



Corso di Laurea Magistrale in Ingegneria Biomedica

Facoltà Ingegneria

Politecnico di Torino

Sviluppo di un sistema automatico di supporto diagnostico per la valutazione del Gleason Score in istologie prostatiche

Autore:

Ivan Tomarchio

Relatori:

Prof. Filippo Molinari

Ing. Massimo Salvi

Marzo, 2021

Sviluppo di un sistema automatico di supporto diagnostico per la valutazione del Gleason Score in istologie prostatiche

Ivan Tomarchio

Relatori:

Prof. Filippo Molinari

Ing. Massimo Salvi

Sommario

L'adenocarcinoma prostatico è una delle patologie tumorali di maggior interesse dal punto di vista della ricerca scientifica essendo la seconda forma tumorale più diagnosticata nell'uomo e la terza per numero di morti ogni anno nel mondo nonostante la sua bassa letalità. Si tratta di una forma tumorale che colpisce prevalentemente la popolazione over 65 ed il suo tasso di sopravvivenza a 5 anni se diagnosticata in fasi iniziali è maggiore del 90%. Per tale ragione a differenza di forme tumorali più aggressive nelle fasi iniziali è possibile attuare delle politiche di vigilanza attiva piuttosto che delle politiche di rimozione radicale dell'organo per migliorare la qualità di vita del paziente. Per rendere possibile ciò è fondamentale disporre di metodologie diagnostiche solide che consentano la classificazione corretta nelle diverse classi di rischio del paziente. Il gold standard diagnostico è l'analisi istologica di agobiopsie che consente di ottenere informazioni circa lo stadio evolutivo del carcinoma prostatico tramite la scala di valutazione di Gleason. La scala di Gleason prevede la valutazione di alcune caratteristiche architetturali dei campioni assegnando ai diversi pattern presenti uno score in funzione del grado di differenziazione delle ghiandole prostatiche partendo da Benigno, perfettamente differenziate, a Gleason 5, perdita completa della struttura ghiandolare. In seguito valutando i due peggiori pattern presenti nei campioni si assegna un Grade Group da 1-5, con livello di rischio per il paziente crescente, che sarà fondamentale nella scelta del percorso teranostico. La criticità principale di tale tecnica è legata alla soggettività di valutazione tra diversi patologi e alle variabilità intrinseche dell'esame. Il presente lavoro di tesi si pone come obiettivo quello di valutare la fattibilità di sviluppo di un sistema di supporto diagnostico per il patologo che lo aiuti nella classificazione. Il sistema sviluppato si basa sull'utilizzo di reti neurali convoluzionali, ad oggi tra i più potenti strumenti di analisi di immagini. Le immagini utilizzate sono state ottenute dal dataset di Arvaniti E. [1]. Per ottimizzare la classificazione nelle diverse classi sono state testate diverse strutture neurali. La validazione del sistema è stata effettuata confrontando la sua classificazione con quella di due patologi esperti.

Abstract

Prostate adenocarcinoma is one of the most interesting tumour for both clinincal and scientific research, being the second most diagnosed form of cancer in humans and the third for number of deaths every year in the world despite its low lethality. This form of cancer mainly affects the population over 65 years old and its 5-year survival rate, if diagnosed in the early stages, is greater than 90 %. For this reason, unlike more aggressive forms of cancer in the initial stages, it is possible to implement active surveillance policies rather than policies of radical removal of the organ to improve the patient's life quality. To make this possible, it is essential to have solid diagnostic methodologies that allow the correct classification in the different classes of patient risk. The gold standard for the diagnosis, the histological analysis of needle biopsies, allows obtaining information about the evolutionary stage of prostate cancer through the Gleason rating scale. The architectural characteristics of the samples are evaluated by assigning to the different patterns a mark according to the degree of differentiation of the prostate glands from Benignant, perfectly differentiated, to Gleason 5, complete loss of the glandular structure. Subsequently, by evaluating the two worst patterns present in the samples, a Grade Group in the range 1-5 is assigned, with an increasing level of risk for the patient, which will be fundamental to choose the right theranostic path. The main drawback of this technique is the subjectivness of the evaluation, dependent on the conclusions of the different pathologist and to the intrinsic variability of the examination. In my disseration I evaluate the feasibility of developing a diagnostic support system for the pathologist. The developed system, based on the usage of convolutional neural networks, was trained with the images obtained from the Arvaniti E. Dataset. Different neural structures were tested to optimize the classification. The system is validated by comparing its classification with that of two expert pathologists.

Riconoscimenti

Innanzitutto ringrazio il professor Molinari per avermi concesso la possibilità di svolgere questo lavoro di tesi. Un ringraziamento particolare all'Ing. Salvi per i consigli e per la sua grande disponibilità durante tutto il percorso. Ringrazio col cuore i miei genitori per il loro supporto e per aver sempre creduto in me. Ringrazio le mie sorelle perchè ci sono state e ci saranno sempre per me. Ringrazio tutti gli amici che hanno reso questi 5 anni a Torino i più belli della mia vita.

Indice

1	Introduzione	14
1.1	Prostata	14
1.1.1	Anatomia	14
1.1.2	Funzioni fisiologiche	16
1.1.3	Patologie	16
1.1.4	Fattori d'insorgenza	18
1.2	Tecniche diagnostiche	19
1.2.1	Dosaggio del PSA	19
1.2.2	Esplorazione rettale	20
1.2.3	Ecografia trans-rettale	20
1.2.4	Risonanza magnetica	20
1.2.5	Agobiopsia ed Esame istologico	21
1.3	Gleason score e Grade Group	22
1.4	Linee guida terapeutiche	26
1.4.1	Vigile Attesa	26
1.4.2	Sorveglianza attiva	26
1.4.3	Trattamenti locali a scopo radicale	26
1.5	Statistiche del carcinoma	27
1.6	Finalità dello studio	28
2	Stato dell'arte	30
2.1	Reti Neurali Convoluzionali	30
2.2	Valutazione del Gleason Score tramite CNN	34
2.2.1	Nagpal	34
2.2.2	Bulten	35
2.2.3	Arvaniti	35
2.3	Studi finalizzati alla classificazione binaria	36
2.3.1	Campanella	36
3	Metodi	40
3.1	Dataset	40
3.2	Organizzazione del lavoro	41
3.3	Ottimizzazione dei dati	41
3.4	Generazione delle maschere	44
3.5	Estrazione smart delle patches	46
3.6	Training	49
3.6.1	Strutture testate	49
3.6.2	Transfer Learning	49

3.6.3	Data Augmentation	50
3.7	Metriche di valutazione dei risultati e validazione	51
3.7.1	Confusion matrix	51
3.7.2	Kappa Cohen	52
3.7.3	Curve ROC	52
3.8	Heatmap e calcolo del Grade Group	54
4	Risultati	56
4.1	Classificazione delle patches del Test Set	56
4.1.1	Sistema 1	56
4.1.2	Sistema 2	58
4.1.3	Sistema 3	60
4.2	Validazione del miglior sistema	61
4.2.1	Classificazione del Grade Group	63
4.2.2	Valutazione delle Heatmap	66
5	Conclusioni	72
5.1	Conclusioni	72
5.2	Sviluppi futuri	73

Elenco delle figure

1.1	Nella <i>fig.(a)</i> è rappresentata l'anatomia del distretto prostatico. Immagine tratta da [3]. Nella <i>fig.(b)</i> la suddivisione della prostata nelle diverse zone: CZ (zona centrale), PZ (zona periferica), TZ (zona transizionale) e FM (zona fibromuscolare). La zona FM risulta priva di tessuto ghiandolare. Immagine tratta da [4].	15
1.2	Immagine istologica con coloranti ematossilina ed eosina di tessuto prostatico. Nella figura sono evidenziate in verde una ghiandola prostatica, in blu una zona ricca di stroma e in rosso le cellule basali.	15
1.3	Nella figura (a) è mostrata una prostata sana. Nella figura (b) è mostrata una prostata ipertrofica a causa di IPB. Immagine tratta da [9].	17
1.4	Rappresentazione fisica degli stadi del carcinoma prostatico. Immagine ottenuta da [10].	18
1.5	(a) Porzione di tessuto prostatico dove sono evidenziate due lesioni ipoecogene. Immagine tratta da [17]. (b) In evidenza una zona nodulare caratterizzata da una minore intensità di segnale. L'analisi istologica ha confermato la presenza di lesione tumorale. Immagine tratta da [18].	21
1.6	(a) Slice ottenuta da un microarray di tessuto. Immagine tratta da Dataset pubblico di Arvaniti E.[1] (b) Porzione di una WSI. Immagine ottenuta da Dataset interno.	22
1.7	Modifiche nel tempo delle caratteristiche istologiche architetturali utilizzate per assegnare il Gleason score. Immagine tratta da [19]. . .	23
1.8	Patch contenenti i diversi pattern riscontrabili in istologia prostatica .	24
1.9	Immagini estratte dal Dataset di Gleason andando a sovrapporre i bordi delle annotazioni manuali del patologo 1. I relativi Grade Group sono: a) GG1 (3 + 4), b) GG2 (4 + 3), c) GG3 (4 + 4), d) GG4 (5+3) e f) GG5 (5 + 4)	25
1.10	Schema di funzionamento di un processo diagnostico che integri al suo interno il sistema sviluppato	28
2.1	Schema architetturale classico di una rete neurale convoluzionale. Immagine tratta da [23].	30
2.2	Rappresentazione visiva dell'operazione di convoluzione	31
2.3	Struttura del sistema proposto da Nagpal	34
2.4	Struttura standard di una rete U-net, costituita da un percorso di contrazione che favorisce la cattura del contesto (sx) e da un percorso di espansione che favorisce la localizzazione (dx)	35

2.5	Andamento dell'errore minimo di validazione in funzione della dimensione della dimensione del dataset. Immagine tratta da [29]	36
2.6	Sistema sviluppato da Campanella per la classificazione binaria di istologie prostatiche. Immagine tratta da [30]	37
3.1	Schema di sviluppo del lavoro	41
3.2	Patches estratte nella zona di tessuto, con dimensione differenti. La patch (a) di dimensione 299x299 non riesce a catturare l'intera architettura della ghiandola prostatica.	42
3.3	a) Valori dei pixel nello spazio RGB. Come si può notare le due colorazioni risultano separabili ma tramite una curva. b) Valori dei pixel nello spazio OD. Le due colorazioni risultano separabili tramite una relazione lineare. Immagine tratta da [31]	43
3.4	Esempio pre-processing. Le variazioni sono minime poichè le istologie sono state acquisite nello stesso centro, con lo stesso scanner e lo stesso lotto di colorante.	44
3.5	Nella figura sono mostrate l'immagine 62 e la relativa maschera del patologo più tutte le maschere generate per ottimizzare l'estrazione delle patches. Come si può notare valutando le sottoimmagini (b), (e) ed (f) si ha una maggiore concentrazione di nuclei nelle regioni annotate dal patologo mentre le concentrazioni più elevate di stroma si trovano al di fuori delle annotazioni.	45
3.6	Flowchart dell'algoritmo di estrazione delle patches	46
3.7	Esempi di estrazione smart delle patches. Nell'immagine di destra si ha attivazione del Criterio B* che garantisce l'estrazione delle ghiandole benigne.	47
3.8	Esempi di figure di debug dell'estrazione delle patches tramite l'algoritmo smart senza valutare l'annotazione manuale con overlap del 50%	48
3.9	Schema di funzionamento del transfer learning	49
3.10	Nella <i>fig.(a)</i> l'immagine originale mentre nella <i>fig.(b)</i> l'immagine augmentata	50
3.11	Esempio di una confusion matrix a 4 classi	51
3.12	Esempio di confusion matrix relativa ad un problema di classificazione binario	52
3.13	Esempio curve ROC e AUC per 3 diversi sistemi.	53
3.14	Pipeline per la generazione del Gleason Score e Grade Group delle immagini intere	54
4.1	Nella figura sono presenti le confusion matrix del classificatore normal <i>fig.(a)</i> e smart <i>fig.(b)</i> relative al sistema 1	57
4.2	Curve ROC dei due classificatori del sistema 1	58
4.3	Nella figura sono presenti le confusion matrix del classificatore normal <i>fig.(a)</i> e smart <i>fig.(b)</i> relative al sistema 2	58
4.4	Curve ROC delle reti a 3 classi del sistema 2	59
4.5	Nella figura sono presenti le confusion matrix del classificatore normal <i>fig.(a)</i> e smart <i>fig.(b)</i> relative al sistema 3	60
4.6	Curve ROC delle reti Gleason 3 vs Gleason 4 del sistema 3	61

4.7	In figura le confusion matrix delle patch predette con maggiore probabilità con due soglie differenti	62
4.8	Confusion Matrix relativa al Grade Group del modello vs Patologo 1	63
4.9	Confusion matrix relative a pattern primario e secondario del modello vs Patologo 1	64
4.10	Confusion Matrix relativa al Grade Group del modello vs Patologo 2	64
4.11	Confusion matrix relative a pattern primario e secondario del modello vs Patologo 2	65
4.12	Confusion Matrix relativa al Grade Group del Patologo 1 vs Patologo 2	65
4.13	Confusion matrix relative a pattern primario e secondario del Patologo 1 vs Patologo 2	66
4.14	Heatmap immagine 33. Patologo 1: Benigno (verde), Patologo 2: Gleason 3(Blu).	67
4.15	Heatmap immagine 51. Patologo 1: Gleason 3 (blu) + Gleason 4 (giallo), Patologo 2: Gleason 4 (giallo)	67
4.16	Heatmap immagine 64. Patologo 1: Gleason 3 (blu), Patologo 2: Gleason 4 (giallo) + Gleason 3 (blu).	68
4.17	Heatmap immagine 119. Patologo 1: Gleason 3 (blu), Patologo 2: Gleason 3 (blu).	69
4.18	Heatmap immagine 120. Patologo 1: Gleason 4 (giallo), Patologo 2: Gleason 4 (giallo).	69
4.19	Heatmap immagine 79. Patologo 1: Gleason 5 (rosso), Patologo 2: Gleason 5 (rosso).	70

Elenco delle tabelle

1.1	Classificazione del carcinoma prostatico mediante il parametro T . . .	18
1.2	Assegnazione del Gleason Grade Group in funzione dei Gleason score	23
2.1	Tabella riassuntiva	38
3.1	Nella tabella di <i>sinistra</i> la distribuzione delle immagini del construction set (training e validation). Nella tabella di <i>destra</i> la distribuzione della immagine del test set del Patologo 1.	40
3.2	Numero di patches estratte dal construction set per ogni classe senza e con estrazione smart	47
3.3	<i>Suddivisione dei due dataset utilizzati per l'allenamento del classificatore a 4 classi con e senza estrazione smart e data augmentation</i> . .	50
4.1	Caratteristiche del sistema 1 normal	57
4.2	Caratteristiche del sistema 1 smart	57
4.3	Caratteristiche del sistema 2 normal	59
4.4	Caratteristiche del sistema 2 smart	59
4.5	Caratteristiche del sistema 3 normal	60
4.6	Caratteristiche del sistema 3 smart	61
4.7	Parametri di valutazione delle patches predette con probabilità di classificazione maggiore dell'80%	62
4.8	Parametri di valutazione delle patches predette con probabilità di classificazione maggiore dell'90%	63
4.9	Valutazione del kappa di Cohen per le diverse coppie	66

Capitolo 1

Introduzione

1.1 Prostata

1.1.1 Anatomia

La ghiandola prostatica è situata nel comparto subperitoneale ed è racchiusa tra diaframma pelvico e cavità peritoneale. In condizioni normali, presenta un diametro di circa 4cm , una lunghezza di 3cm e una larghezza di 2cm . Trovandosi posteriormente rispetto alla sinfisi pubica ed anteriormente rispetto al retto, risulta possibile esaminarla tramite palpazione digitale. Con la sua forma conica, circonda l'uretra prossimale al di fuori della vescica. È composta da una base, un apice e da quattro diverse superfici: anteriore, posteriore e due infero-laterali. La base è attaccata al collo della vescica e l'uretra entra nella parte mediana vicino la superficie anteriore. L'apice è situato sulla superficie superiore del diaframma urogenitale a contatto con la superficie mediale del muscolo elevatore anale. La superficie posteriore, triangolare e piatta, giace sulla parete anteriore del retto. Le superfici infero-laterale e anteriore si trovano in contatto con il fascio muscolare elevatore dell'ano al di sopra del diaframma urogenitale. La prostata umana è composta da elementi ghiandolari e stroma, ovvero tessuto fibromuscolare. Le diverse ghiandole prostatiche sono collegate mediante dotti escretori ai seni prostatici dell'uretra prostatica, dove riversano il loro contenuto durante l'eiaculazione. La prostata è suddivisa in 3 zone che presentano ghiandole: centrale (CZ) 25% volume prostatico, transizione (TZ) 5% del volume prostatico, periferica (PZ) 70% del volume prostatico. Queste diverse zone presentano caratteristiche istologiche ed anatomiche differenti ed una differente suscettibilità ad eventi patologici. Il 70% delle patologie tumorali prostatiche si origina nella zona PZ, il 25% nella zona TZ mentre la zona CZ, originata da origini embrionali differenti dalle altre due risulta molto meno colpita. La zona TZ ha origini embrionali molto simili alla zona PZ, ma presenta una incidenza tumorale molto inferiore, ed è possibile spiegare ciò valutando la differente componente di stroma delle due zone. La TZ, con una maggiore incidenza di *IPB*¹, risulta maggiormente fibromuscolare e ciò ha portato a pensare che la *IPB* sia una malattia dello stroma fibromuscolare [2]. Lo stroma fibromuscolare anteriore è privo di ghiandole.

¹Iperplasia prostatica benigna

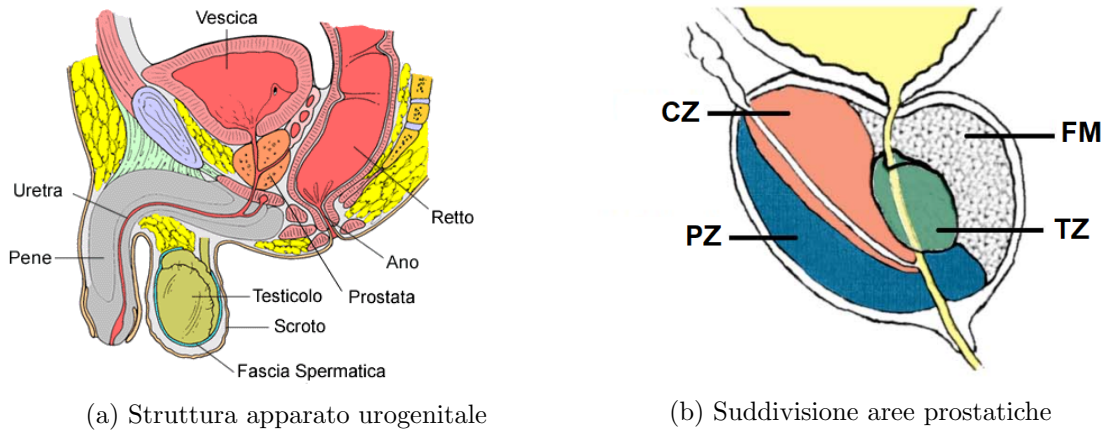


Figura 1.1: Nella *fig.(a)* è rappresentata l'anatomia del distretto prostatico. Immagine tratta da [3]. Nella *fig.(b)* la suddivisione della prostata nelle diverse zone: CZ (zona centrale), PZ (zona periferica), TZ (zona transizionale) e FM (zona fibromuscolare). La zona FM risulta priva di tessuto ghiandolare. Immagine tratta da [4].

La prostata è costituita da ghiandole tubuloacinari supportate da stroma connettivale, a sua volta attraversato da vasi linfatici ed ematici, che ha il compito di fornire stabilità strutturale. Le proporzioni tra questi due variano nelle diverse zone prostatiche. Le ghiandole presentano un epitelio a doppio strato: uno strato colonnare secretorio ed uno strato profondo cuboidale. È proprio quest'ultimo a subire delle modificazioni con l'insorgenza di patologie maligne. I dotti prostatici hanno dimensioni variabili nelle differenti zone, da diametri di $0.15-0.3\text{mm}$ nelle zone PZ/TZ fino a raggiungere 0.6mm nella zona CZ [5]. I bordi di dotti ed acini risultano poligonali e con prominenti ondulazioni delle pareti epiteliali, accentuate nelle zone PZ e TZ, per consentire l'espansione del lume ed avere una funzione di accumulo delle secrezioni ghiandolari. Variazioni delle pareti, che assumono forme più circolari o ovali, sono un criterio per la diagnosi di carcinoma ad alto tasso di differenziazione [6]. Come per gli altri organi ghiandolari, le cellule secretorie, gli acini, sono separati dalla membrana basale e dallo stroma da uno strato di cellule basali di aspetto filiforme, con nucleo scuro e un citoplasma difficilmente distinguibile.

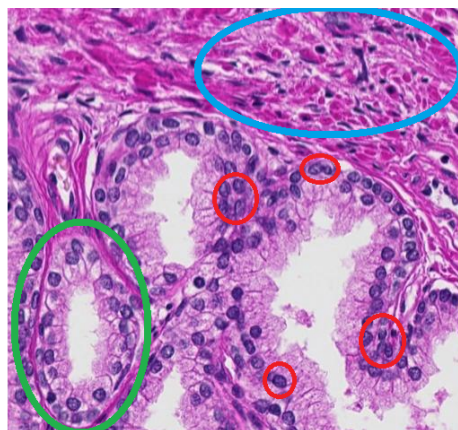


Figura 1.2: Immagine istologica con coloranti ematossilina ed eosina di tessuto prostatico. Nella figura sono evidenziate in verde una ghiandola prostatica, in blu una zona ricca di stroma e in rosso le cellule basali.

