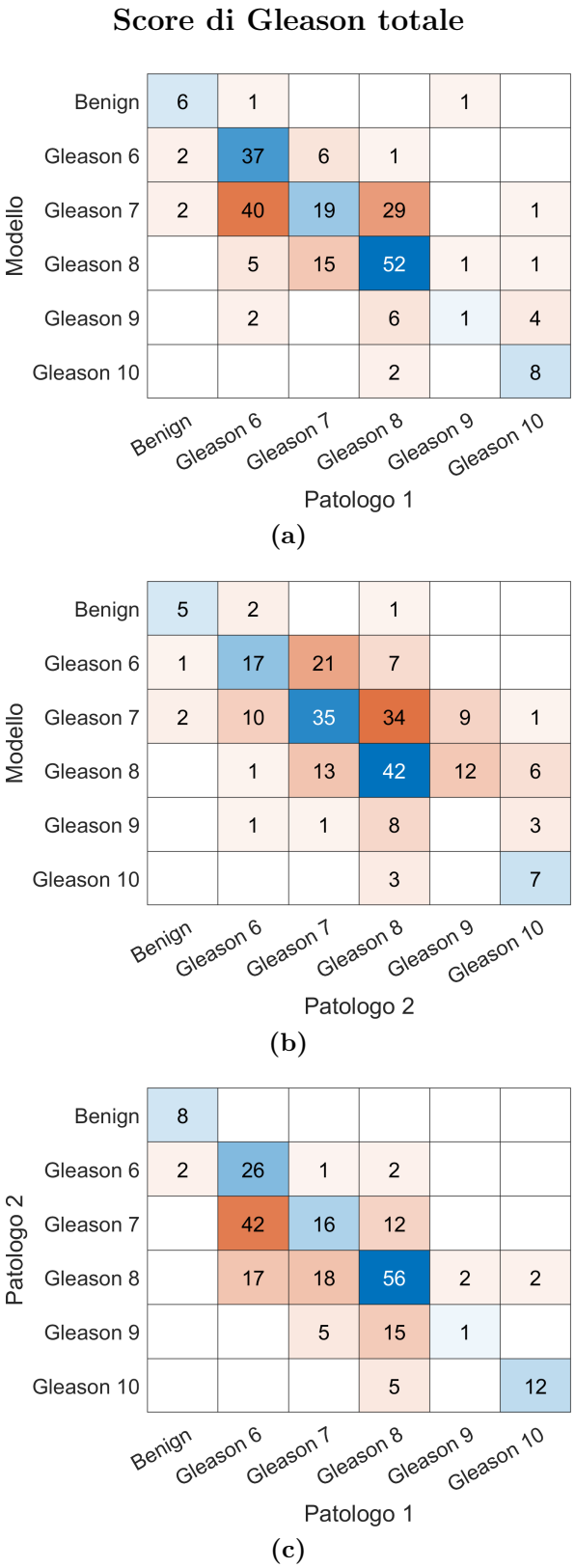


Figura 3.2: Confusion matrix sullo score di Gleason globale: a) modello vs patologo 1, b) modello vs patologo 2, c)patologo 2 vs patologo 1.

Il classificatore sviluppato presenta un'accuratezza del 50.83% rispetto al patologo 1 e del 43.80% rispetto al patologo 2. Tra i due patologi l'accuratezza è del 49.17%. Le prestazioni del classificatore rispetto al patologo 1 sono comparabili al consenso tra i due patologi. Rispetto al patologo 2 le prestazioni sono leggermente inferiori e questo comportamento poteva essere previsto per il fatto che il classificatore è stato allenato su immagini annotate dal patologo 1. Analizzando la confusion matrix del modello rispetto al patologo 1 si nota che il 18.59% delle immagini vengono sottostimate ma di queste immagini il 64.44% sono Gleason 8 che sono stati classificati come Gleason 7. Questo mette nuovamente in luce le problematiche nello standardizzare il riconoscimento tra il pattern 3 e il pattern 4. Solamente 2 immagini tumorali vengono classificate come benigne, di cui 1 è un Gleason 6 con un pattern molto vicino al pattern benigno. Il Gleason 8 classificato come Benigno invece è un errore grave ma rimane un caso isolato che sicuramente andrebbe sottoposto all'attenzione di un patologo.
In generale in tutte le confusion matrix gli errori decrescono rapidamente allontanandosi dalla classe di riferimento e questo comportamento crea una confusion matrix con gli elementi addensati verso la diagonale principale.
Per capire al meglio le cause degli errori e soprattutto la loro gravità, lo score di Gleason globale calcolato sulle immagini è stato separato per permettere lo studio della ricerca del pattern principale e del pattern secondario. Per questo sono state costruite le confusion matrix del modello rispetto ai patologi separatamente per i due pattern.