1.1.2 Funzioni fisiologiche

La prostata è una ghiandola secretoria *esocrina*² la cui principale funzione è quella di produrre ed immagazzinare il liquido prostatico, componente dello sperma per circa il 20-30%, fonte di nutrimento e veicolo di trasporto per gli spermatozoi. Nell'uretra il liquido prostatico si miscela con gli spermatozoi, il fluido seminale e le secrezioni di altre ghiandole, per formare lo sperma. Il liquido prostatico, leggermente acido ($\text{PH} \approx 6$), rende lo sperma ospitale per le cellule spermatiche introducendo in esso enzimi, zinco e citrato. Un importante enzima, prodotto dalla prostata, è il *PSA*³ che ha lo scopo di rendere lo sperma più diluito e fluido, favorendone il passaggio attraverso l'uretra e la sopravvivenza al viaggio fino all'ovocita [7]. Un importante funzione meccanica della prostata è quella di ostruire il collegamento tra uretra e vescica, grazie alla sua contrazione durante l'eiaculazione, con due effetti: impedire il mescolamento con l'urina e tramite aumento della pressione incrementare la velocità dello sperma in uscita. Inoltre svolge anche un ruolo attivo nel metabolismo ormonale del testosterone, che qui viene ridotto tramite l'enzima 5α -reduttasi, in *DHT*⁴, ormone responsabile del mantenimento dei tratti specifici sessuali maschili [8].

1.1.3 Patologie

Prostatite

La prostatite è una malattia infiammatoria della prostata, in forma acuta o cronica, che può avere origine batterica o essere aspecifica. L'eziopatogenesi per la forma:

- **Batterica:** batteri gram negativi del tratto urinario che raggiungono il tessuto prostatico e proliferano nel tessuto intra-ghiandolare. Può inoltre essere causata da TBC⁵ polmonare o genito-urinaria
- **Aspecifica:** causa sconosciuta, frequente in persone con età >50 anni

La manifestazione clinica consiste nell'insorgenza di dolore alla minzione, dolore perineale, brividi e febbre. Si sospetta la correlazione tra ripetuti eventi infiammatori prostatici nell'arco della vita e insorgenza del carcinoma prostatico anche se non ancora dimostrata.

Iperplasia prostatica benigna

Si tratta di un'iperplasia nodulare delle componenti ghiandolari e stromali della prostata. Interessa circa il 50% degli over 50-60 con un alto grado di incidenza nei paesi industrializzati. Le cause dell'iperplasia sono da ricercare nell'invecchiamento cellulare che porta all'aumento della sensibilità delle cellule epiteliali e dello stroma agli androgeni ed estrogeni.

²Riversa il proprio contenuto verso l'esterno

³Prostate specific antigen

⁴ 5α -Diidrottestosterone, metabolita biologicamente attivo dell'ormone testosterone, di cui risulta più potente grazie all'affinità maggiore con i recettori androgeni

⁵Tubercolosi

La manifestazione clinica di tale patologia consiste in un ingrossamento della prostata con conseguente:

- Esitazione iniziale durante la minzione
- Diminuzione dell'intensità urinaria fino alla possibile occlusione
- Svuotamento incompleto della vescica

Solitamente quando si sospetta la presenza di IPB viene effettuata una diagnosi differenziale tramite agobiopsia di tutti i quadranti prostatici per escludere l'adenocarcinoma.

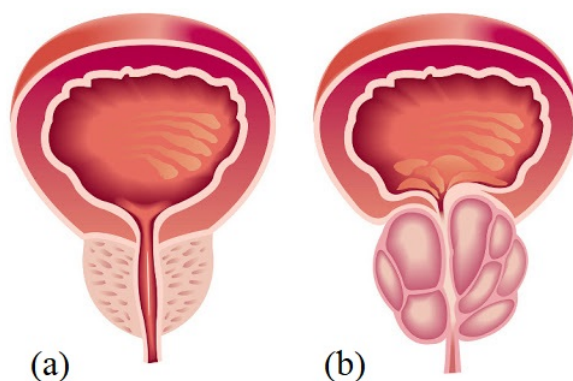


Figura 1.3: Nella figura (a) è mostrata una prostata sana. Nella figura (b) è mostrata una prostata ipertrofica a causa di IPB. Immagine tratta da [9].

Adenocarcinoma prostatico

Si tratta di una lesione tumorale del tessuto prostatico con maggior grado di incidenza nella zona periferica. In fase iniziale è pressoché asintomatico e l'apparizione dei primi sintomi corrisponde ad una fase già avanzata della sua evoluzione. Per questa ragione è fondamentale attuare delle politiche di screening volte ad identificarlo in maniera precoce. E' originato da una lesione pre-neoplastica chiamata *Neoplasia prostatica intraepiteliale* che gradualmente porta alla rarefazione dello strato basale fino alla sua totale scomparsa. In seguito si ha una alterazione morfologica del tessuto prostatico dove le singole ghiandole prostatiche non sono più facilmente distinguibili e si formano ammassi irregolari di cellule neoplastiche. Si sospetta un ruolo degli androgeni nello sviluppo di tale patologia. Si pensa che il cancro alla prostata si origini da popolazioni minori di cellule staminali a bassa proliferazione situati nel compartimento basale [2]. Per valutare l'aggressività del tumore si utilizza la stadiazione tramite classificazione TMN, dove:

- **T**: è relativo all'estensione del carcinoma (vedi Tab. 1.1)
- **N**: si riferisce all'interessamento dei linfonodi, varia da 0 a 3, con gravità crescente
- **M**: si riferisce alla assenza/presenza di metastasi (vale 0-1)

T	Caratteristiche
T1	Tumore clinicamente non apprezzabile, non palpabile, ne visibile con le tecniche di imaging
T2	Tumore limitato alla prostata che interessa uno o entrambi i lobi
T3	Tumore esteso al di fuori della prostata e che interessa il collo vescicale o le vesciche seminali
T4	Tumore fisso che invade le strutture adiacenti oltre le vesciche seminali: collo della vescica, sfintere, retto, muscoli elevatori e/o parete pelvica

Tabella 1.1: Classificazione del carcinoma prostatico mediante il parametro T

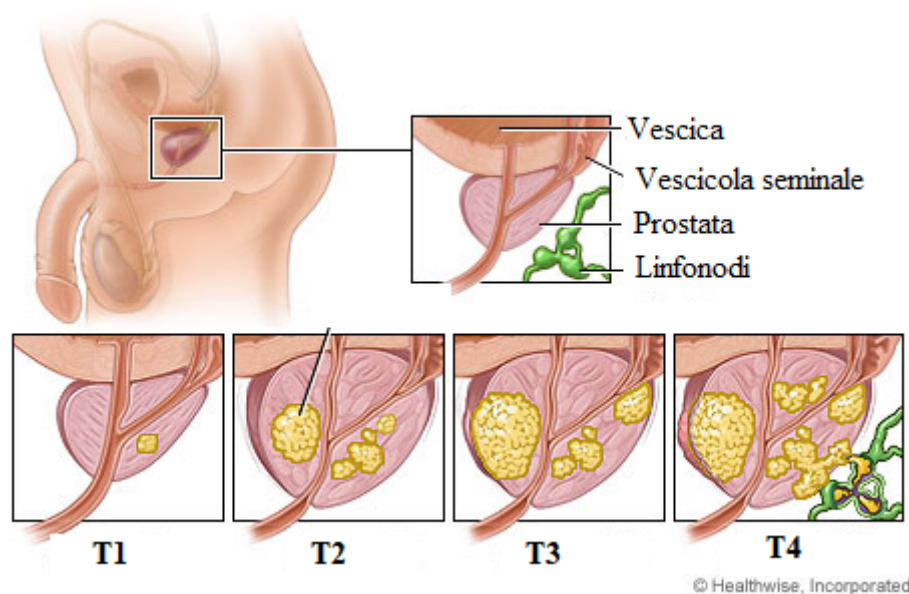


Figura 1.4: Rappresentazione fisica degli stadi del carcinoma prostatico. Immagine ottenuta da [10].

1.1.4 Fattori d'insorgenza

Di rischio

L'eziologia del carcinoma prostatico è multifattoriale, risultato di una complessa interazione tra fattori genetici ed ambientali tra cui:

- **Età:** l'incidenza del tumore aumenta con l'età
- **Etnia:** le persone di origine africana o latina presentano un maggior rischio a causa dei diversi livelli di androgeni, DHT e di 5-alfa reduttasi[11][12]
- **Fattori ormonali:** elevati livelli di testosterone e di IGF-1⁶ predispongono all'insorgenza del tumore[13]
- **Storia familiare:** circa il 25% dei casi
- **Fattori genetici:** 9% di forme ereditarie[11]

⁶*Somatomedina C* ormone prodotto dal fegato sotto stimolo dell'ormone della crescita

- **Stile di vita:** eccessivo apporto calorico e di grassi[13], eccessivo apporto di calcio, obesità e fumo
- **Storia clinica:** presenza di patologie infiammatorie croniche è correlata all'eziopatogenesi e alla progressione tumorale

Protettivi

Sono stati analizzati diversi fattori che potrebbero essere coinvolti con la diminuzione del rischio di insorgenza:

- **Attività fisica:** favorisce un migliore stato di salute
- **Abitudini alimentari:** consumo di caffè, pomodori, pesce e integrazione della vitamina D
- **Farmaci:** utilizzo di statine per abbassare i livelli di colesterolo

1.2 Tecniche diagnostiche

Non essendo completamente chiaro quali siano i meccanismi da adottare per effettuare una corretta prevenzione primaria, senza dubbio, la prevenzione secondaria è lo strumento più adeguato per diminuire la mortalità della malattia. Considerando l'elevato tasso di incidenza sulla popolazione di tale patologia è fondamentale l'impiego di meccanismi di screening.

1.2.1 Dosaggio del PSA

In condizione fisiologiche solo quantità minime di questo antigene raggiungono il circolo ematico. A causa delle modificazioni della normale istoarchitettura prostatica si ha un incremento ematico di PSA, che può essere dunque inteso come marcatore di patologia prostatica. Viene sfruttato come screening "opportunistico" a livello globale. Dal 1990 ad oggi, quest'esame, è stato valutato come meccanismo di screening in Europa dall'ERSPC⁷[14][15] i cui principali risultati in evidenza sono stati:

- **Diminuzione del rischio di morte:** è stato valutato un indice HR⁸ pari a 0.8 ($p=0.04$) a favore del gruppo di screening rispetto al gruppo di controllo
- **Bassa diminuzione della mortalità globale:** si ha solo una differenza di 0.71 morti ogni 1000 uomini sottoposti a screening
- **Elevato rischio di sovradiagnosi:** e di conseguente rischio di sovratrattamento
- **Aumento nel tempo della resa dello screening:** durante la durata complessiva dello studio
- **Riduzione della mortalità cancro-specifica:** del 29%

In seguito a tali conclusioni, ad oggi, le linee guida si contrappongono all'adozione di politiche di screening di popolazione basate sulla misurazione del PSA. Tuttavia è comune l'utilizzo spontaneo della misurazione del PSA come test di screening

⁷European Randomized Study of Screening for Prostate Cancer

⁸*Hazard rate* che si valuta tra due differenti gruppi

opportunistico a partire dai 50 anni e fino ai 70-75 anni. Bisogna tenere in conto, oltre all'eccessivo numero di falsi positivi generati dal test, che portano a sovratratteggio anche una percentuale di falsi negativi del 20-25% di tumori in fase iniziale caratterizzati da bassi valori di PSA [16].

1.2.2 Esplorazione rettale

L'esplorazione rettale (ER) è da intendere come il primo approccio diagnostico al paziente che presenti sintomatologia riferibili ad un eventuale patologia prostatica. Essa consente di rilevare al tatto lesioni che presentino dei volumi superiori a circa 0.2 mL. Valutando in maniera congiunta il risultato dell'ER e del dosaggio del PSA è possibile incrementare il valore predittivo positivo dell'esame, che comunque mantiene un livello di sensibilità e specificità abbastanza basse. Motivo per il quale è necessario combinare altre metodologie diagnostiche.

1.2.3 Ecografia trans-rettale

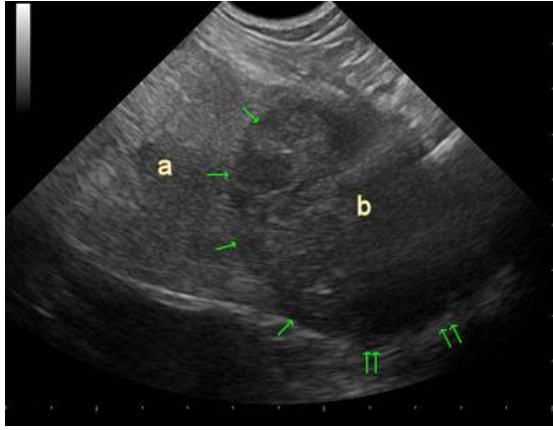
L'ecografia trans-rettale sfrutta la diversa ecogenicità del carcinoma prostatico. È molto efficace se utilizzata in sinergia con ER e dosaggio del PSA. Uno dei principali limiti di tale tecnica è il fatto che, nonostante una buona parte delle lesioni prostatiche siano ipoecogene, alcune forme di carcinoma della prostata sono iperecogene o addirittura isoecogene. Tale tecnica è molto importante come guida real-time alla biopsia prostatica.

1.2.4 Risonanza magnetica

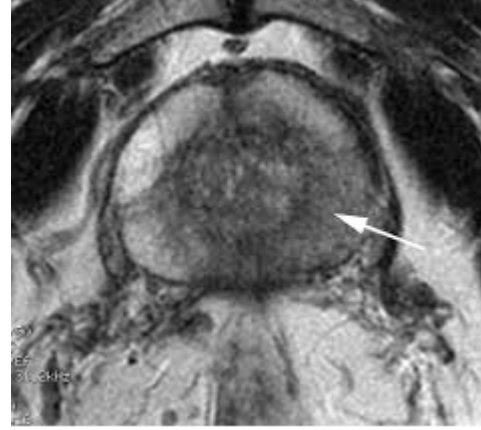
La RM multiparametrica con sequenze T2-pesate, T1-pesate in studio di perfusione (DCE-MRI⁹) e in diffusione (DWI¹⁰) fornisce delle informazioni anatomiche, biologiche e funzionali che consentono una migliore discriminazione delle lesioni sospette. Nello specifico lo studio della diffusione molecolare dell'acqua (DWI) consente di stabilire il grado di proliferazione della lesione tumorale e lo studio della perfusione-dinamica tramite mezzo di contrasto (DCE-MRI) permette la caratterizzazione delle lesioni in funzione dell'indice di vascolarizzazione. Viene fortemente indicata per pazienti con biopsie negative e persistente sospetto clinico di cancro della prostata come strumento di guida alla selezione delle aree di interesse durante la successiva biopsia.

⁹*Dynamic Contrast Enhanced Magnetic Resonance Imaging*

¹⁰*Diffusion Weighted Imaging*



(a) Ecografia Trans-rettale



(b) MRI assiale pesata T2

Figura 1.5: (a) Porzione di tessuto prostatico dove sono evidenziate due lesioni ipoecogene. Immagine tratta da [17]. (b) In evidenza una zona nodulare caratterizzata da una minore intensità di segnale. L'analisi istologica ha confermato la presenza di lesione tumorale. Immagine tratta da [18].

1.2.5 Agobiopsia ed Esame istologico

Rappresenta l'unica metodologia per ottenere la certezza diagnostica di neoplasia prostatica. Prima di sottoporre un paziente a agobiopsia¹¹ bisogna valutare aspettativa di vita ed eventuali comorbidità presenti. Durante una biopsia si effettua un numero di prelievi compreso tra 8-12 campioni, si prelevano inoltre campioni di ulteriori lesioni osservate durante l'esame. Essa risulta molto più efficace se indirizzata tramite le tecniche di imaging descritte in precedenza. Il campione prelevato dovrà successivamente essere trattato prima di poter essere analizzato dal patologo. La prima operazione da eseguire dopo il prelievo è la fissazione. Si utilizzano dei processi chimico-fisici per disattivare gli enzimi autolitici e modificare la consistenza del tessuto. Un fissativo comune è la formalina¹². Successivamente si tratta il tessuto con paraffina che conferisce resistenza, prima di procedere alla fase di inclusione. L'inclusione consiste nell'andare a sezionare il tessuto in sottili strati dello spessore di $2-5\mu m$, tramite microtomo a rotazione, per renderlo trasparente alla luce. Infine si ha la fase di colorazione che consente, utilizzando coloranti affini a diversi costituenti del tessuto, di aumentare il contrasto tra le diverse strutture. La colorazione più utilizzata nell'analisi istologica di tessuto prostatico è quella ematossilina-eosina. Essa sfrutta il diverso valore di pH dei tessuti e degli organelli cellulari.

- **Ematossilina:** l'ematossilina ossidata (ematina¹³) è un colorante basico che tende a legarsi a molecole acide come quelle presenti nel nucleo, conferendogli il caratteristico colore blu-violaceo
- **Eosina:** è un colorante acido che si lega ai granuli eosinofili delle cellule, abbondanti nel citoplasma e nei tessuti muscolari/connettivi conferendogli il caratteristico colore rosa.

¹¹Procedura eseguita mediante ago che permette di prelevare porzioni di tessuto a fine diagnostico

¹²Soluzione di formaldeide acquosa

¹³Si forma quando l'ematossilina incontra ioni metallici

L'ematossilina è un prodotto naturale estratto dal tronco di un albero, ciò rende difficile la standardizzazione su diversi lotti del prodotto. Inoltre il colorante è soggetto a precipitazione che potrebbe indurre variazioni giornaliere all'interno dello stesso laboratorio di analisi. Ad oggi è pratica comune effettuare l'acquisizione digitale dei vetrini tramite scanner prima di procedere con la refertazione, facilitando il lavoro del patologo e la conservazione a lungo termine dell'esame istologico. Le variabilità imposte dal colorante, dal metodo di preparazione e dallo scanner di acquisizione comportano una consistente variabilità di colore tra diverse analisi istologiche (vedi *Fig. 1.6*).

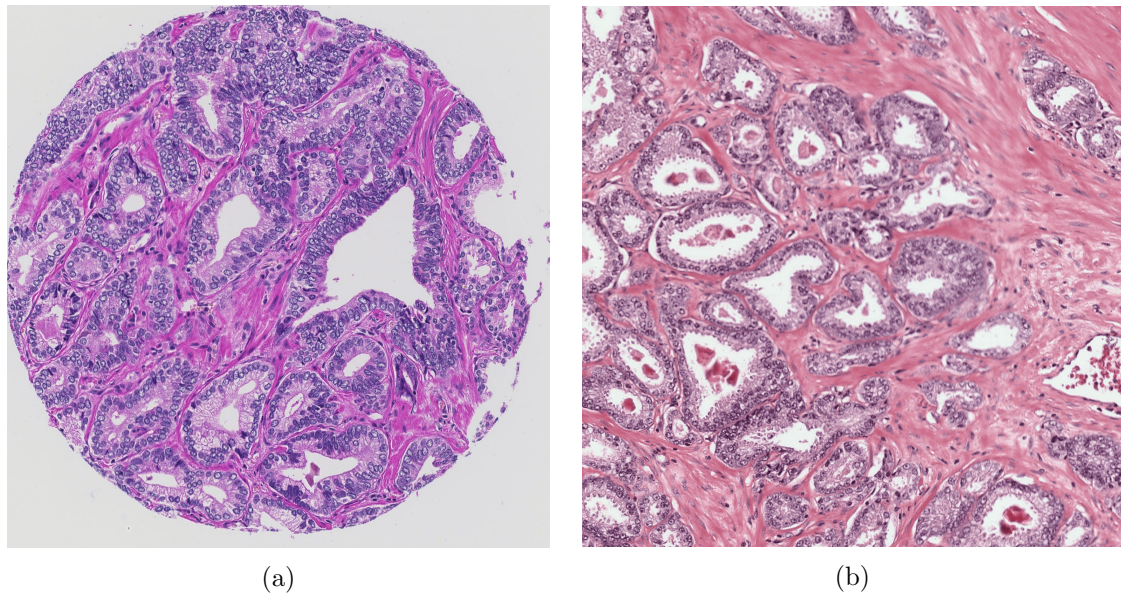


Figura 1.6: (a) Slice ottenuta da un microarray di tessuto. Immagine tratta da Dataset pubblico di Arvaniti E.[1] (b) Porzione di una WSI^a. Immagine ottenuta da Dataset interno.

^aWhole slide image

1.3 Gleason score e Grade Group

Per effettuare la stadiazione dell'adenocarcinoma viene utilizzato il sistema di gradazione di Gleason che, in funzione di determinate caratteristiche architettoniche dei tessuti ghiandolari prostatici, assegna uno score da 1 a 5 ai diversi pattern architettonici. Le linee guida utilizzate per l'assegnazione dello score sono state riviste nel tempo dalla ISUP¹⁴ (vedi *fig. 1.7*).

¹⁴International society of Urological Pathology

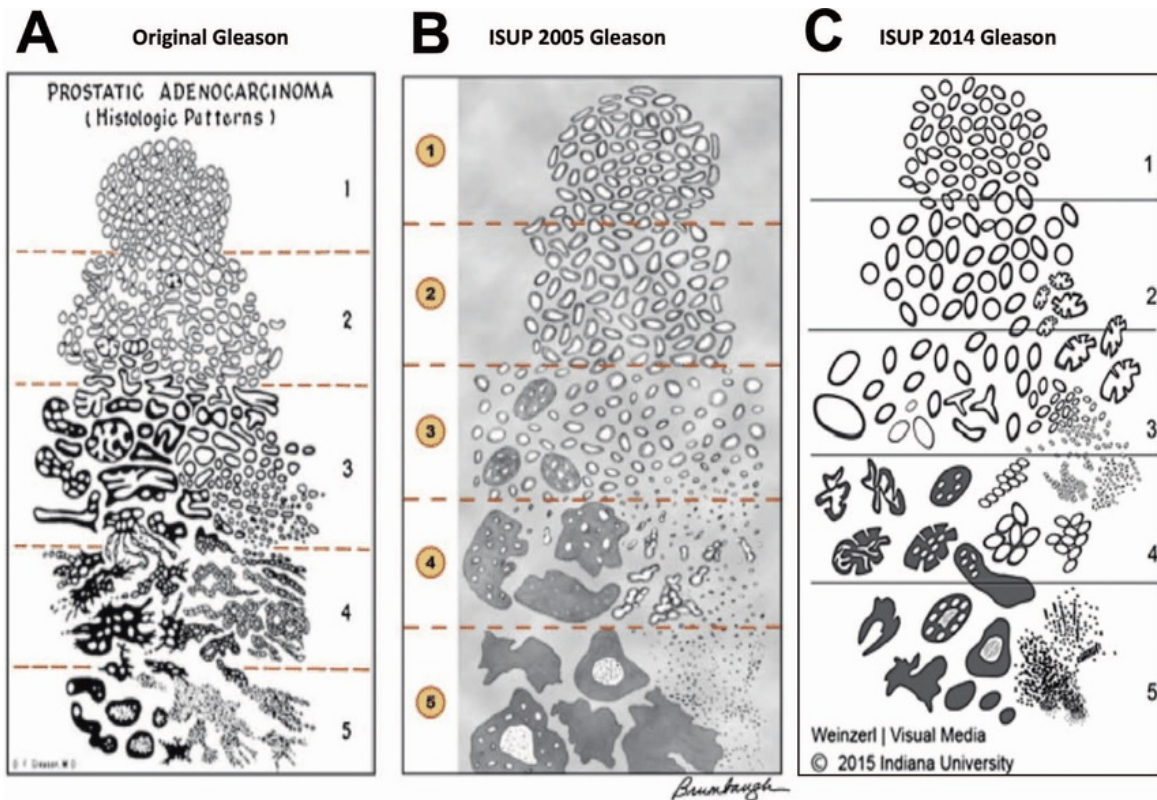


Figura 1.7: Modifiche nel tempo delle caratteristiche istologiche architetturali utilizzate per assegnare il Gleason score. Immagine tratta da [19].

Inizialmente veniva utilizzato unicamente il Gleason score del pattern più presente all'interno del tessuto prostatico. Nel 2016 vi è stata l'introduzione del Gleason Grade Group che si ottiene considerando il Gleason score dei due pattern più rappresentati all'interno dei tessuti analizzati (vedi *tab. 1.2*). Questo indicatore è di fondamentale importanza in quanto indirizza i successivi trattamenti prognostici e terapeutici.

Grade Group	Gleason score	Rischio
Grade Group I	≤ 6	Basso
Grade Group II	7 (3 + 4)	Moderato favorevole
Grade Group III	7 (4 + 3)	Moderato sfavorevole
Grade Group IV	8	Alto
Grade Group V	9-10	Alto

Tabella 1.2: Assegnazione del Gleason Grade Group in funzione dei Gleason score

Quindi l'esperto anatomo-patologo, valutando le caratteristiche strutturali, a diversi ingrandimenti del campione (da 4x a 40x) assegnerà i corrispondenti score ai due pattern più presenti nel campione. Qualora non vi sia un pattern secondario il punteggio sarà due volte il pattern primario identificato e verrà così ottenuto il Gleason Grade Group. Le caratteristiche istologiche corrispondenti ad ogni diverso gruppo sono:

- **GG I:** Alto grado di differenziazione, solo ghiandole singole e ben distinte

- **GG II:** Prevalenza di ghiandole regolari con un piccola componente di ghiandole fuse e/o cribiformi (3+4)
- **GG III:** Prevalenza di ghiandole irregolari, fuse e/o cribiformi con una piccola componente di ghiandole regolari (4+3)
- **GG IV:** Solo ghiandole irregolari, fuse e/o cribiformi (4+4). Prevalenza di ghiandole regolari con una componente minore di tessuto privo di formazione di ghiandole (3+5). Prevalenza di tessuto con assenza di formazione ghiandolare con una minore componente di ghiandole regolari (5+3)
- **GG V:** Assenza di formazione ghiandolare o presenza di ghiandole necrotiche con o senza ghiandole irregolari, fuse e/o cribiformi (5+5, 5+4)

Nella *Fig.1.9* sono mostrate immagini appartenenti al Dataset di Arvaniti con diversi Gleason score e di conseguenza diverso Gleason Grade Group.

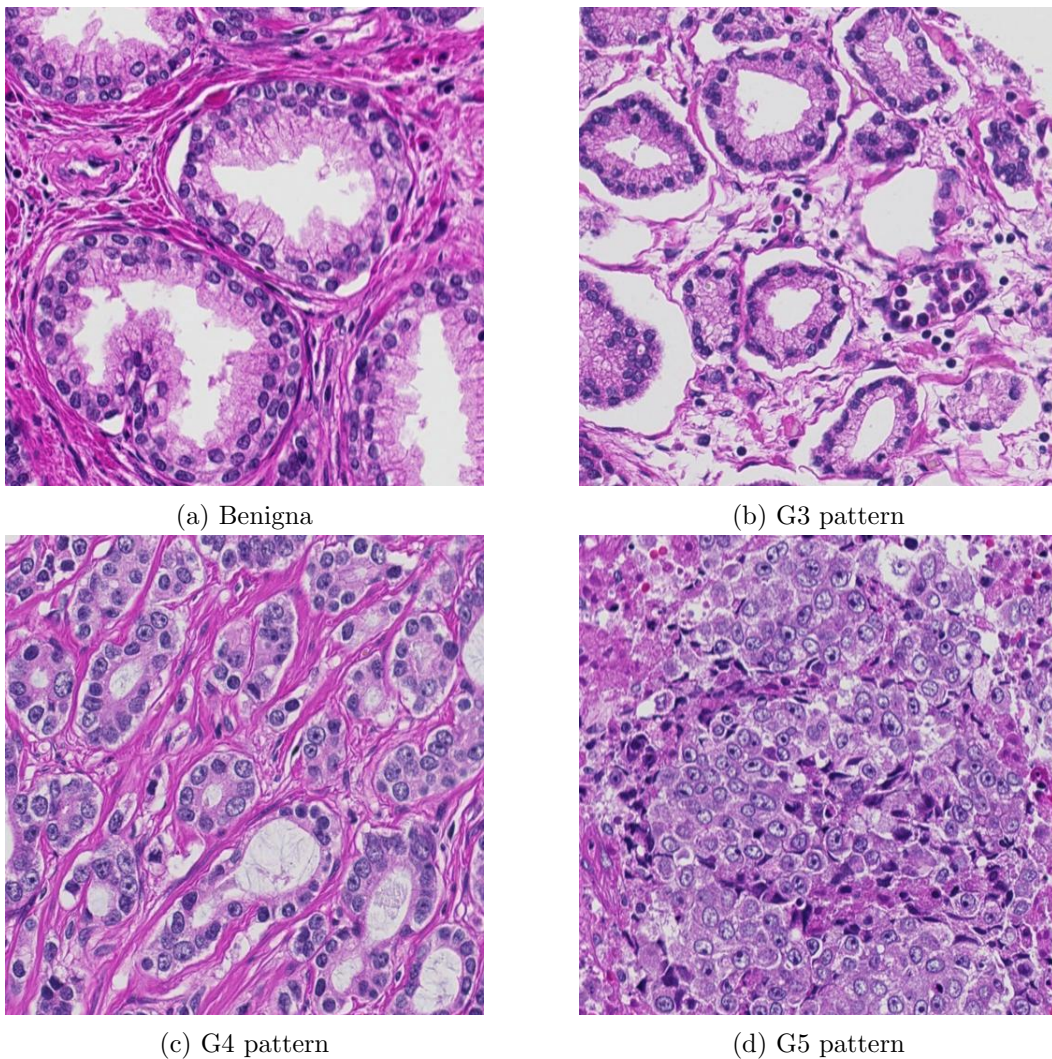


Figura 1.8: Patch contenenti i diversi pattern riscontrabili in istologia prostatica

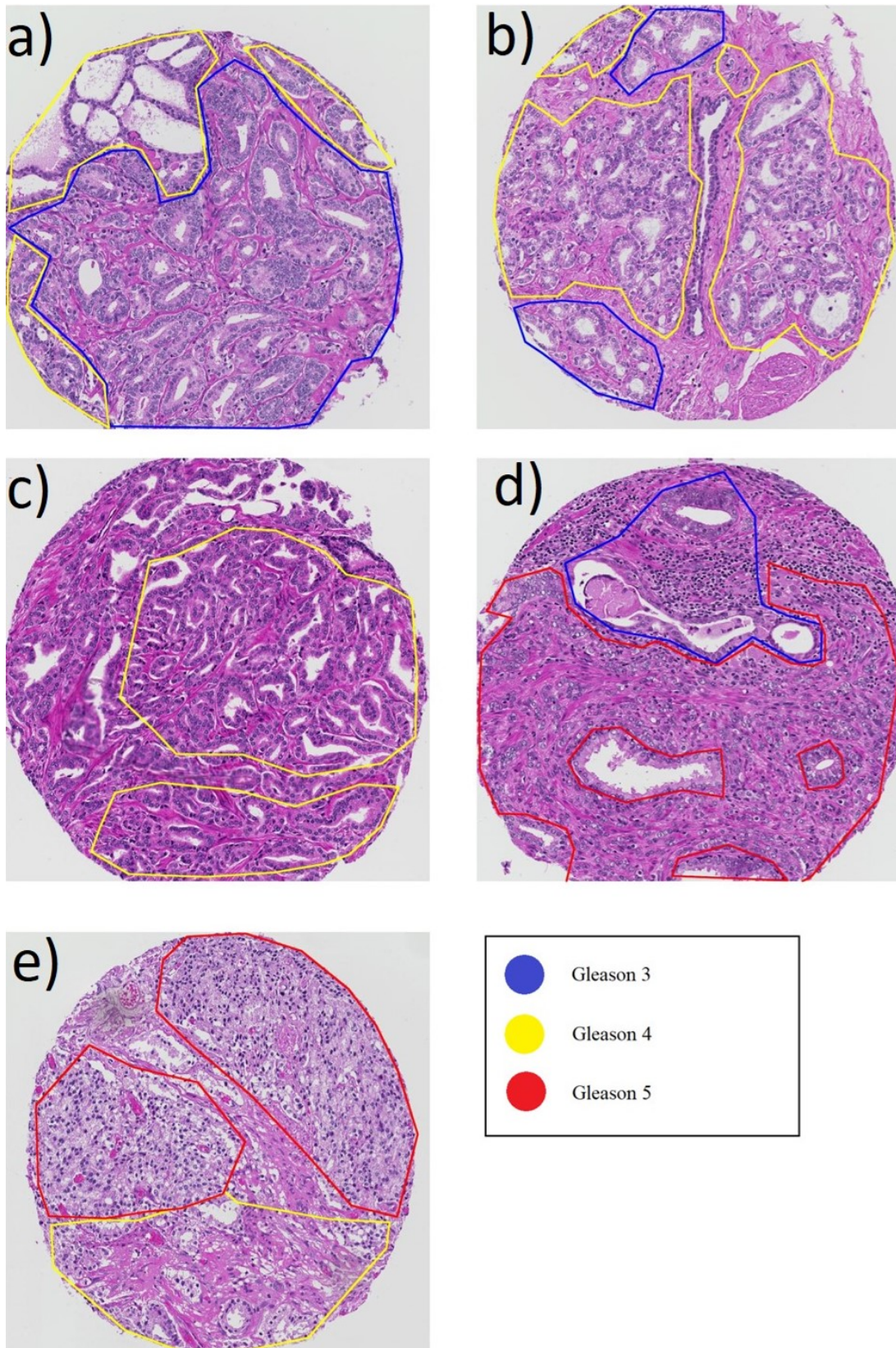


Figura 1.9: Immagini estratte dal Dataset di Gleason andando a sovrapporre i bordi delle annotazioni manuali del patologo 1. I relativi Grade Group sono: a) GG1 (3 + 4), b) GG2 (4 + 3), c) GG3 (4 + 4), d) GG4 (5+3) e f) GG5 (5 + 4)

1.4 Linee guida terapeutiche

In base alla classificazione TMN del tumore, ai risultati ottenuti dall'esame istologico e all'aspettativa di vita del paziente si delinea l'iter terapeutico.

1.4.1 Vigile Attesa

Circa il 40% di pazienti cui viene diagnostico una neoplasia prostatica hanno un'aspettativa di vita inferiore a 10 anni a causa dell'età avanzata o a causa dell'eventuale presenza di comorbidità con una letalità superiore al carcinoma prostatico. Per questa classe di pazienti, per i quali il trattamento del tumore non è in grado di impattare l'aspettativa di vita, si stabilisce una politica di sorveglianza, in assenza di controlli sistematici, volta al trattamento della sintomatologia a puro scopo palliativo.

1.4.2 Sorveglianza attiva

È una strategia di trattamento differito, che viene proposta ai pazienti con rischio di malattia basso-molto basso al momento della diagnosi (GG1-2) che consiste nel monitoraggio continuo: ripetizione periodica di biopsie, visite, analisi del PSA ed eventuali RM multiparametriche al fine di rilevare tempestivamente la progressione della patologia. I vantaggi di tale politica sono a carico del paziente, che ha dei miglioramenti dello stile di vita. Risulta però fondamentale riconoscere in maniera tempestiva la progressione della malattia al fine di avviare trattamenti locali con intento radicale.

1.4.3 Trattamenti locali a scopo radicale

Prostatectomia radicale

Si tratta di una pratica chirurgica di rimozione della prostata, delle ampolle deferenziali e delle vescicole seminali. Ad oggi è comune utilizzare tecniche mini-invasive: laporoscopica o laparoscopica robot-assistita. L'obiettivo principale è quello di eradicare la malattia e se possibile preservare continenza e funzionalità erettile. Le complicazioni legate a tali pratica sono:

- **Intraoperatorie:** perdite ematiche, perforazione della parete rettale e lesioni uretrali
- **Postoperatorie precoci** (< 30gg): trombosi venose, embolia polmonare, complicanze cardiache, infezioni, linforrea¹⁵ e linfocele¹⁶
- **Postoperatorie tardive** (> 30gg): incontinenza, impotenza e stenosi della connessione vescico-uretrale

Tale pratica operatoria è più efficace nei tumori confinati all'interno della ghiandola prostatica (T1 e T2). Diversi studi hanno dimostrato vantaggi in termine di soprav-

¹⁵Legata all'asportazione dei linfonodi, la linfa non viene eliminata e porta alla formazione di lesioni della cute

¹⁶Raccolta di linfa dovuta all'asportazione dei linfonodi

vivenza globale, cancro-specifica, riduzione del rischio di progressione e diffusione a distanza del tumore nei pazienti in classe di rischio intermedia[20].

Radioterapia e Brachiterapia

Si tratta di un'opzione terapeutica con fini radicali. Le tecniche radioterapiche più utilizzate sono: radioterapia a fasci esterni e brachiterapia¹⁷. Lo studio PROTECT[21] ha dimostrato che in pazienti con classe di rischio bassa, molto-bassa e intermedia la radioterapia produce indici di sopravvivenza globale e ricaduta biochimica sovrapponibili alle tecniche chirurgiche. I parametri di azione della radioterapia sono:

- **Dose totale** : si hanno risposte positive sfruttando dosi elevate ($> 75 \text{ Gy}$ ¹⁸)
- **Dose per seduta** : solitamente compresa tra 1-2 Gy
- **Volume di trattamento** : il volume all'interno del quale viene somministrata la dose prefissata

Nel caso di brachiterapia si inseriscono degli impianti permanenti o degli aghi radioattivi realizzati in titanio che sfruttano il decadimento radioattivo degli isotopi dello Iodio e del Palladio per il trattamento del tessuto. Di solito la radioterapia viene coadiuvata da un trattamento ormonale di deprivazione degli androgeni che inizia mesi prima dell'inizio del trattamento e prosegue fino al termine dello stesso. La radioterapia può inoltre essere utilizzata come trattamento post-chirurgico per diminuire il rischio di recidiva.

1.5 Statistiche del carcinoma

Il tumore della prostata è la prima neoplasia maligna più diffusa in Italia per il genere maschile (35.000 nuovi casi nel 2020) ed è al terzo posto per mortalità (dietro a polmone e colon-retto). Il tasso di sopravvivenza a 5 anni dalla diagnosi vale circa il 91% ed è in costante aumento mentre la mortalità è in costante diminuzione (-1,9% ogni anno) grazie a:

- Diffusione dello screening opportunistico basato sul PSA
- Maggiore sensibilizzazione della popolazione maschile riguardo a tale problematica
- Miglioramento delle linee guida al trattamento terapeutico

Una delle problematiche di maggior interesse riguardanti il tumore alla prostata è il tasso di sovradiagnosi e sovratrattamento[22] causato in parte dall'utilizzo dell'analisi del PSA come screening opportunistico.

¹⁷Posizionamento della sorgente radioattiva in posizione adiacente al tessuto da trattare

¹⁸Gray

1.6 Finalità dello studio

Il lavoro di tesi svolto mira a valutare l'efficacia di un sistema automatico di analisi delle istologie prostatiche basato su reti neurali convoluzionali che possa in futuro essere utilizzato come strumento di supporto decisionale dagli anatomo-patologi per una corretta e riproducibile stadiazione della malattia. In *Fig. 1.10* è mostrato come il sistema proposto possa essere integrato all'interno del processo diagnostico.

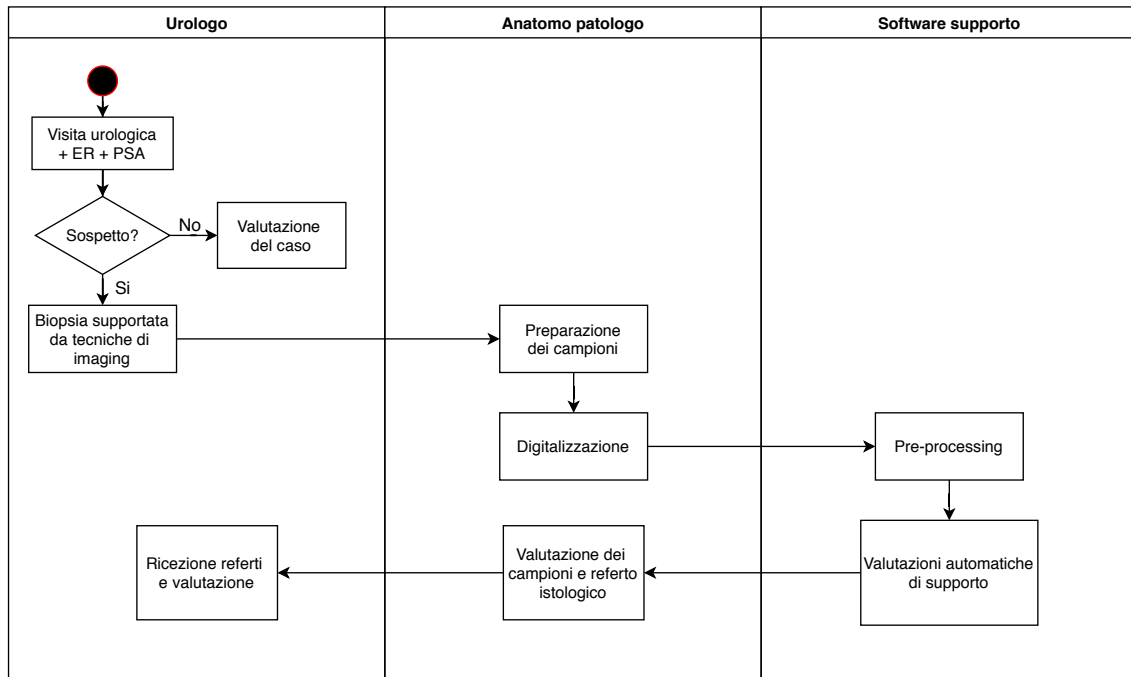


Figura 1.10: Schema di funzionamento di un processo diagnostico che integri al suo interno il sistema sviluppato

I vantaggi legati all'implementazione di tale tecnologia sono:

- Minore variabilità tra acquisizioni digitali differenti grazie all'utilizzo algoritmi di normalizzazione delle immagini
- Processi diagnostici più rapidi con relativi benefici per i pazienti
- Supporto decisionale all'anatomo patologo
- Diminuzione del carico di lavoro dei centri di anatomia patologica

Nell'ipotesi finale che tale sistema venga implementato all'interno di un software proprietario potrebbe essere aggiunta la possibilità di richiedere ulteriori consulti da parte di altri specialisti, migliorando così la qualità diagnostica dell'esame.

Capitolo 2

Stato dell'arte

Il campo della Digital Pathology ha subito un estremo sviluppo negli ultimi 10 anni grazie all'avvento delle reti neurali convoluzionali che hanno rivoluzionato il campo dell'analisi delle immagini.