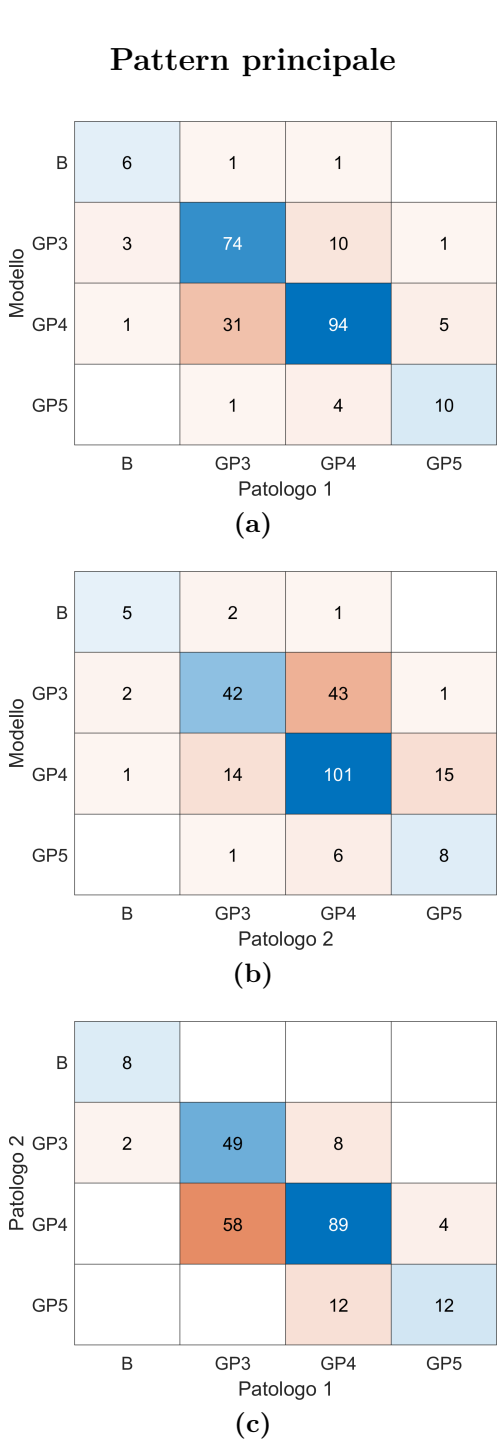


Figura 3.3: Confusion matrix sul pattern principale: a) modello vs patologo 1, b) modello vs patologo 2, c)patologo 2 vs patologo 1.

Nel riconoscimento del pattern principale, il classificatore presenta un'accuratezza del 76.03% rispetto al patologo 1 e del 64.46% rispetto al patologo 2. Tra i due patologi l'accuratezza è del 65.28%. Il modello, probabilmente a causa di un leggero overfitting, si trova in accordo con il patologo 1 con un grado maggiore rispetto al patologo 2. Analizzando la confusion matrix del modello rispetto al patologo 1, gli errori si presentano maggiormente tra il pattern 3 e il pattern 4 e solamente il 7.4% dei pattern principali sono stati sottostimati. Tra il modello e il patologo 2 gli elementi sottostimati aumentano ma la maggior parte sono pattern di tipo 4 classificato come pattern 3.
In generale, le prestazioni del classificatore sulla ricerca del pattern principale sono comparabili e in linea con la variabilità presente tra i due patologi. L'algoritmo mostra una buona affidabilità nella ricerca del pattern tumorale principale.