2.1 Reti Neurali Convoluzionali

Una rete neurale convoluzionale (CNN) è un tipo di rete neurale artificiale profonda "feed-forward", che emulando delle caratteristiche dei neuroni che compongono la corteccia visiva è utilizzata per la classificazione di immagini. Essa è tipicamente composta da 3 strutture principali:

- **Blocco recettivo input:** si tratta del primo strato della CNN, che ha funzionamento analogo ai neuroni corticali che ricevono l'input visivo
- **Struttura elaborazione nascosta:** una successione di blocchi composti da un layer convoluzionale ed uno di pooling. La sua funzione è di estrarre delle informazioni/caratteristiche dalle immagini che verranno utilizzate dal blocco di classificazione per generare l'output.
- **Blocco classificazione di output:** è composta da uno o più strati FC¹

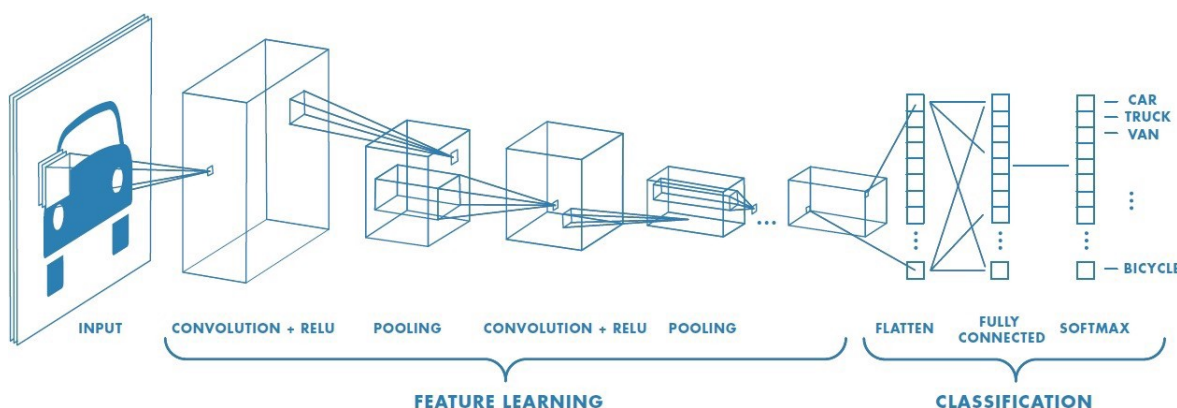


Figura 2.1: Schema architetturale classico di una rete neurale convoluzionale. Immagine tratta da [23].

¹Fully connected

Layer Convoluzionale

La convoluzione è un'operazione che tramite l'utilizzo di un kernel (filtro) che viene fatto scorrere sulla matrice di input genera un valore di output eseguendo il prodotto scalare tra il kernel e i corrispondenti punti della matrice di partenza (vedi Fig. 2.2). I coefficienti dei filtri sono parametri incogniti della CNN che dovranno essere ottimizzati durante la fase di training. Ogni strato convoluzionale possiede n filtri pari alla terza dimensione dello strato precedente[24].

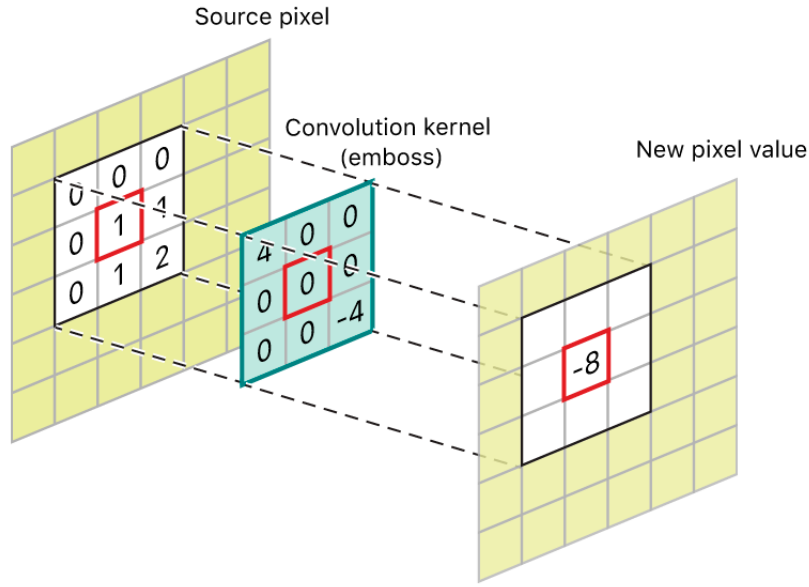


Figura 2.2: Rappresentazione visiva dell'operazione di convoluzione

Immediatamente dopo un layer convoluzionale è presente un layer di attivazione non-lineare che ha lo scopo di andare ad aggiustare o limitare gli output del layer convoluzionale. Le funzioni di attivazione più utilizzate sono: sigmoide, tangente iperbolica e ReLU²

La convoluzione e di conseguenza la dimensione degli output di un layer è regolata da due diversi parametri:

- **Stride**: passo tra le diverse operazioni di convoluzione, mi consente di diminuire la dimensione degli output
- **Padding**: consiste nell'andare ad aggiungere dei valori nulli su tutte l'estremità dell'immagine, al fine di mantenere le informazioni contenute nel bordo e poter mantenere la dimensione di partenza per l'output

È dunque possibile calcolare la dimensione di output di uno strato convoluzionale come:

$$O = 1 + \frac{N + 2P - F}{S} \quad (2.1)$$

²Rectified linear unit, $f(x) = \max(0, x)$

Nell' *eq.2.1* le lettere indicano:

- **O**: dimensione matrice output
- **N**: dimensione matrice input
- **P**: padding
- **F**: dimensione del filtro
- **S**: stride

Layer di Pooling

Il concetto che sta alla base di quest'operazione come si può osservare in *Fig.2.2* è il sotto-campionamento. È un'operazione molto simile a quella in analisi delle immagini di diminuzione della risoluzione. Di fatto si analizzano delle sottoregioni dell'output ottenuto dal layer convoluzionale sfruttando delle operazioni di Max-Pooling (output è il massimo della sottoregione) o Average-Pooling (output è la media della sottoregione).

Layer fully connected

È una struttura simile a quella delle comuni reti neurali tradizionali. In questa struttura ogni nodo è direttamente connesso a tutti i nodi del layer precedente e successivo tramite delle connessioni. Ognuna di queste connessioni rappresenta un parametro da allenare della rete. Una tecnica molto utile da utilizzare per andare a velocizzare l'allenamento di questa tipologia di strati è la tecnica Dropout che consiste nell'andare ad eliminare durante l'allenamento con una certa probabilità determinati neuroni e determinate connessioni[24].

Apprendimento di una CNN

Ogni layer è caratterizzato da un numero differente di neuroni, le unità fondamentali di una rete neurale. Ognuno di questi neuroni della struttura è caratterizzato da un vettore pesi w e un bias b . In funzione degli input x il suo output z sarà:

$$z = wx + b \quad (2.2)$$

Al quale verrà applicata una funzione di attivazione σ ed y sarà l'output del layer di attivazione:

$$y = \sigma(z) \quad (2.3)$$

Così gli input x , i dati del training set, saranno propagati di neurone in neurone, di layer in layer fino all'uscita della rete neurale dove otterremo il valore predetto dalla rete y . Noto il valore atteso \hat{y} è possibile definire una funzione di perdita (loss function):

$$J(W, b) = \frac{1}{n} \sum_{i=1}^n L(\hat{y}_i, y_i) \quad (2.4)$$

L'obiettivo dell'apprendimento sarà dunque quello di minimizzare tale funzione mantenendo una buona capacità di generalizzazione e minimizzando l'overfitting³. Minimizzare la funzione di costo, corrisponde a trovare i vettori pesi e bias dei diversi neuroni che ottimizzano il problema. Esistono diversi algoritmi di ottimizzazione che implementano diverse tecniche di correzione dei pesi e dei bias. Una delle più utilizzate è lo SGD⁴ che sfrutta la derivata della funzione di costo e un parametro chiamato learning rate α per propagare in direzione retrograda le correzioni:

$$w' = w - \alpha \frac{dJ(w, b)}{dw} \quad (2.5)$$

$$b' = b - \alpha \frac{dJ(w, b)}{db} \quad (2.6)$$

Il learning rate è uno degli iperparametri che regola l'apprendimento di una rete neurale. Un alto valore di learning rate porta ad overfitting mentre un valore troppo basso non consente alla rete di ottenere prestazioni sufficienti. Nelle tecniche di ottimizzazione più raffinate non si utilizza un valore costante di learning rate, ma si pianifica un decremento nel tempo volto a diminuire il fenomeno dell'overfitting. Una altra tecnica utilizzata per diminuire l'overfitting è la regolarizzazione. Si modifica la funzione di costo introducendo termine di regolarizzazione R e un coefficiente di regolarizzazione costante λ :

$$J(W, b) = \frac{1}{n} \sum_{i=1}^n L(\hat{y}_i, y_i) + \lambda R \quad (2.7)$$

Il parametro R è funzione dei vettori pesi dei neuroni e in base alla regolarizzazione utilizzata:

$$L1 : R = \sum_{i=1}^m |w_i| \quad (2.8)$$

$$L2 : R = \sum_{i=1}^m w_i^2 \quad (2.9)$$

Questi termini di regolarizzazione vanno a sommare una penalità alla funzione di costo che è direttamente collegata ai valori dei vettori pesi dei neuroni. L'effetto globale di queste tecniche è quello di forzare i vettori pesi dei neuroni ad abbassare i propri valori vicino allo zero.

³Adattamento eccessivo di una rete ai dati di allenamento

⁴Stochastic gradient descent

2.2 Valutazione del Gleason Score tramite CNN

2.2.1 Nagpal

Nello studio di Nagpal [25] viene proposto un sistema a doppio livello (vedi *Fig. 2.3*) che sfrutta:

- **CNN** : per ottenere i gleason pattern dalle varie patch estratte da una WSI
- **Classificatore k-NN⁵** : per ottenere la valutazione del Grade Group globale della WSI

Il dataset utilizzato contiene annotazioni di un patologo dei diversi Gleason pattern presenti e del Gleason Grade di ogni WSI. Per migliorare la capacità di generalizzazione della CNN sfrutta una tecnica chiamata Hard Negative Mining, che consiste nell'andare a dare un maggior peso durante la fase di training alle immagini più difficili da classificare.

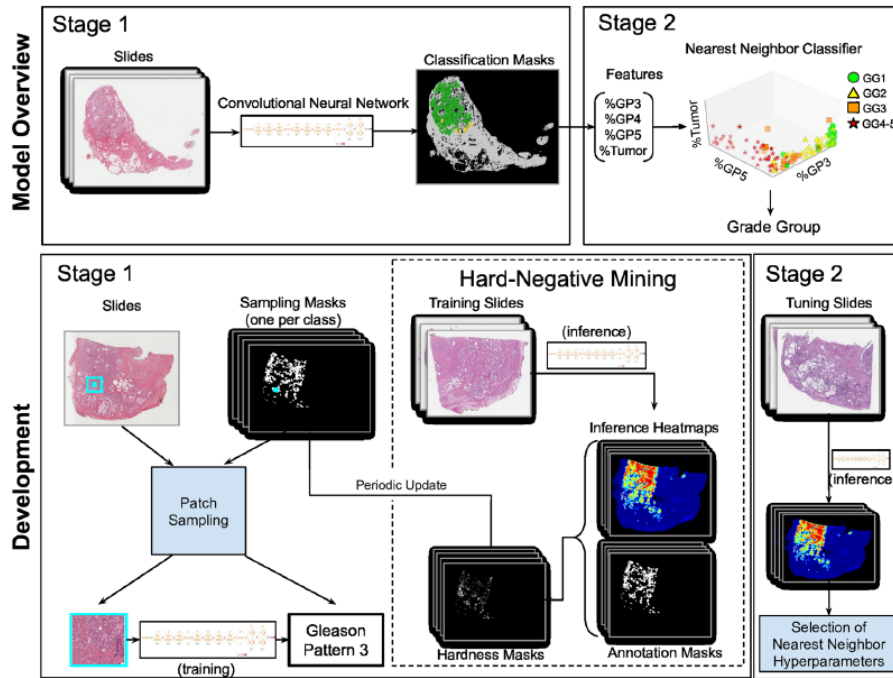


Figura 2.3: Struttura del sistema proposto da Nagpal

⁵k-Nearest neighbors

2.3 Studi finalizzati alla classificazione binaria

2.3.1 Campanella

Campanella ed il suo team [29] a partire da un dataset di grandi dimensioni 12.160 WSI (25 volte più pixel di quelli contenuti nell'ImageNet dataset⁸) allenano un modello utilizzando solo annotazioni di alto livello sull'intera slide. Le loro immagini sono interamente ottenute da biopsie prostatiche e non da prostatectomie, il che fornisce un maggior esempio di ghiandole benigne e di conseguenza una maggiore robustezza al classificatore. Utilizzano un modello MIL⁹. La classificazione viene effettuata analizzando le diverse patch una per volta (instance) ed aggregando tutti i risultati per ottenere il risultato su tutta la slide (bag). Dimostrano sperimentalmente quanto l'errore di classificazione sia collegato alla dimensione del dataset utilizzato per l'allenamento (*vedi Fig. 2.5*).

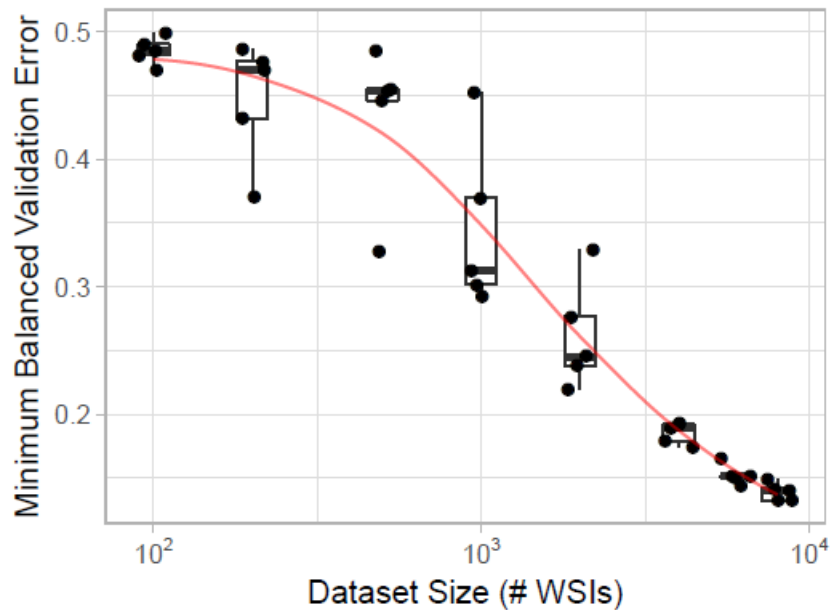


Figura 2.5: Andamento dell'errore minimo di validazione in funzione della dimensione della dimensione del dataset. Immagine tratta da [29]

Successivamente Campanella propone un miglioramento del suo sistema[30] utilizzando una RNN¹⁰ che aggregando le probabilità predette dalla rete convoluzionale per le singole istanze effettua la classificazione finale (*vedi Fig. 2.6*).

⁸Uno dei dataset più importanti nell'ambito del deep learning utilizzato spesso per il trasferimento learning

⁹Multiple instance learning: solo l'annotazione di alto livello sulla slide è necessaria

¹⁰Recursive neural network

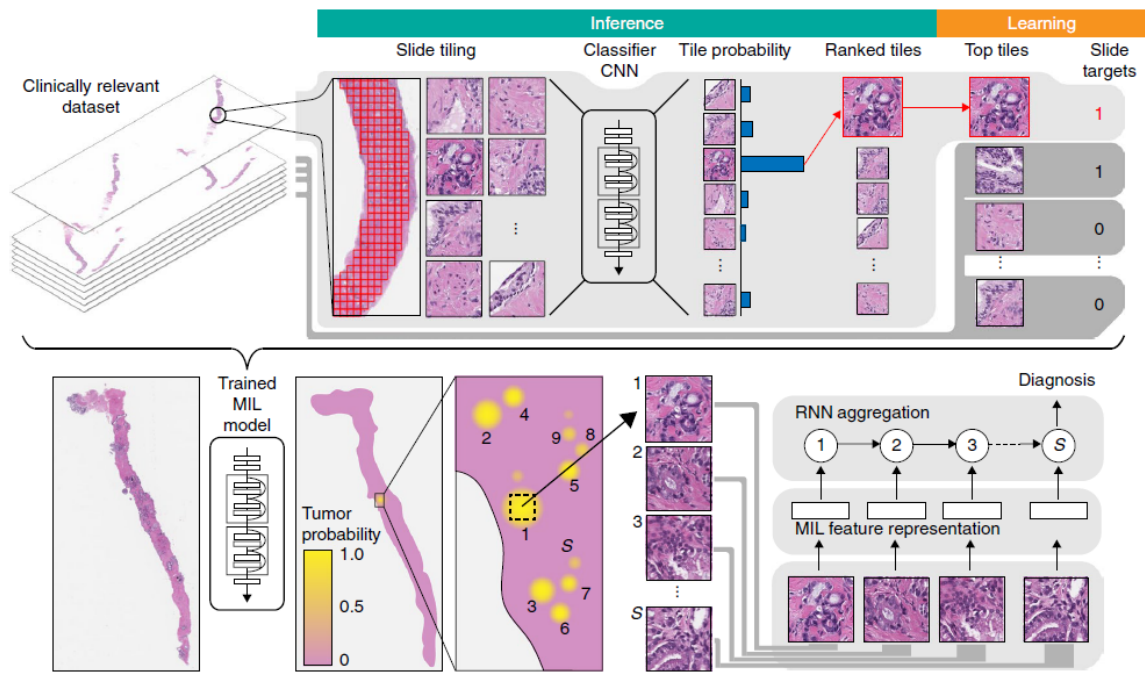


Figura 2.6: Sistema sviluppato da Campanella per la classificazione binaria di istologie prostatiche. Immagine tratta da [30]

Tabella 2.1: Tabella riassuntiva

Autore	Nagpal	Bulten	Arraniti	Campanella
Scopo	Grade group Prostatectomie 1243 WSI	Grade group Biopsie 1143 WSI	Grade group Prostatectomie 886 TMA spot	Benigno vs Maligno Biopsie 12160 WSI
Dataset				
Ingrandimento				
Pre-processing				
Data augmentation				
Transfer learning	Si	Si	Si	Si
Reti testate	Inception V3	U-Net	Inception-V3, ResNet-50, MobileNet	AlexNet, Resnet, VGG-11
Risultati	AUC(GG;2) = 0.95	AUC(binario)=0.99	Cohen K vs Pato.1 = 0.55 Cohen K vs Pato.2 = 0.49 Cohen K P1 vs P2 = 0.67	2018: AUC = 0.967
	AUC(GG;3) = 0.96			2019: AUC = 0.991
	AUC(GG;4) = 0.95			

Capitolo 3

Metodi

3.1 Dataset

Per lo svolgimento della tesi è stato sfruttato il dataset pubblico di Arvaniti E.[1]. Esso comprende 5 tissue microarrays inizialmente composti da 200-300 campioni di tessuto l'uno trattati con la colorazione ematossilina-eosina. In seguito sono stati eliminati tutti i campioni contenenti artefatti o tessuto non prostatico. Un patologo ha annotato l'intero dataset classificando i diversi pattern tumorali se presenti. Inoltre il TMA80, utilizzato come test set, è stato annotato da un secondo patologo per poter effettuare delle valutazioni di variabilità inter-operatore. I vari campioni sono stati digitalizzati mediante scanner NanoZoomer-XR con una risoluzione a 40x ($0,23\mu m$ per pixel). Inizialmente si è valutato se utilizzare il Dataset sfruttando la risoluzione originale o se effettuare un ricampionamento. Per aumentare le prestazioni delle CNN si è ricampionato l'intero dataset ad un ingrandimento di 20x ($0,46\mu m$ per pixel).

Primario	Secondario				
	Pattern	Ben	G3	G4	G5
	Ben	103	0	0	0
	G3	0	201	46	0
	G4	0	25	138	11
G5	0	2	11	95	

Primario	Secondario				
	Pattern	Ben	G3	G4	G5
	Ben	12	0	0	0
	G3	0	78	25	1
	G4	0	20	85	0
G5	0	1	0	16	

Tabella 3.1: Nella tabella di *sinistra* la distribuzione delle immagini del construction set (training e validation). Nella tabella di *destra* la distribuzione della immagine del test set del Patologo 1.

Sono state scartate 9 immagini dal construction set e 7 dal test set in quanto non superavano la soglia minima di annotazione manuale ottenendo così un totale di 632 immagini per il construction set e di 238 per il test set.

3.2 Organizzazione del lavoro

Il lavoro di tesi svolto è stato suddiviso in due fasi:

- **Ottimizzazione dei dati** : ricerca del miglior ingrandimento, dimensione delle patches, valutazione dell'impiego di pre-processing, estrazione smart delle patches e data augmentation
- **Ottimizzazione delle strutture** : valutazione di diverse architetture di CNN, ottimizzazione degli iperparametri di training e valutazione dell'utilizzo di diverse configurazioni di CNN in cascata

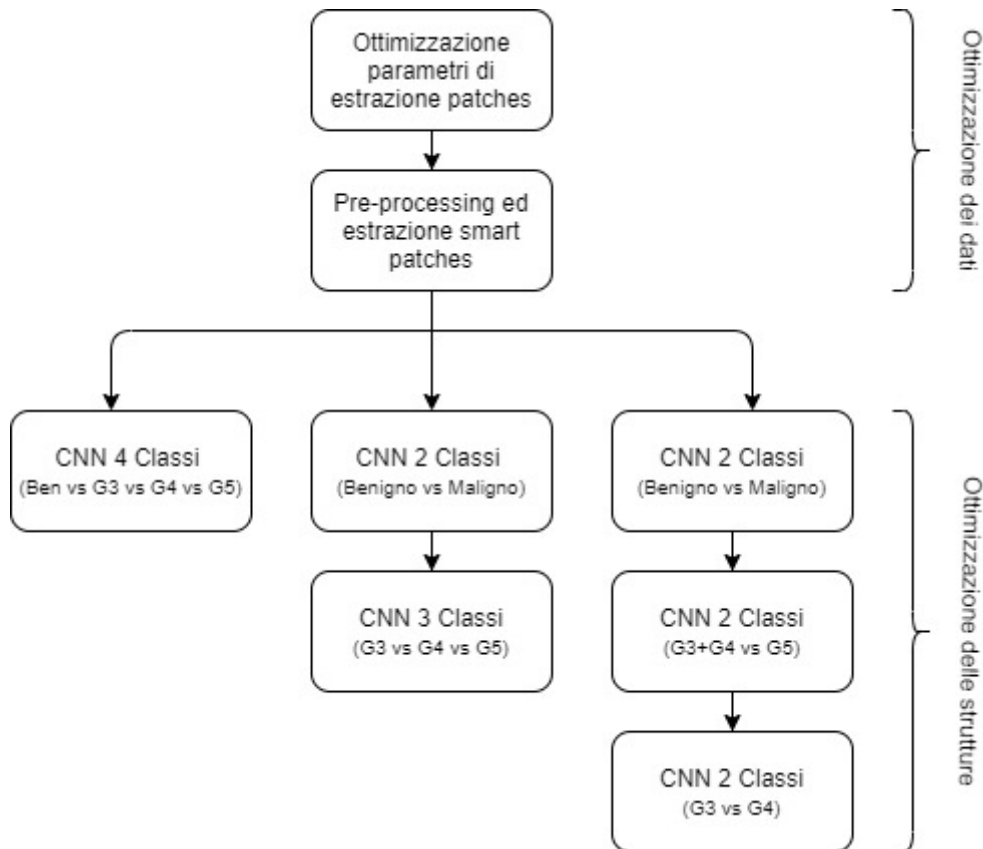


Figura 3.1: Schema di sviluppo del lavoro

3.3 Ottimizzazione dei dati

Nel campo del deep learning il corretto funzionamento di un classificatore dipende principalmente dalla qualità dei dati utilizzati per il suo allenamento. Le caratteristiche che deve avere un buon training set sono:

- Omogeneità
- Bassa variabilità tra elementi appartenenti alla stessa classe
- Alto numero di campioni per ogni classe
- Contenere tutti i diversi prototipi delle diverse classi

Estrazione delle patches

I parametri di estrazione delle patches giocano un ruolo fondamentale in quanto determinano il grado di informazione contenuto dalle diverse patches ed il numero di elementi che andrà a comporre i diversi set. Nel campo dell'analisi di istologie prostatiche, essendo la classificazione, basata unicamente su caratteristiche architettoniche, è fondamentale che la scala di dimensione utilizzata consenta alle diverse patches di contenere al proprio interno le caratteristiche necessarie alla rete per la loro classificazione (vedi *Fig. 3.2*).

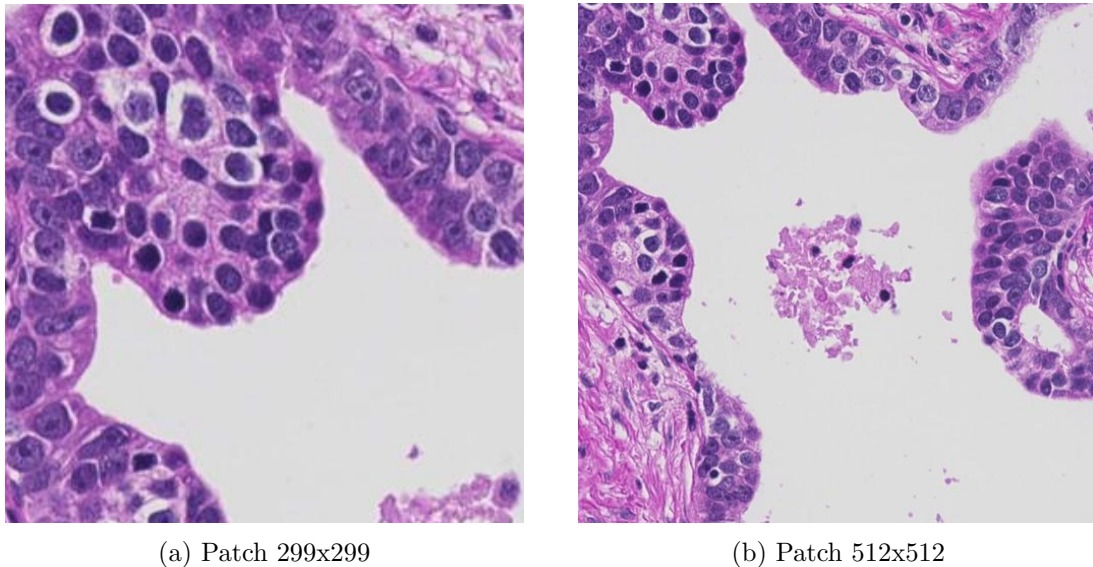


Figura 3.2: Patches estratte nella zona di tessuto, con dimensione differenti. La patch (a) di dimensione 299x299 non riesce a catturare l'intera architettura della ghiandola prostatica.

Un ulteriore parametro fondamentale durante l'estrazione delle patches è l'overlap¹ che regola in maniera diretta il numero di patches estratte da ogni immagine. Un suo valore errato può determinare:

- **Troppo basso** : un numero insufficiente di esempi per allenare correttamente il sistema
- **Troppo elevato** : una scarsa capacità di generalizzazione del sistema dovuta all'eccessivo adattamento ai dati ridondanti del training set

Al fine di bilanciare i due possibili effetti è stato scelto un valore di overlap pari al 50% ed una dimensione delle patches di 512x512 pixels.

¹Percentuale di sovrapposizione tra patches adiacenti

Pre-processing

Nel campo della digital pathology è comune utilizzare delle tecniche di pre-processing che permettano di normalizzare un'immagine sfruttandone una di riferimento. Come detto in precedenza la colorazione ematossilina-eosina presenta una variabilità intrinseca di trattamento, di conservazione e di acquisizione digitale dei campioni. Per uniformare lo stain vector delle diverse immagini è stato utilizzato il metodo di normalizzazione proposta da Macenko et al.[31]. Prima di tutto si calcola a partire dall'immagine RGB la densità ottica dell'immagine:

$$OD = -\log_{10}(I) \quad (3.1)$$

Con la matrice RGB dell'immagine normalizzata tra $[0,1]$. Si sfrutta la OD in quanto rende separabili linearmente i due componenti di colore dell'immagine (vedi Fig. 3.3).

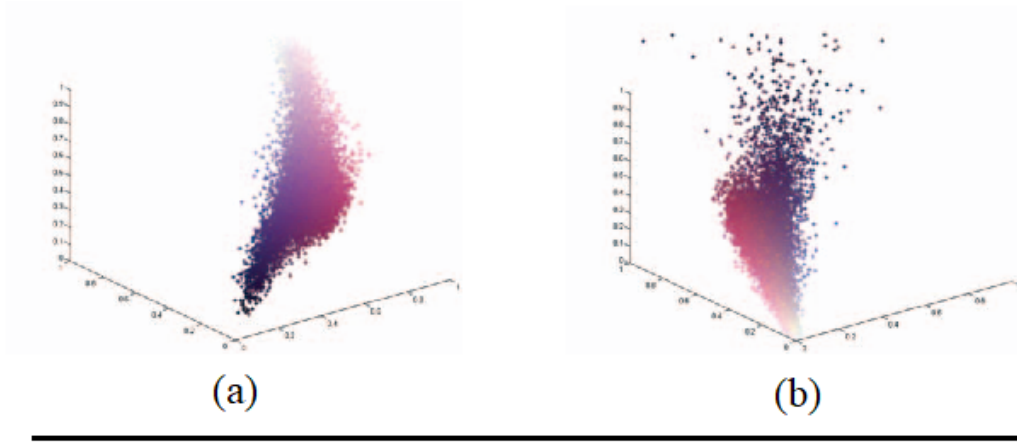


Figura 3.3: a) Valori dei pixel nello spazio RGB. Come si può notare le due colorazioni risultano separabili ma tramite una curva. b) Valori dei pixel nello spazio OD. Le due colorazioni risultano separabili tramite una relazione lineare. Immagine tratta da [31]

Note le relazioni:

$$OD = WH \quad (3.2)$$

$$H = W^{-1}OD \quad (3.3)$$

Dove W è la matrice del vettore di stain ed H è la matrice di concentrazione di ogni stain. Per procedere alla normalizzazione dell'immagine bisogna ottenere una stima della matrice V . Prima di tutto si effettua un thresholding della matrice OD per rimuovere i pixel dove nessuna colorazione ha aderito. Si fattorizza la matrice OD tramite una tecnica chiamata SVD (scomposizione in valori singolari). Si genera allora un piano utilizzando i due vettori corrispondenti ai due valori singolari maggiori ottenuti tramite SVD e si proiettano tutti i valori dei pixel su questo piano normalizzandoli. Vengono calcolati in seguito i diversi angoli tra le direzioni di questi punti e la direzione principale SVD. Si valutano successivamente i due estremi (massimo e minimo) come 99 ed 1 percentile. In seguito i due estremi vengono riconvertiti nel piano OD, ottenendo così la matrice W . Utilizzando la relazione (3.3) si ottiene la

matrice H . Nella pratica per procedere alla normalizzazione si sfrutta un'immagine target della quale tramite il procedimento precedente si ottengono le matrici W ed H . Si ripete lo stesso procedimento sull'immagine da normalizzare ottenendo W_S e H_S . In seguito si effettua la normalizzazione della matrice delle concentrazioni H_S utilizzando gli pseudo massimi per righe (99 percentile) delle due matrici delle concentrazioni (3.4). In seguito si ottiene l'OD normalizzata moltiplicando la matrice delle concentrazioni normalizzata per la matrice degli stain dell'immagine target (3.5). In seguito si ottiene l'immagine RGB normalizzata (3.6).

$$H_S^{norm}(i, :) = \frac{H_S(i, :)}{H_S^{99th}(i, :)} H^{99th}(i, :) \quad (3.4)$$

$$OD_S^{norm} = W H_S^{norm} \quad (3.5)$$

$$I_S^{norm} = 10^{-OD_S^{norm}} \quad (3.6)$$

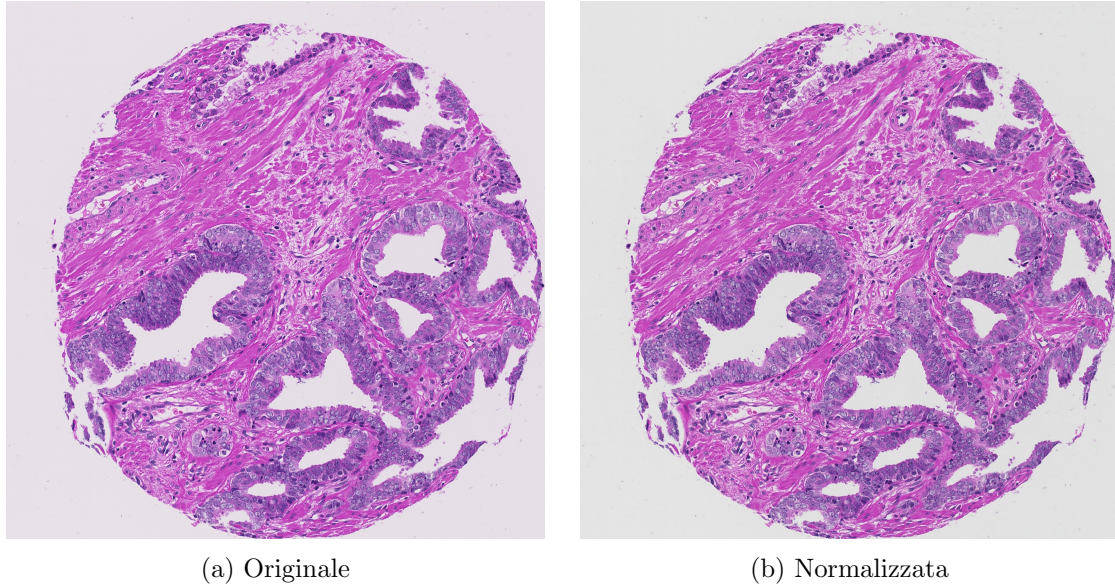
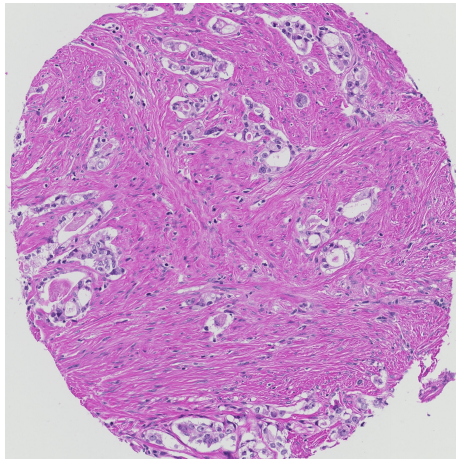


Figura 3.4: Esempio pre-processing. Le variazioni sono minime poichè le istologie sono state acquisite nello stesso centro, con lo stesso scanner e lo stesso lotto di colorante.

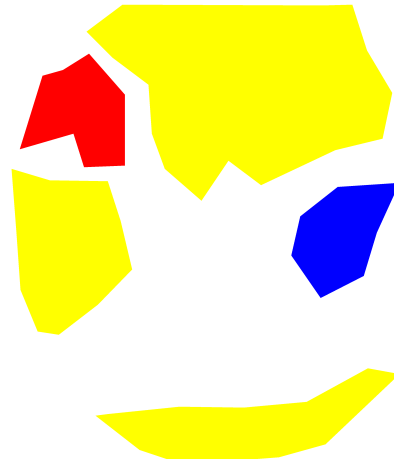
3.4 Generazione delle maschere

A partire dalle immagini normalizzate, tramite appositi algoritmi di segmentazione già sviluppati da studenti negli anni precedenti, sono state ottenute delle maschere:

- **Bianchi**
- **Zone da non considerare** : si ottiene eliminando piccoli artefatti dalla maschera dei bianchi
- **Nuclei** : fondamentali per identificare le ghiandole prostatiche
- **Stroma** : per evitare di considerare come benigne zone di tessuto ad alto contenuto di stroma



(a) Immagine normalizzata



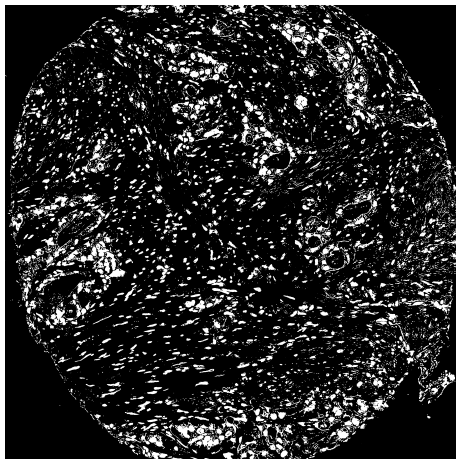
(b) Maschera Patologo 1



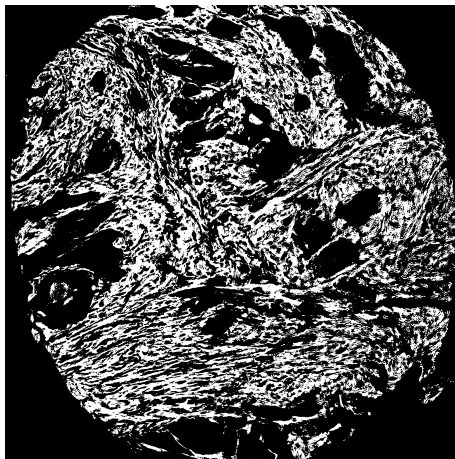
(c) Maschera dei bianchi



(d) Maschera zone non considerare



(e) Maschera dei nuclei



(f) Maschera dello stroma

Figura 3.5: Nella figura sono mostrate l'immagine 62 e la relativa maschera del patologo più tutte le maschere generate per ottimizzare l'estrazione delle patches. Come si può notare valutando le sottoimmagini (b), (e) ed (f) si ha una maggiore concentrazione di nuclei nelle regioni annotate dal patologo mentre le concentrazioni più elevate di stroma si trovano al di fuori delle annotazioni.

3.5 Estrazione smart delle patches

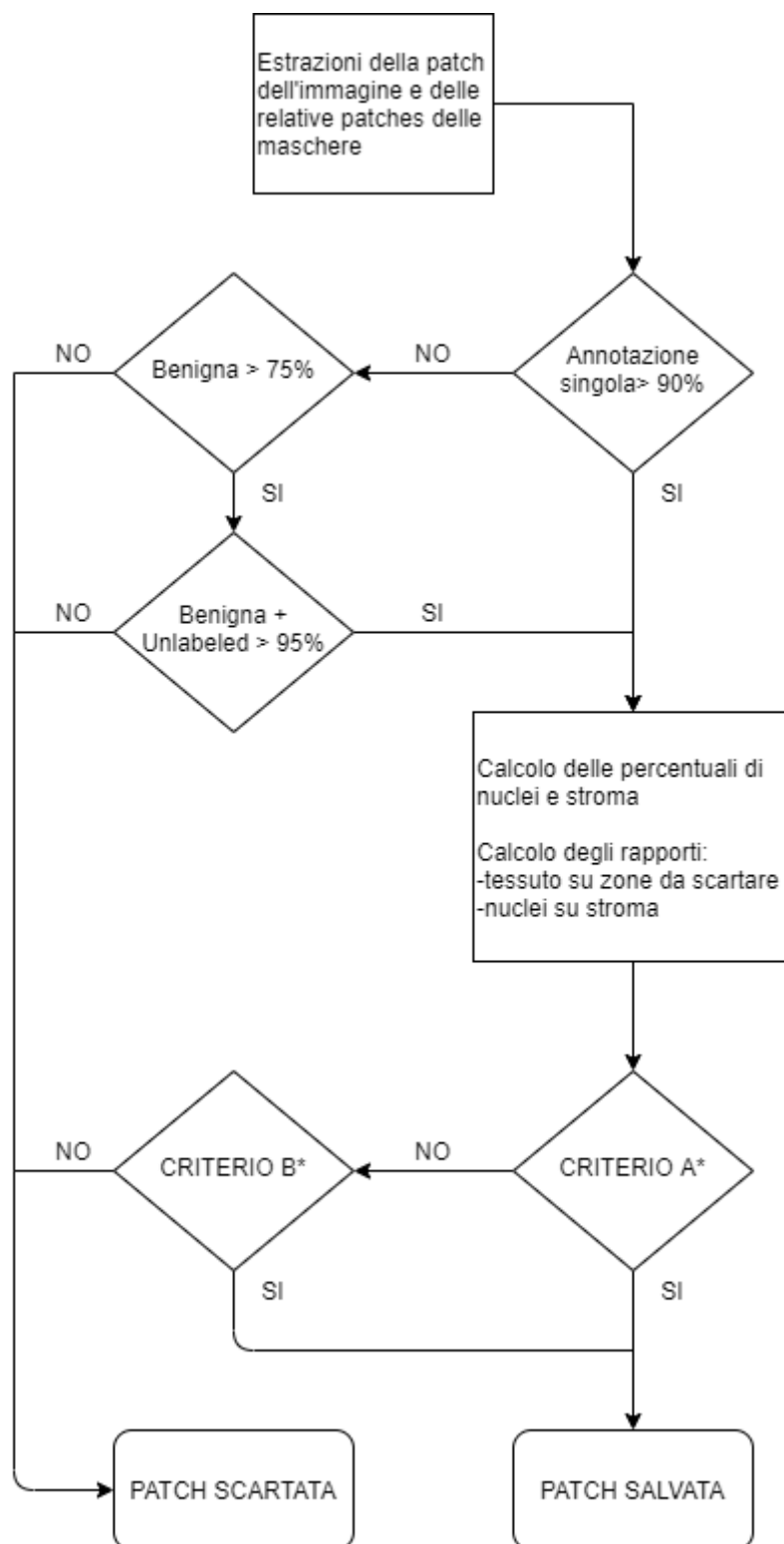


Figura 3.6: Flowchart dell'algoritmo di estrazione delle patches

Il primo criterio scelto, fondamentale per la costruzione di un training set robusto, è quello della percentuale minima di annotazione singola all'interno della patch. Si è utilizzata una soglia molto selettiva del 90%. Nel caso in cui l'annotazione sia benigna ($> 75\%$) si ammette una percentuale massima di patch non classificata del 20% per garantire un maggior numero di esempi della classe benigna (la meno rappresentata). Vengono calcolati successivamente degli score ottenuti a partire dalle maschere ottenute in precedenza. Le condizioni di ammissibilità si basano su questi score:

- **CRITERIO A*** : valuta la quantità di tessuto in relazione ai bianchi (20% massimo di bianchi), il rapporto nuclei su stroma ($> 10\%$) e la quantità di nuclei ($> 15\%$). Garantisce l'estrazione delle ghiandole prostatiche presenti.
- **CRITERIO B*** : utilizza una soglia meno selettiva in relazione alla quantità di bianco presente (3 volte inferiore) ma più selettiva rispetto allo score nuclei su stroma ($> 25\%$) e percentuali di nuclei ($> 25\%$).

La scelta delle soglie è stata ottimizzata per garantire un equilibrio tra selettività dell'algoritmo e quantità di patches estratte.

Classe	Annotazione manuale	Estrazione smart
Ben	794	501
G3	2388	2149
G4	1661	1468
G5	867	702

Tabella 3.2: Numero di patches estratte dal construction set per ogni classe senza e con estrazione smart

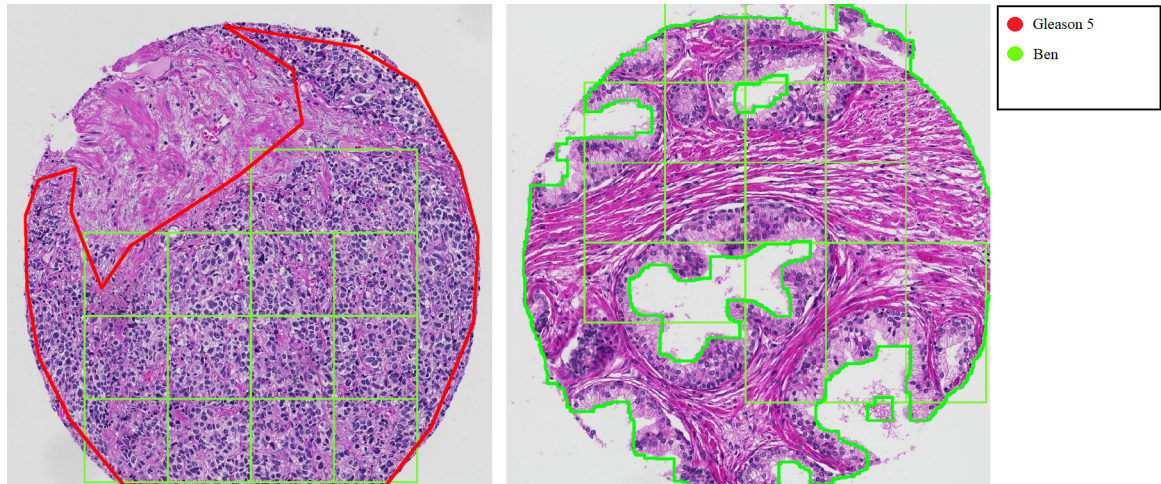


Figura 3.7: Esempi di estrazione smart delle patches. Nell'immagine di destra si ha attivazione del Criterio B* che garantisce l'estrazione delle ghiandole benigne.