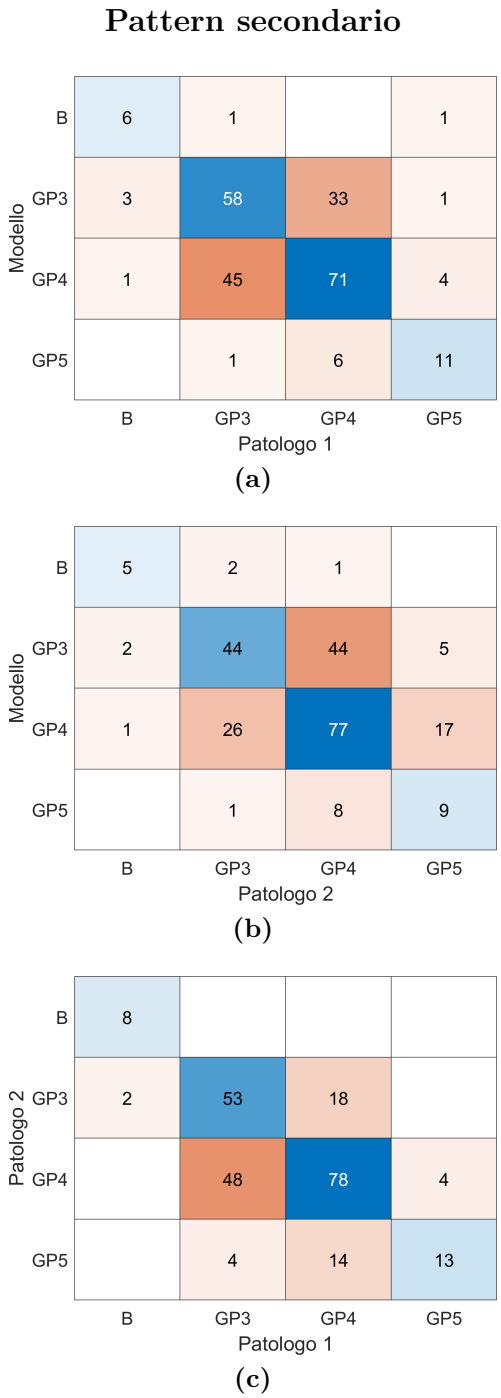


Figura 3.4: Confusion matrix sul pattern secondario: a) modello vs patologo 1, b) modello vs patologo 2, c)patologo 2 vs patologo 1.

Per quanto riguarda il pattern secondario, quindi il secondo pattern più

presente nell'immagine, l'accuratezza tra il modello e il patologo 1 raggiunge il 60.33% mentre tra il modello e il patologo 2 l'accuratezza è 55.78%. Tra i due patologi invece l'accuratezza è 62.80%. Quindi ci sono più problemi nel riconoscere il pattern secondario rispetto al pattern primario ma anche in questo caso si può facilmente notare come la maggior parte degli errori sia come al solito tra il pattern 3 e il pattern 4. Anche per il pattern secondario si può concludere che, analizzando le confusion matrix, le prestazioni del classificatore sviluppato raggiungono il livello del grado di accordo tra i due patologi.

Capitolo 4

Conclusioni e sviluppi futuri

4.1 Conclusioni

In questo lavoro di tesi è stato sviluppato un algoritmo automatico per l'assegnazione dello score di Gleason in immagini istopatologiche prostatiche. L'algoritmo è in grado di ricevere in input un'immagine di un campione di TMA e restituire il Gleason score dell'intera immagine.
Il sistema realizzato, basato sul machine learning ed in particolare sull'apprendimento delle reti neurali convoluzionali, ha dimostrato una discreta generalizzazione del problema. Può tornare utile come strumento di supporto e come seconda opinione per un patologo esperto. L'algoritmo presenta un'affidabilità sul calcolo del Gleason score sull'intera biopsia che rispecchia la variabilità tra gli operatori.
La soggettività è un limite della scala di Gleason poichè patologi differenti utilizzano parametri diversi, arrivando molto spesso a conclusioni differenti. L'algoritmo sviluppato può fornire un immediato parere che può indirizzare le valutazioni del patologo verso determinate caratteristiche e di conseguenza ridurre i tempi di analisi delle biopsie.
Il problema della variabilità nell'assegnazione dello score di Gleason è eviden-