L'estrazione smart delle patches è stata utilizzata in fase di validazione per estrarre le patches dalle immagini del Test Set da utilizzare per la generazione delle heat-map. Per tale scopo è stato disattivato il controllo sull'annotazione manuale. In figura 3.8 è mostrata l'estrazione smart delle patches dopo aver disattivato il controllo

sull'annotazione manuale. L'algoritmo estrae correttamente le patches contenenti le ghiandole anche al di fuori dell'annotazione manuale.

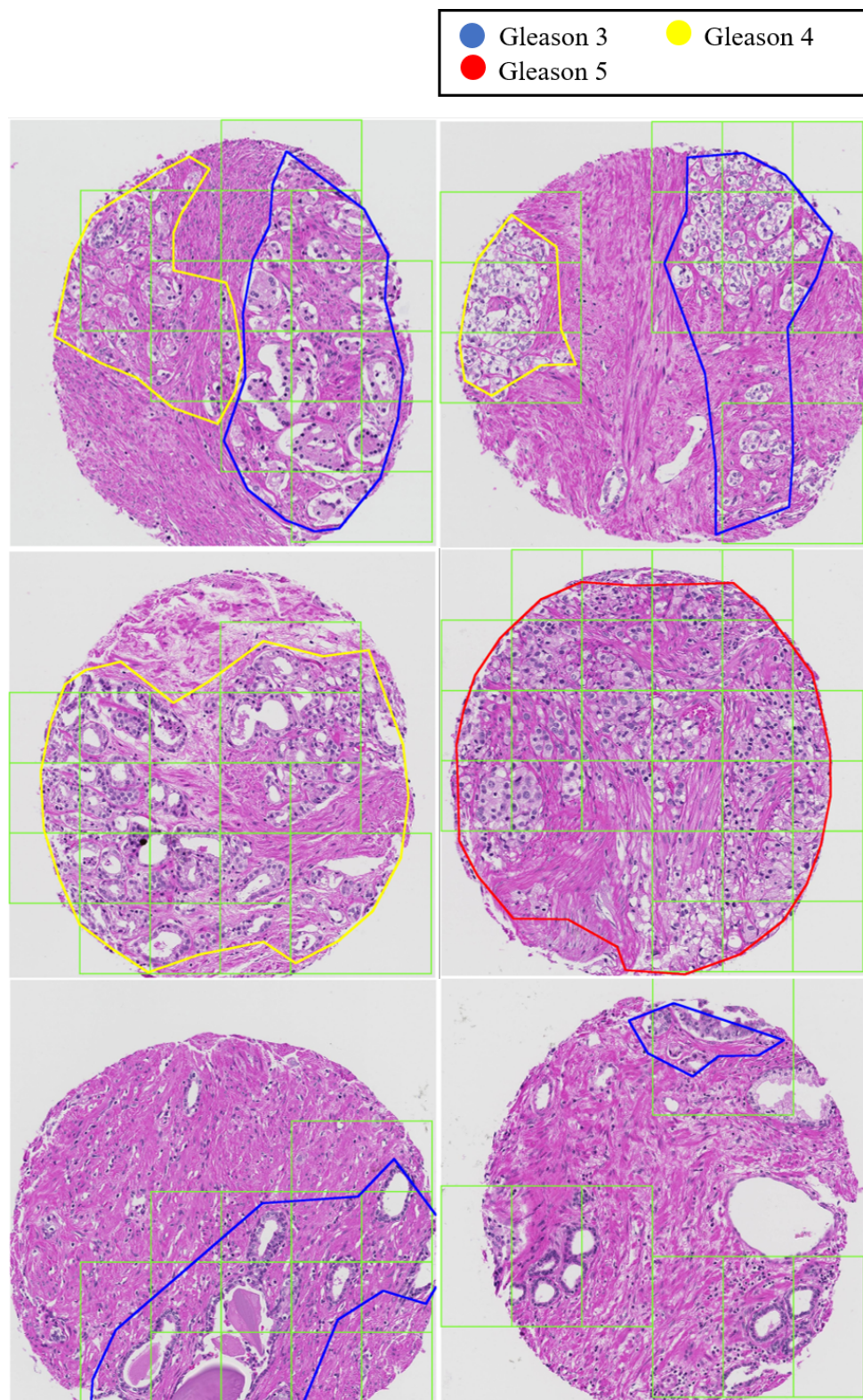


Figura 3.8: Esempi di figure di debug dell'estrazione delle patches tramite l'algoritmo smart senza valutare l'annotazione manuale con overlap del 50%

3.6 Training

3.6.1 Strutture testate

Sono stati testati principalmente tre differenti modelli esistenti di CNN: InceptionV3, InceptionResNetV2 e ResNet50. La differenza principale tra questi differenti modelli sta in degli schemi utilizzati durante la fase di apprendimento, nella differente struttura convoluzionale e nella differente profondità e complessità del modello. Modelli più profondi, possiedono un numero di parametri maggiore, ciò porta a:

- Migliore capacità della rete di risolvere task di classificazione più complessi
- Maggior rischio di overfitting se si utilizzano parametri di apprendimento errati, in quanto la rete possiede un numero sufficiente di parametri per adattarsi ai dati di allenamento

Per limitare il rischio di overfitting risulta fondamentale l'utilizzo di tecniche come il dropout in fase di training.

3.6.2 Transfer Learning

Nel campo del Deep Learning una tecnica ampiamente utilizzata è quella del Transfer Learning. Consiste nell'andare a caricare come pesi dei neuroni della CNN dei pesi ottenuti da altre persone su altri dataset. Il senso dietro questa tecnica è che CNN che sono state allenate anche per task differenti su grandi dataset abbiano già la capacità di estrarre feature dalle immagini nei layer convoluzionali. In questo approccio quindi si utilizzano i layer convoluzionali della rete pre-allenata, ma si sostituiscono i layer di classificazione in quanto andranno allenati per il proprio task specifico. Successivamente si procede all'allenamento dei layer di classificazione ed all'eventuale riallenamento dei layer convoluzionali, il cui adattamento al task specifico impiegherà meno tempo di un eventuale allenamento effettuato da zero. Arvaniti nel suo lavoro utilizza reti pre-allenate sull'ImageNet dataset ed evidenzia l'importanza del transfer learning per incrementare le performance dei sistemi[28].

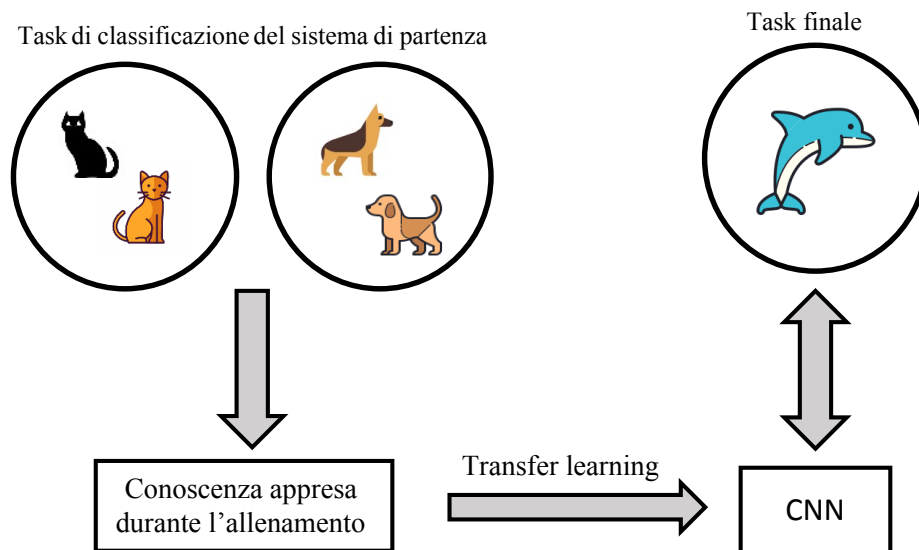


Figura 3.9: Schema di funzionamento del transfer learning

3.6.3 Data Augmentation

Il data augmentation è una tecnica che consiste nell'andare a effettuare trasformazioni morfologiche delle immagini per ottenerne delle nuove differenti. E' utilissima per far apprendere alla rete delle feature immuni alla rotazione, alla traslazione e alle oscillazioni di caratteristiche del colore.

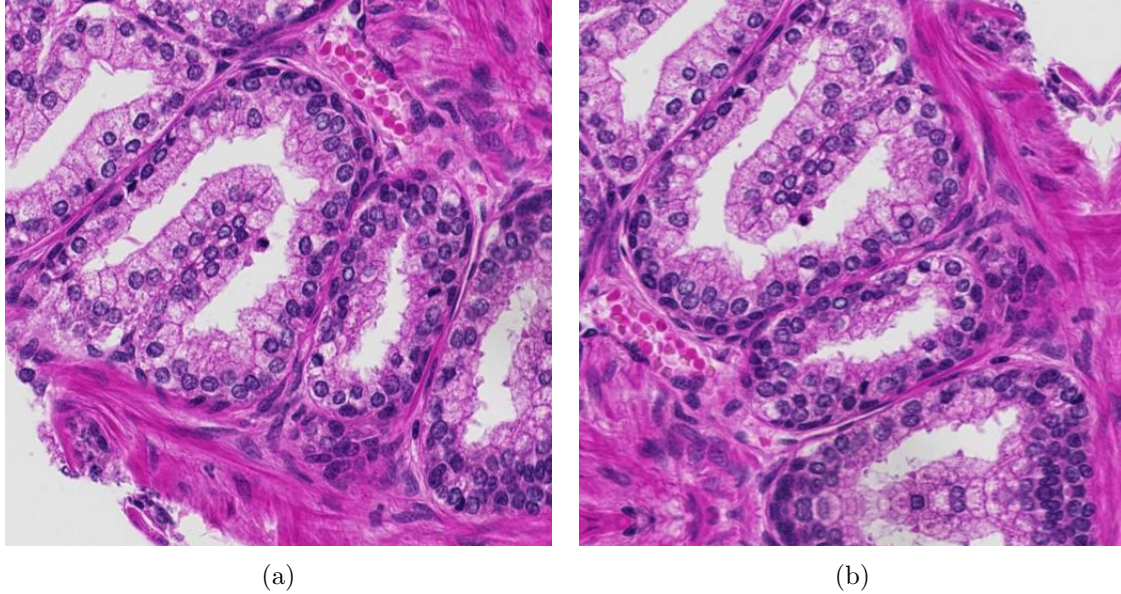


Figura 3.10: Nella *fig.(a)* l'immagine originale mentre nella *fig.(b)* l'immagine aumentata

È pratica comune utilizzare data augmentation real-time durante la fase di training delle CNN[32][33]. In questo lavoro di tesi si è sfruttato sia il data augmentation online che il data augmentation offline.

Classe	Annotazione manuale		Estrazione smart	
	Training	Validation	Training	Validation
Ben	715	79	1353	150
G3	2149	239	1935	214
G4	1495	166	1322	146
G5	780	87	1264	137

Tabella 3.3: *Suddivisione dei due dataset utilizzati per l'allenamento del classificatore a 4 classi con e senza estrazione smart e data augmentation*

Lo sbilanciamento delle diverse classi all'interno del training set viene risolto tramite l'utilizzo di un parametro della rete neurale che consente di considerare con un peso differente durante la fase di training le immagini appartenenti alle diverse classi in funzione del numero di elementi presenti per ogni classe.

3.7 Metriche di valutazione dei risultati e validazione

3.7.1 Confusion matrix

Al fine di confrontare i risultati ottenuti dai tre diversi sistemi di classificazione costruiti si sono valutate le confusion matrix. Si tratta di un metodo di valutazione delle performance di predizione del classificatore rispetto alla classe reale identificata dall'annotazione del patologo. La matrice avrà un numero eguale di righe e colonne che dipende dal numero di classi. Sulle colonne sono contenuti i valori predetti dal classificatore in relazione alla classe reale che è indicata sulle righe. Nel caso di classificazione multiclasse è possibile definire tre gruppi di dati:

- **Correttamente classificati:** gli elementi contenuti sulla diagonale
- **Sovrastimati:** gli elementi al di sopra della diagonale
- **Sottostimati:** gli elementi al di sotto della diagonale

		<i>Predetta</i>			
		Benigna	Gleason 3	Gleason 4	Gleason 5
<i>Reale</i>	Benigna	✓	↑	↑	↑
	Gleason 3	↓	✓	↑	↑
	Gleason 4	↓	↓	✓	↑
	Gleason 5	↓	↓	↓	✓

Figura 3.11: Esempio di una confusion matrix a 4 classi

È possibile definire dei parametri di valutazione globale:

$$ACC = \frac{\sum_{i=j=1}^N M_{ij}}{N_{tot}} \quad (3.7)$$

$$\%Sovrastima = \frac{\sum Elementisovrastimati}{N_{tot}} \quad (3.8)$$

$$\%Sottostima = \frac{\sum Elementisottostimati}{N_{tot}} \quad (3.9)$$

Per effettuare dei confronti analizzando i risultati ottenuti su una specifica classe da due diversi classificatori si utilizzano:

- **Recall:** elementi correttamente predetti in classe i-esima diviso il numero totale di elementi appartenenti alla classe i-esima
- **Precision:** elementi correttamente predetti in classe i-esima diviso il numero totale di elementi predetti in classe i-esima
- **F1-score:** è la media armonica tra recall e precision (*vedi eq. 3.10*)

$$F1 = \frac{2 \times (P \times R)}{P + R} \quad (3.10)$$

3.7.2 Kappa Cohen

Il kappa di Cohen è un coefficiente statistico utilizzato per misurare l'affidabilità inter-operatore. Ha un significato più robusto del coefficiente di concordanza in quanto tiene in conto della probabilità di concordanza casuale. Il suo valore varia nell'intervallo $[-1,1]$, dove valori minori di 0 indicano mancanza completa di concordanza ed 1 indica concordanza massima.

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (3.11)$$

Nell'eq. 3.11 i termini:

- **Pr(a)**: somma degli elementi della diagonale della confusion matrix diviso il numero totale degli elementi. Rappresenta il giudizio concorde.
- **Pr(e)**: prodotto dei totali positivi sommato a quello dei totali negativi (nel caso di classificazione binaria) diviso il quadrato del totale degli elementi. Rappresenta la probabilità di accordo casuale.

Si utilizza il metodo di calcolo pesato lineare del kappa di Cohen in quanto tiene in conto il grado di disaccordo.

3.7.3 Curve ROC

Le curve ROC² sono uno strumento grafico utilizzato per verificare le performance di classificazione al variare del valore di soglia del classificatore. Riportano su ordinate e ascisse rispettivamente TPR³ e FPR⁴. In un problema di classificazione binaria (*fig. 3.12*) è possibile definire:

$$TPR = \frac{TP}{TP + FN} \quad (3.12)$$

$$Spec. = \frac{TN}{TN + FP} \quad (3.13)$$

$$FPR = \frac{FP}{TN + FP} \quad (3.14)$$

		Predetta	
		Classe	
Reale	Benigno	TN	FP
	Maligno	FN	TP

Figura 3.12: Esempio di confusion matrix relativa ad un problema di classificazione binario

Il più importante parametro che è possibile estrarre dalle curve ROC è l'AUC⁵ che rappresenta una misura diretta della capacità di separazione tra le due classi del classificatore.

²Receiver operating characteristic

³True positive rate, sensibilità

⁴False positive rate

⁵Area under the curve

Il suo valore varia nel range $[0,1]$ dove:

- **AUC=0**: peggior modello possibile, classifica positivi come negativi e viceversa
- **AUC=0.5**: il modello non possiede capacità di separazione (separazione casuale)
- **AUC=1**: massimo grado di separazione, tutti gli elementi sono classificati correttamente

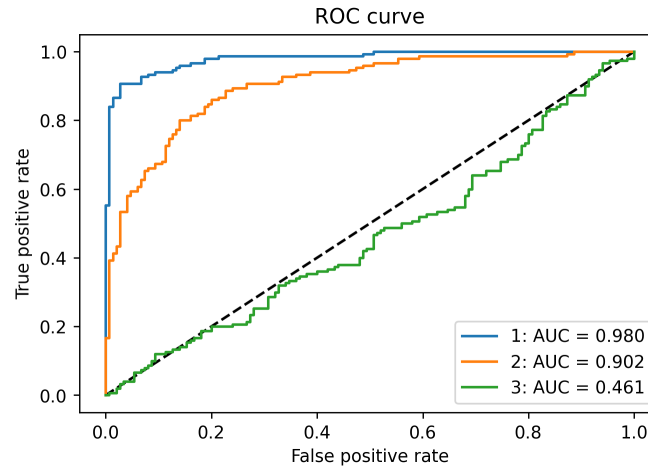


Figura 3.13: Esempio curve ROC e AUC per 3 diversi sistemi.

Nel caso di problemi di classificazione multiclasse la curva ROC diventa una superficie in uno spazio di dimensione N che dipende dal numero di classi:

$$N = 2 \times (n_{classi} - 1) \quad (3.15)$$

L'aumento di dimensione provoca un forte aumento della complessità computazionale di calcolo in caso di problemi multiclasse. In questi casi è possibile utilizzare un approccio a coppie, come suggerito da Landgrebe e Duin [34], o un approccio One vs All ottenendo la binarizzazione di problemi multiclasse [35]. Trattandosi di un sistema di supporto alla diagnosi di carcinoma prostatico nel caso di classificatori multiclasse si è scelto di valutare le curve ROC:

- **Benigno vs G3, G4, G5**: per valutare la capacità di discriminazione tra pattern benigni e tumorali
- **G3 vs G4, G5**: per valutare la capacità di discriminazione tra pattern tumorali a basso ed alto grado di differenziazione che determinano outcome terapeutici differenti (nel caso di pattern G4 si procede alla prostatectomia)

3.8 Heatmap e calcolo del Grade Group

L'obiettivo finale di classificazione è quello di assegnare ad ogni spot appartenente al test set una valutazione globale del Gleason Score e del Grade Group, parametri fondamentali per l'orientamento teranostico dei pazienti. Dopo aver valutato quale dei tre sistemi garantisce la maggiore robustezza sulle patches estratte dal test set è stata utilizzata la pipeline mostrata in *Fig. 3.14* per ottenere la classificazione finale.

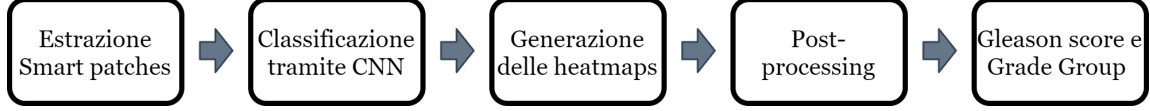


Figura 3.14: Pipeline per la generazione del Gleason Score e Grade Group delle immagini intere

Prima di tutto si effettua l'estrazione smart delle patches di dimensione $512 \times 512 \times 3$ dall'immagine target utilizzando un overlap più elevato rispetto ai parametri di training. Dopo aver ottenuto le patches si ottiene la predizione tramite il modello per ogni singola patch. Successivamente note le coordinate di appartenenza delle patches, vengono assegnati quattro valori ad ogni pixel che corrispondono alla media della probabilità predetta di appartenenza alle quattro diverse classi di ogni patches contenente il pixel-*i*-esimo. Si ottiene così una mappa di probabilità per ognuna delle quattro classi. Il post-processing delle mappe di probabilità consente l'identificazione del pattern primario e dell'eventuale pattern secondario se presente.

Si sfrutta una maschera di segmentazione dei tessuti per limitare l'estensione delle mappe di probabilità alle sole aree contenenti tessuto. Prima di procedere alla classificazione si effettua un global thresholding per eliminare le predizioni con bassa probabilità (<50%). Il passo successivo consiste nel calcolo dell'istogramma delle probabilità per ogni classe. L'istogramma rappresenta sulle ascisse la probabilità e sulle ordinate le occorrenze di pixel associati ad un determinato valore di probabilità. Si ottiene uno score finale per ognuna delle 4 classi sommando il prodotto tra occorrenze di pixel e probabilità associata (vedi *eq. 3.16*).

$$Score_i = \sum_{k=1}^M n_k P_k \quad (3.16)$$

Lo score maggiore sarà identificato come pattern primario. Il secondo score maggiore, se supera un valore di soglia o è appartenente ad una classe maggiore del pattern primario, è valutato come pattern secondario.

Capitolo 4

Risultati

4.1 Classificazione delle patches del Test Set

Si valutano le prestazioni dei classificatori, basati su reti neurali convoluzionali sviluppati secondo i tre diversi schemi architetturali (vedi *Fig. 3.1*), sulle patches estratte dal test set ed annotate dal Patologo 1. Ognuno dei tre diversi sistemi è stato allenato sia sul dataset estratto valutando le sole annotazioni manuali (Normal) che sul dataset ottenuto tramite l'estrazione smart delle patches (Smart) per valutare l'eventuale miglioramento delle prestazioni. Il sistema migliore viene scelto valutando l'accuratezza, la percentuale di sovrastima e principalmente la percentuale di sottostima. Trattandosi di classificazione tumorale è molto importante non classificare in difetto, ma piuttosto nei casi sospetti classificare in eccesso.

4.1.1 Sistema 1

Il sistema 1 è costituito da una sola rete convoluzionale a 4 classi. Analizzando la confusion matrix (*Fig. 4.1*) si osserva che l'accuratezza globale dei due classificatori è molto simile. Il classificatore smart ottiene delle prestazioni di classificazione più bilanciate, con un'accuratezza sensibilmente superiore nel riconoscimento dei pattern benigni e Gleason 5 (rispettivamente +15,2% e +12,7%). La percentuale di patch sovrastimate risulta del 13,5% e 14,1% rispettivamente per il classificatore normal ed il classificatore smart. La percentuale di patch sottostimate del 8,1% e del 7,1%. La maggior parte degli errori commessi dai classificatori è correlata alla corretta distinzione tra Gleason 3 e Gleason 4, che rappresenta una delle maggiori criticità per i patologi. Il classificatore smart sottostima il minor numero di patches Gleason 4 rispetto a tutti gli altri sistemi sviluppati, caratteristica molto interessante in quanto la presenza di Gleason 4 in analisi istologica indirizza verso il trattamento chirurgico radicale. I due sistemi presentano un'elevata percentuale di sovrastima della classe Benigna e Gleason 3. La maggior parte delle patches benigne sovrastimate viene classificata come Gleason 3, cosa che non comporterebbe eventuali trattamenti terapeutici non necessari.

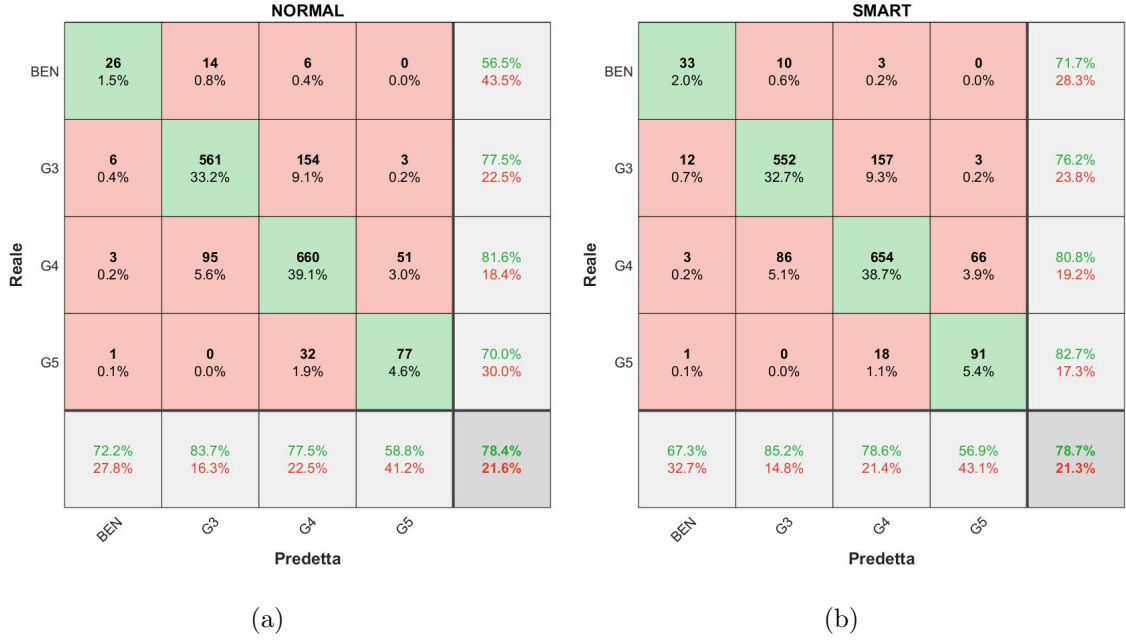


Figura 4.1: Nella figura sono presenti le confusion matrix del classificatore normal *fig.(a)* e smart *fig.(b)* relative al sistema 1

	Normal			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	56,5	0,69	43,5	/
Gleason 3	77,5	0,8	21,8	0,8
Gleason 4	81,6	0,8	6,3	12,11
Gleason 5	70	0,67	/	30
Globale	78,4	0,79	13,5	8,1

Tabella 4.1: Caratteristiche del sistema 1 normal

	Smart			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	71,7	0,69	28,26	/
Gleason 3	76,2	0,8	22,1	1,6
Gleason 4	80,8	0,8	8,16	11
Gleason 5	82,7	0,67	/	17,27
Globale	78,4	0,79	14,15	7,1

Tabella 4.2: Caratteristiche del sistema 1 smart

In *Fig.4.2* sono mostrate le curve ROC ottenute tramite il metodo one vs all. I valori elevati di AUC ottenuti da entrambi i sistemi sottolineano la buona separazione tra le classi maligne e la classe benigna. Solo 16 patches su 1643 maligne sono state predette come benigne.

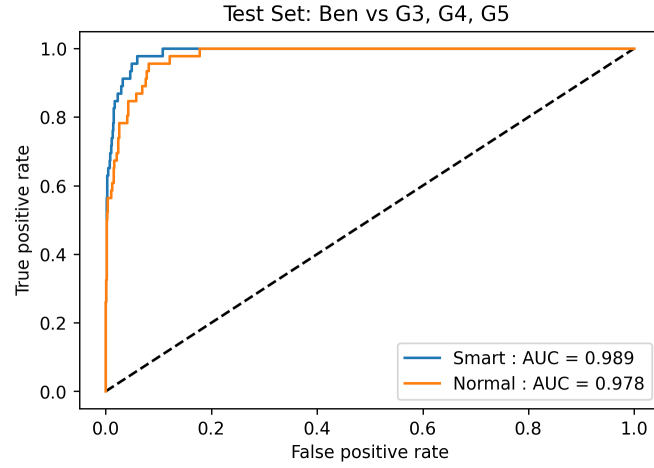


Figura 4.2: Curve ROC dei due classificatori del sistema 1

4.1.2 Sistema 2

Il sistema due è costituito dalla successione di due reti neurali convoluzionali in cascata:

- **2 CLASSI:** benigno vs maligno
- **3 CLASSI:** separare le patches identificate come maligne nelle classi Gleason

Il classificatore smart ottiene delle prestazioni superiori di accuratezza globale (75,7% contro 70,1%) e di accuratezza delle singole classi. Sottostima un numero significativamente inferiore di patches (10% contro 15,7%), ottenendo però una percentuale di sottostima superiore al sistema 1 smart (7,1%).

		NORMAL							SMART				
Reale	BEN	28 1.7%	8 0.5%	10 0.6%	0 0.0%	60.9% 39.1%		BEN	33 2.0%	10 0.6%	3 0.2%	0 0.0%	71.7% 28.3%
	G3	43 2.5%	494 29.2%	185 11.0%	2 0.1%	68.2% 31.8%		G3	12 0.7%	523 31.0%	189 11.2%	0 0.0%	72.2% 27.8%
	G4	20 1.2%	145 8.6%	610 36.1%	34 2.0%	75.4% 24.6%		G4	3 0.2%	103 6.1%	663 39.3%	40 2.4%	82.0% 18.0%
	G5	11 0.7%	0 0.0%	47 2.8%	52 3.1%	47.3% 52.7%		G5	1 0.1%	1 0.1%	49 2.9%	59 3.5%	53.6% 46.4%
		27.5% 72.5%	76.4% 23.6%	71.6% 28.4%	59.1% 40.9%	70.1% 29.9%			67.3% 32.7%	82.1% 17.9%	73.3% 26.7%	59.6% 40.4%	75.7% 24.3%
		Predetta							Predetta				
		BEN	G3	G4	G5				BEN	G3	G4	G5	

Figura 4.3: Nella figura sono presenti le confusion matrix del classificatore normal *fig.(a)* e smart *fig.(b)* relative al sistema 2

Come si può osservare confrontando le tabelle sottostanti con quelle relative al sistema 1 (Tab. 4.1/4.2) l'utilizzo di due reti neurali in cascata non apporta alcun miglioramento alla classificazione.

	Normal			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	60,9	0,38	39,1	/
Gleason 3	68,2	0,72	25,8	5,9
Gleason 4	75,4	0,73	4,2	20,4
Gleason 5	47,3	0,53	/	52,7
Globale	70,1	0,71	14,15	15,7

Tabella 4.3: Caratteristiche del sistema 2 normal

	Smart			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	71,7	0,69	28,26	/
Gleason 3	72,2	0,77	26,1	1,7
Gleason 4	82	0,77	4,9	12,7
Gleason 5	53,6	0,56	/	46,4
Globale	75,7	0,76	14,32	10

Tabella 4.4: Caratteristiche del sistema 2 smart

In Fig.4.4 si può osservare la curva ROC ottenuta con il metodo one vs all delle reti a 3 classi relative al sistema 2. Nonostante il sistema 2 non ottenga delle buone prestazioni globali, comparate al sistema 1, la seconda rete neurale presenta un'ottima capacità di discriminazione tra pattern tumorali di basso grado (Gleason 3) e pattern tumorali di alto grado (Gleason 4 e Gleason 5). L'allenamento tramite l'estrazione smart delle patches porta ad un miglioramento di classificazione come si può notare dalle curve ROC e dai valori di AUC.

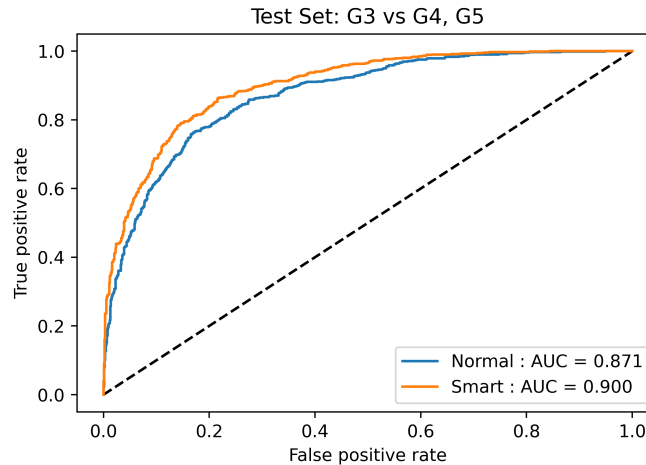


Figura 4.4: Curve ROC delle reti a 3 classi del sistema 2

4.1.3 Sistema 3

Il sistema due è costituito dalla successione di tre reti neurali convoluzionali in cascata:

- **2 CLASSI:** benigno vs maligno
- **2 CLASSI:** classifica le patches in Gleason 3 + Gleason 4 vs Gleason 5
- **2 CLASSI:** classifica le patches in Gleason 3 vs Gleason 4

Analizzando le confusion matrix del sistema 3 (*Fig.4.5*) si nota che i due classificatori hanno un accuratezza globale comparabile, con il sistema smart che ottiene prestazioni più bilanciate. L'utilizzo di tre reti in cascata dove l'ultima effettua la classificazione Gleason 3 vs Gleason 4 non porta a dei miglioramenti a causa dell'errore cumulato di classificazione delle tre reti, nonostante le prestazioni della rete Gleason 3 vs Gleason 4 siano buone (*vedi Fig.4.6*). Le percentuali di sotto-stima e sovrastima risultano essere peggiori del sistema 1 smart (+1,2% e +7,1% rispettivamente).

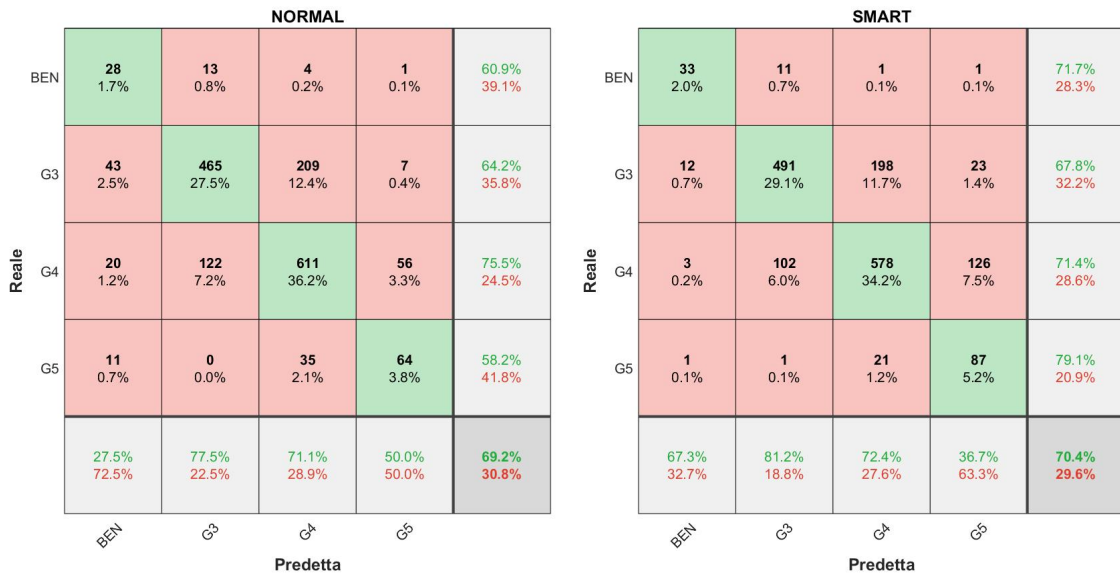


Figura 4.5: Nella figura sono presenti le confusion matrix del classificatore normal *fig.(a)* e smart *fig.(b)* relative al sistema 3

	Normal			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	60,9	0,38	39,1	/
Gleason 3	64,2	0,70	29,8	5,9
Gleason 4	75,5	0,73	6,9	17,8
Gleason 5	58,2	0,54	/	41,8
Globale	69,2	0,70	17,16	13,7

Tabella 4.5: Caratteristiche del sistema 3 normal

	Smart			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	71,7	0,69	28,3	/
Gleason 3	67,8	0,74	30,5	1,7
Gleason 4	71,4	0,72	15,6	13
Gleason 5	79,1	0,50	/	20,9
Globale	70,4	0,71	21,3	8,3

Tabella 4.6: Caratteristiche del sistema 3 smart

Analizzando la curva ROC relativa al classificatore Gleason 3 vs Gleason 4 si evidenzia la difficoltà di distinzione tra le due classi. L'utilizzo dell'estrazione smart delle patches migliora la capacità di distinzione tra gleason 3 e gleason 4.

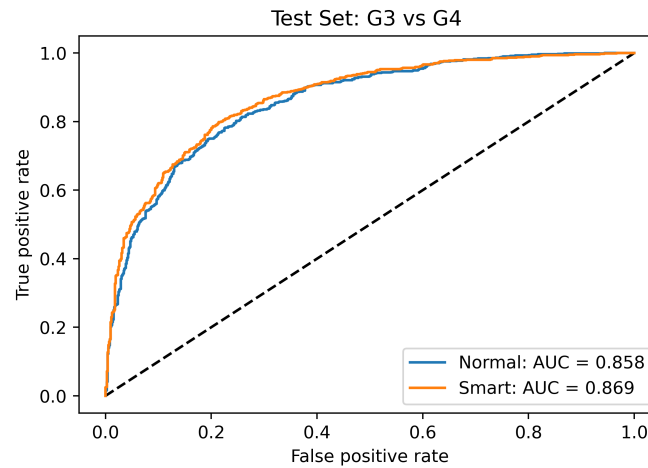


Figura 4.6: Curve ROC delle reti Gleason 3 vs Gleason 4 del sistema 3

4.2 Validazione del miglior sistema

Il sistema che fornisce le migliori prestazioni è il sistema 1 che utilizza una sola rete neurale convoluzionale a 4 classi per discriminare i diversi pattern. Il sistema 1 fornisce dei buoni risultati sul test set che uniti al guadagno computazionale di utilizzo di una sola rete neurale lo rendono la migliore scelta. Per valutare la robustezza del classificatore si è scelto di analizzare le patches predette con maggiore probabilità (*Fig.4.7*). Un totale di 860 patches su 1689, 51% del test set, viene predetta con una probabilità maggiore dell'80% con un'accuratezza di classificazione media del 90,9%. Di queste 509, 30% del test set, sono state predette con una probabilità maggiore del 90% ed un'accuratezza di classificazione media del 95,3%. Questi risultati dimostrano la capacità di generalizzazione e la solidità del modello convoluzionale che presenta alte percentuali di accuratezza sulle patch predette con maggiore confidenza.

Test Set SMART

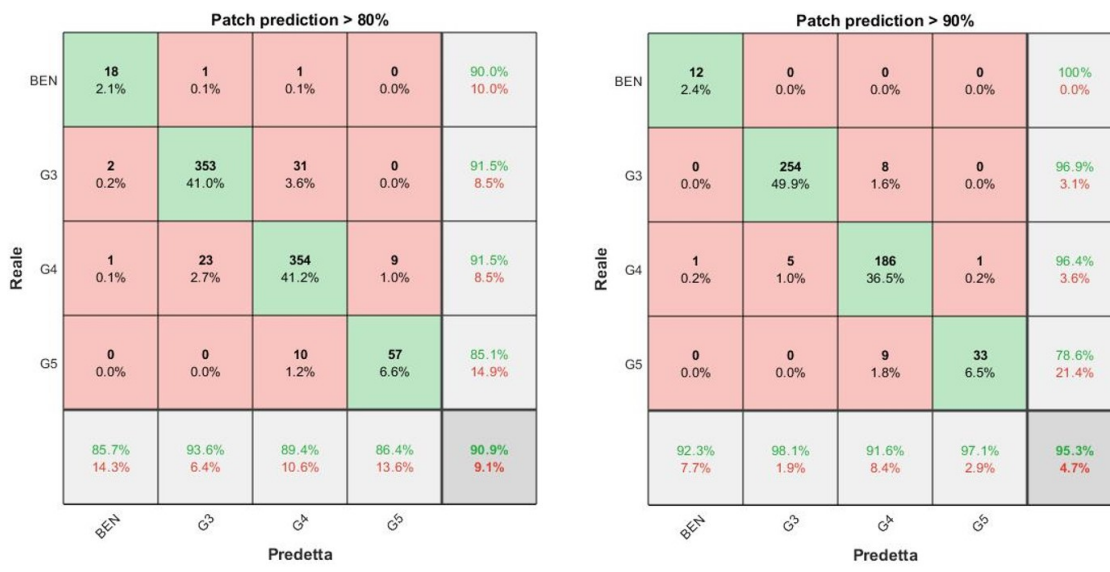


Figura 4.7: In figura le confusion matrix delle patch predette con maggiore probabilità con due soglie differenti

Le percentuali di sovrastima e sottostima delle patches predette con alta probabilità sono sensibilmente inferiori alle percentuali globali. Nel caso delle patches con probabilità maggiore dell'80% si ha una sovrastima del 4,9% ed una sottostima del 4,2% (-9,2% e -2,9% rispetto al totale). Le patches con probabilità maggiore dell'90% hanno una sovrastima del 1,8% ed una sottostima del 2,9% (-12,3% e -4,2% rispetto al totale). Nel caso di Gleason 4 il sistema sottostima solo 6 patches su 193 (3,1%). Gli errori di sottostima di Gleason 5 in Gleason 4 risultano meno significativi in quanto non comportano conseguenze terapeutiche.

	Patch con probabilità maggiore 80%			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	90	0,88	10	/
Gleason 3	91,5	0,93	8	0,5
Gleason 4	91,5	0,90	2,3	6,2
Gleason 5	85,1	0,86	/	14,92
Globale	90,9	0,91	4,90	4,2

Tabella 4.7: Parametri di valutazione delle patches predette con probabilità di classificazione maggiore dell'80%

	Patch con probabilità maggiore 90%			
Classe	Accuratezza [%]	F1-score	%Sovrastimati	%Sottostimati
Benigno	100	0,96	0	/
Gleason 3	96,9	0,98	2,2	0
Gleason 4	96,4	0,94	0,5	3,1
Gleason 5	78,6	0,87	/	21,4
Globale	95,3	0,95	1,8	2,9

Tabella 4.8: Parametri di valutazione delle patches predette con probabilità di classificazione maggiore dell'90%

4.2.1 Classificazione del Grade Group

Si sfrutta il modello convoluzionale costruito capace di effettuare classificazione delle singole patches per andare ad ottenere la classificazione dei pattern presenti all'interno delle immagini intere del test set. Si estraggono tramite l'algoritmo smart le patches e si sfruttano le predizioni della rete sulle diverse patches per generare una mappa di probabilità. Tali mappe consentono di identificare i pattern presenti e di ottenere il Gleason Score ed il Grade Group. Il test set è stato annotato da un secondo patologo consentendo il confronto inter-operatore.

Modello vs Patologo 1

Il sistema sviluppato presenta un'accuratezza di predizione del Grade Group pari all'60,5% rispetto al patologo 1. Gli errori di classificazione si concentrano rispetto alle classi reali GG 2 e GG 3, dove il sistema non riconosce correttamente presenza ed estensione del pattern secondario (Gleason 3 o Gleason 4).

		Grade Group				
Modello	GG1	62	8	1	4	0
	GG2	13	4	2	2	0
	GG3	11	7	9	14	0
	GG4	4	6	8	56	3
	GG5	0	0	0	11	13
		GG1	GG2	GG3	GG4	GG5
		Patologo 1				

Figura 4.8: Confusion Matrix relativa al Grade Group del modello vs Patologo 1

Per analizzare le criticità del sistema si sono valutate separatamente le prestazioni di riconoscimento del pattern principale e secondario (*Fig. 4.9*). Il classificatore

possiede un accuratezza di predizione del 77,8% rispetto al pattern primario che scende al 64,7% per il pattern secondario. Tale peggioramento può essere causato dai pattern secondari di dimensione minore della dimensione delle patches e di conseguenza difficili da identificare correttamente e da errori nella metrica di valutazione della presenza del pattern secondario. Un miglioramento potrebbe essere ottenuto sviluppando col supporto di un anotomo patologo delle metriche stabili per stabilire la presenza di pattern secondario.