ziata soprattutto tra le classi molto simili come il Gleason 3 e il Gleason 4 e di conseguenza anche il classificatore implementato mostra diverse limitazioni da questo punto di vista.
Le grandi potenzialità che presenta l'algoritmo, oltre all'utilizzo del deep learning, si trovano nell'estrazione di parti di immagini per la classificazione con approccio a patch. Il metodo di estrazione sviluppato sfrutta le caratteristiche anatomiche microscopiche della biopsia per ottenere in maniera robusta le patch più utili con cui allenare le reti neurali. In questo modo l'apprendimento risultato molto più mirato verso le caratteristiche fondamentali sulle quali basare la classificazione. Inoltre la dimensione scelta delle patch permette una visione completa delle strutture ghiandolari, indicatore fondamentale per la classificazione nella scala di Gleason.
In generale, nonostante sia stato scelto di testare sulle immagini di test solamente il secondo approccio caratterizzato da due reti neurali in cascata, tutti gli approcci provati mostrano un potenziale che può essere proposto per ulteriori approfondimenti e studi. L'algoritmo sviluppato può fornire una buona base per sviluppi e miglioramenti futuri e raggiungere un livello di attendibilità tale da poter affiancare un patologo esperto nell'assegnazione dello score di Gleason.

4.1.1 Sviluppi futuri

L'algoritmo sviluppato ha raggiunto buoni risultati ma presenta alcune limitazioni da superare per poter fornire un completo supporto alla prognosi del tumore prostatico. Dalla corretta classificazione del cancro infatti dipende il percorso terapeutico per i pazienti. Il dataset pubblico utilizzato è composto da 846 immagini di TMA di cui 641 sono state usate come training set e 244 (il completo TMA 80) come test set. Un dataset più ampio permetterebbe una maggiore raccolta di patch e di conseguenza una più ampia visione delle casistiche possibili.

Fondamentale per questo tipo di studi è sicuramente la bontà del dataset. E' necessario che le annotazioni manuali vengano fatte con estrema precisione per avere un riferimento (ground truth) affidabile ed evitare al più possibile patch inutili all'allenamento della rete neurale convoluzionale.
Un sistema di questo tipo, supportato da annotazioni molto precise di zone tumorali ognuna contrassegnata dal proprio score di Gleason, può essere aggiornato per fornire una diagnosi segmentando il tumore e successivamente assegnandone anche lo score di Gleason.
Sicuramente questo algoritmo può essere modificato e migliorato per poter essere applicato sulle WSI (whole slide image), quindi sulle immagini delle intere biopsie.

Ringraziamenti

Vorrei prima di tutto ringraziare il prof. Filippo Molinari per avermi concesso l'occasione di svolgere questo lavoro lavoro di tesi. Un ringraziamento particolare va all'Ing. Massimo Salvi per la sua disponibilità e il suo continuo supporto mostrato durante tutto il lavoro svolto. Grazie inoltre al Biolab per l'incredibile accoglienza.

Grazie alla mia famiglia, ai miei genitori, a mio fratello Daniele e a mia sorella Giulia per essere stati un punto di riferimento e per avermi permesso di intraprendere questo viaggio di 6 anni senza farmi mai mancare nulla.

Grazie a Nonna Anna per essersi dimostrata, forse senza accorgersene, una fonte di forza ed energia inestimabile. E grazie a tutti i nonni che non ci sono più per farmi sentire, anche oggi, il loro appoggio costante.

Grazie alla mia Robertina, conosciuta forse troppo tardi, che non ha mai smesso di credere in me e che mi ha accompagnato come solo lei sa fare negli ultimi anni di questa avventura. Adesso viene il bello!

Grazie a Vali, una scoperta speciale che ha ufficialmente occupato un posto nel mio cuore.

Grazie a Stefano, ritrovato dopo anni (grazie al Politecnico), è diventato un compagno prima e un appoggio unico ed inimitabile dopo.

Grazie a Carlo, Gianni e Totò che hanno arricchito il viaggio con una simpatia fuori dal comune e con un pizzico di ignoranza.

Grazie agli amici di sempre di Scuola Materna per essere stati un aiuto ed una presenza fissa durante l'intera durata degli studi.

Grazie a Federicone per essere stato la mia anima gemella per 5 anni e per aver raggiunto insieme, si può proprio dire, questo traguardo.
Grazie a Mogio, Franceschini e tutto il fedozzo per essere stati compagni di viaggio straordinari ed aver reso decisamente più dolci questi anni.
Grazie all'Atletico Carlone, la mia squadra del cuore nella quale ho trovato persone speciali.

Bibliografia

[1] Arvaniti E. «Automated Gleason grading of prostate cancer tissue microarrays via deep learning». In: *Scientific Reports* 8 (2018).

[2] Institute for Quality e Efficiency in Health Care (IQWiG). «How does the prostate work?» In: (2011). Available online. URL: https://www.ncbi.nlm.nih.gov/books/NBK279291/.

[3] World Cancer Research Fund. Available online. 2018. URL: https://www.wcrf.org/dietandcancer/cancer-trends/prostate-cancer-statistics.

[4] Associazione Italiana Oncologia Medica. Available online. 2018. URL: https://www.aiom.it/.

[5] Jennifer Gordetsky e Jonathan Epstein. «Grading of prostatic adenocarcinoma: current state and prognostic implications». In: *Diagnostic Pathology* (2016).

[6] Jonathan Epstein e Lars Egevad. «The 2014 Interantional Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma». In: *The American Journal of Surgical Pathology* (2014).

[7] Human Pathology. Available online. 2012. URL: https://www.humpath.com/spip.php?article18060.

[8] William C. Allsbrook et al. «Interobserver reproducibility of Gleason grading of prostatic carcinoma: Urologic pathologists». In: *Human Pathology* (2001).

[9] Singh R. et al. «Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists». In: *Indian Journal of Cancer* (2011).

[10] Anna Gummeson et al. «Automatic Gleason grading of H and E stained microscopic prostate images using deep convolutional neural networks». In: *Medical Imaging: Digital Pathology* (2017).

[11] Hongming Xu, Sunho Park e Tae Hyun Hwang. «Automatic Classification of Prostate Cancer Gleason Scores from Digitized Whole Slide Tissue Biopsies». In: *bioRxiv* (2018).

[12] W. Han et al. «Automatic high-grade cancer detection on prostatectomy histopathology images». In: 10956 (2019).

[13] Marc Macenko et al. «A Method for Normalizing Histology Slides for Quantitative Analysis.» In: *Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009* 9 (2009), pp. 1107–1110.

[14] Abhishek Vahadane et al. «Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images». In: *IEEE Transactions on Medical Imaging* 35 (2016).

[15] Massimo Salvi e Filippo Molinari. «Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images». In: *BioMedical Engineering OnLine* 17 (2018), pp. 1107–1110.

[16] A. Arsenov et al. «Evolution of convolutional neural network architecture in image classification problems». In: *Ceur workshop processing* (2018).

[17] Towards Data Science. Available online. 2018. URL: https://www.towardsdatascience.com.

[18] DeepLearning.ai. Available online. URL: https://www.coursera.org/learn/convolutional-neural-networks.

[19] Jeremy West, Dan Ventura e Sean Warnick. «Spring Research Presentation: A Theoretical Foundation for Inductive Transfer». In: (2007).

[20] Mathworks. Available online. URL: https://it.mathworks.com/discovery/transfer-learning.html.

[21] Wei Wang et al. «Development of convolutional neural network and its application in image classification: a survey». In: *Optical Engineering* 58.4 (2019).

[22] Min Lin, Qiang Chen e Shuicheng Yan. «Network In Network». In: (2013).

[23] Christian Szegedy et al. «Going Deeper with Convolutions». In: (2014).

[24] Deep Learning. Available online. URL: https://www.deeplearning.ai/ai-notes/optimization.

[25] Jesse K. McKenney et al. «The Potential Impact of Reproducibility of Gleason Grading in Men With Early Stage Prostate Cancer Managed by Active Surveillance: A Multi-Institutional Study». In: *The Journal of Urology* 186.2 (2011), pp. 465–469.

[26] Towards Data Science. Available online. 2018. URL: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.