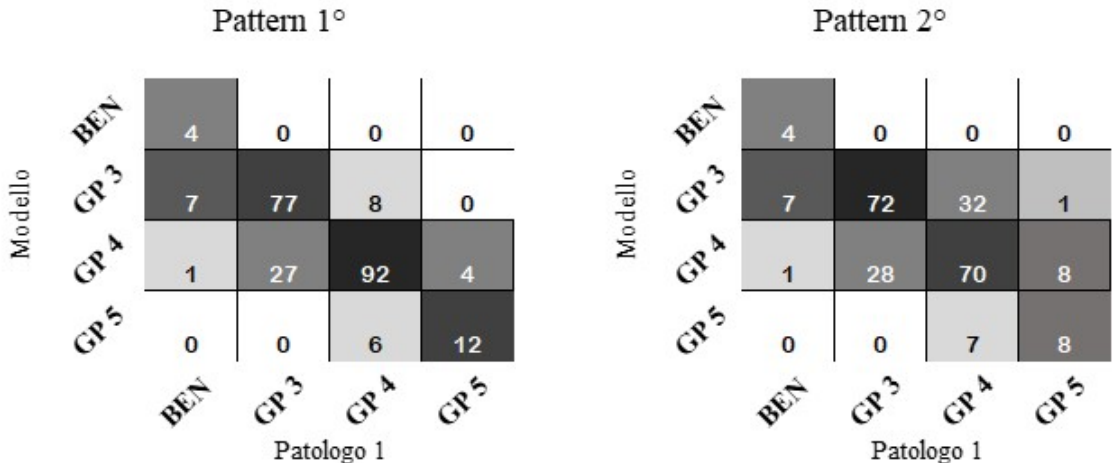


Figura 4.9: Confusion matrix relative a pattern primario e secondario del modello vs Patologo 1

Modello vs Patologo 2

L'inter-agreement rispetto al patologo 2 si attesta al 45,4%, sensibilmente inferiore all'inter-agreement rispetto al patologo 1 (60,5%), ma molto simile all'inter-agreement dei due patologi pari al 45,8%.

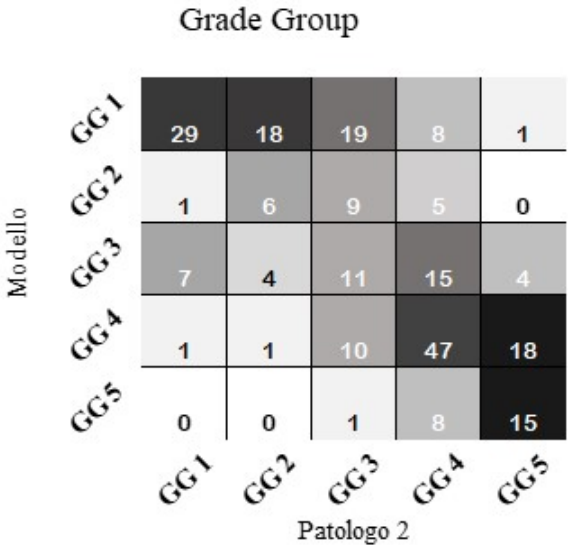


Figura 4.10: Confusion Matrix relativa al Grade Group del modello vs Patologo 2

L'accuratezza di identificazione del pattern primario è pari al 66,8% e scende al 56,8% per il pattern secondario. Come atteso il sistema presenta una concordanza maggiore con il patologo 1 che ha realizzato le annotazioni delle immagini di test.

		Pattern 1°						Pattern 2°			
Modello	BEN	3	1	0	0	Modello	BEN	3	1	0	0
	GP 3	6	44	41	1		GP 3	6	56	47	3
	GP 4	1	12	100	11		GP 4	1	20	68	18
	GP 5	0	0	6	12		GP 5	0	1	6	8
		BEN	GP 3	GP 4	GP 5			BEN	GP 3	GP 4	GP 5
		Patologo 2						Patologo 2			

Figura 4.11: Confusion matrix relative a pattern primario e secondario del modello vs Patologo 2

Patologo 1 vs Patologo 2

L'inter-agreement tra i due patologi si attesta al 45,8%, dimostrando l'estrema variabilità e soggettività di valutazione tramite scala di Gleason. Il patologo 2 classifica 39,5% delle patches in una classe superiore a quella del patologo 1, caratteristica che giustifica l'aumento degli elementi sottostimati da parte del sistema rispetto al patologo 2.

		Grade Group				
Patologo 1	GG 1	34	21	23	12	0
	GG 2	2	2	9	12	0
	GG 3	0	2	8	4	6
	GG 4	2	4	10	52	19
	GG 5	0	0	0	3	13
		GG 1	GG 2	GG 3	GG 4	GG 5
		Patologo 2				

Figura 4.12: Confusion Matrix relativa al Grade Group del Patologo 1 vs Patologo 2

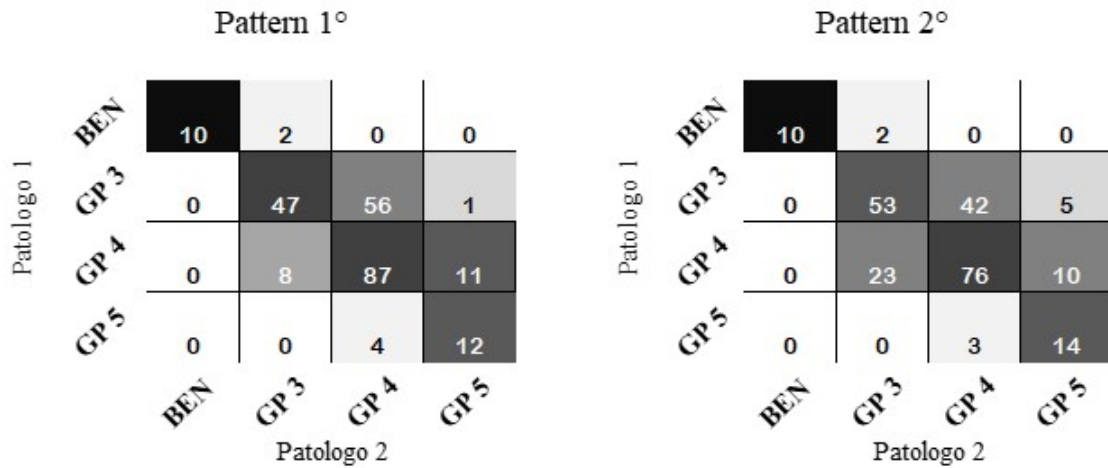


Figura 4.13: Confusion matrix relative a pattern primario e secondario del Patologo 1 vs Patologo 2

Kappa Cohen			
Valutazione	Modello vs P1	Modello vs P2	P1 vs P2
Grade Group	0,66	0,49	0,48
Pattern 1°	0,67	0,50	0,53
Pattern 2°	0,48	0,37	0,50

Tabella 4.9: Valutazione del kappa di Cohen per le diverse coppie

Analizzando i risultati presenti in *Tab.4.9* si evince che il sistema sviluppato, come atteso, presenta un miglior tasso di concordanza con il patologo 1, superiore al tasso di concordanza inter-patologo. Il tasso di concordanza con il patologo 2 è comparabile a quello inter-patologo. Valutando i valori di kappa relativi al pattern secondario si nota la difficoltà del sistema nel riconoscere e classificare correttamente il pattern secondario rispetto ad entrambi i patologi.

4.2.2 Valutazione delle Heatmap

Sono state selezionate le heatmaps di 6 immagini per dimostrare le capacità di classificazione e generalizzazione del sistema sviluppato ed analizzare criticamente gli errori di classificazione commessi dalla rete. Per ognuna delle heatmap sono mostrate le annotazioni dei patologi con indicati i bordi dei pattern individuati (Verde: Benigno, Blu: Gleason 3, Giallo: Gleason 4 e Rosso: Gleason 5). In *Fig.4.14* è mostrato un errore di classificazione della rete rispetto al patologo 1. Il sistema riconosce in concordanza con il patologo 1 la maggior parte dell'immagine come benigna ma una porzione viene classificata come Gleason 3 in accordo al patologo 2.

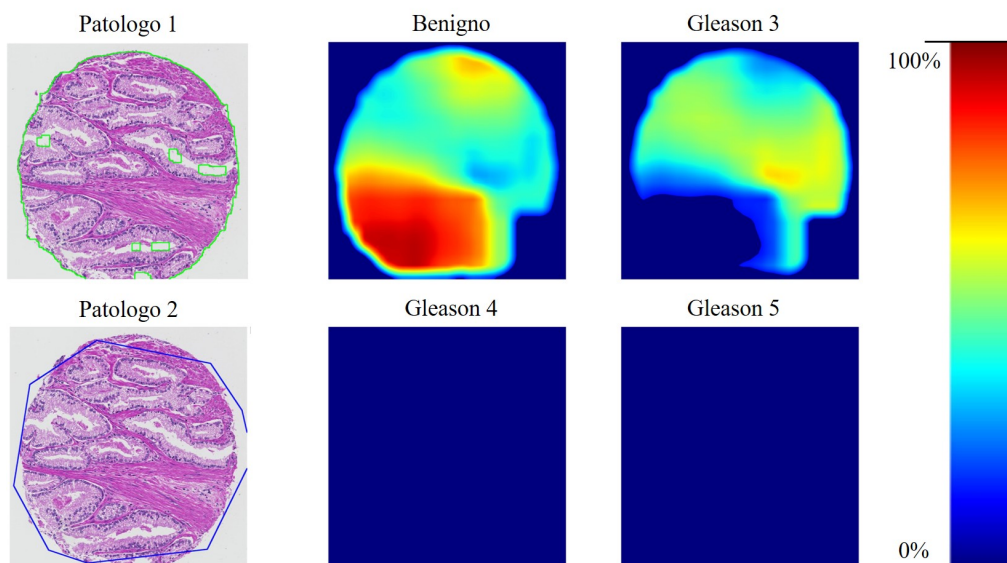


Figura 4.14: Heatmap immagine 33. Patologo 1: Benigno (verde), Patologo 2: Gleason 3(Blu).

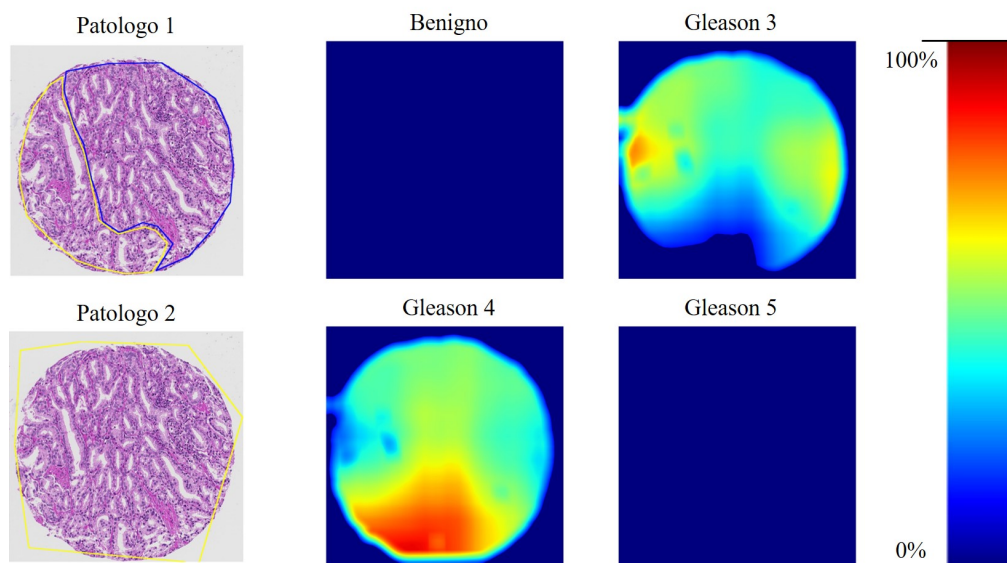


Figura 4.15: Heatmap immagine 51. Patologo 1: Gleason 3 (blu) + Gleason 4 (giallo), Patologo 2: Gleason 4 (giallo)

Nelle Fig.4.15/ 4.16 sono mostrati due casi nei quali i due patologi valutano in maniera differente la presenza di Gleason pattern 3 e 4. Nella prima immagine il sistema riconosce la presenza di Gleason pattern 3 e 4 in accordo al patologo 1 ed in disaccordo con il patologo 2 il quale riconosce solamente il Gleason pattern 4. Nella

seconda immagine il sistema in accordo al patologo 1 riconosce il Gleason pattern 3 come principale ma in accordo al patologo 2 riconosce la presenza di Gleason pattern 4, anche se si trova in disaccordo sull'estensione.

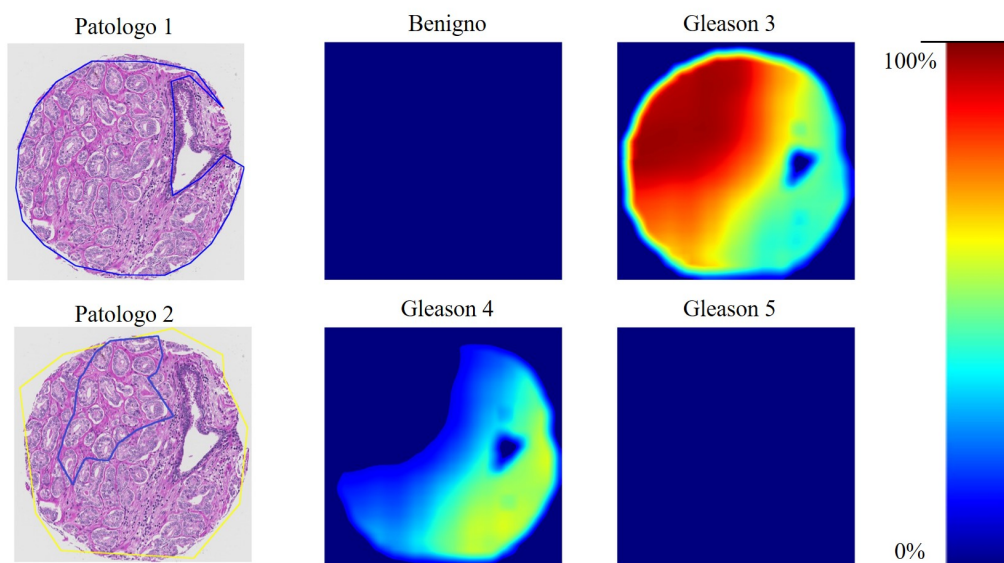


Figura 4.16: Heatmap immagine 64. Patologo 1: Gleason 3 (blu), Patologo 2: Gleason 4 (giallo) + Gleason 3 (blu).

Nelle *Fig.4.17/ 4.18/ 4.19* sono mostrati tre casi nei quali vi è il pieno accordo dei patologi sul pattern presente e sulla sua estensione uno per ognuna delle classi tumorali. Il sistema classifica la quasi totalità delle zone corrispondenti ad annotazione manuale con una probabilità prossima al 100% dimostrando le buone capacità di riconoscimento di pattern singoli e con elevato numero di ghiandole.

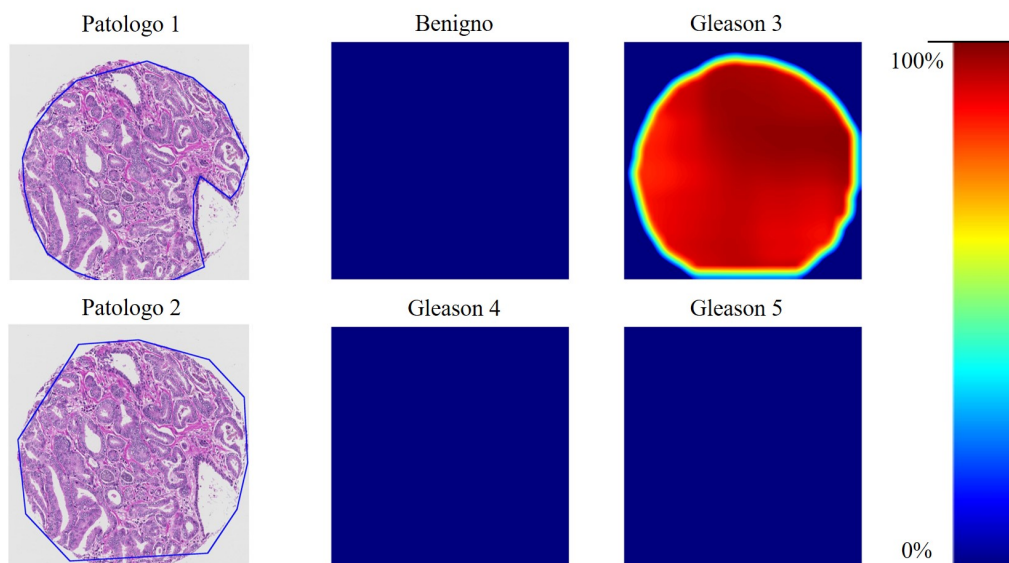


Figura 4.17: Heatmap immagine 119. Patologo 1: Gleason 3 (blu), Patologo 2: Gleason 3 (blu).

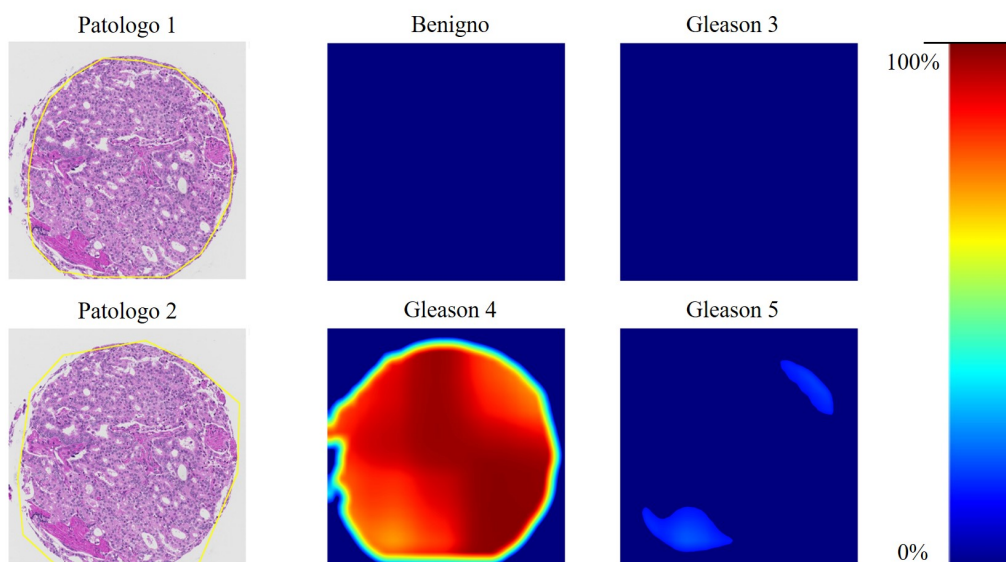


Figura 4.18: Heatmap immagine 120. Patologo 1: Gleason 4 (giallo), Patologo 2: Gleason 4 (giallo).

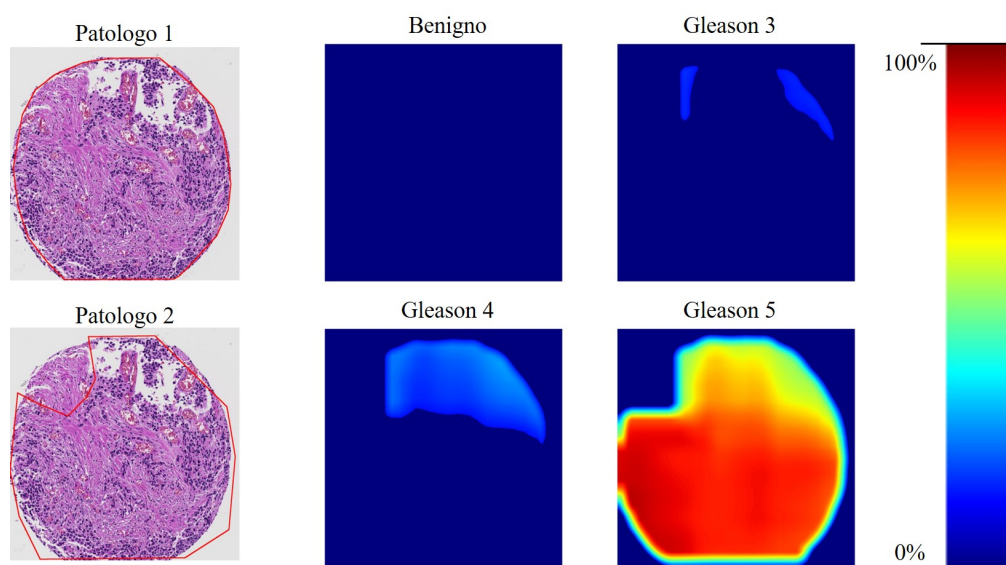


Figura 4.19: Heatmap immagine 79. Patologo 1: Gleason 5 (rosso), Patologo 2: Gleason 5 (rosso).

Capitolo 5

Conclusioni

5.1 Conclusioni

Nel mio lavoro di tesi ho elaborato un sistema di analisi automatica di istologie prostatiche digitali, tramite il quale è possibile ottenere una valutazione del Gleason Score, parametro fondamentale per la valutazione della prognosi di pazienti con sospetta patologia cancerosa della prostata.

Nel campo della Digital Pathology, l'utilizzo di reti neurali convoluzionali, rappresenta ad oggi lo standard per lo sviluppo di sistemi automatici a supporto dell'analisi istologica. La capacità dei sistemi convoluzionali di riconoscere correttamente le caratteristiche delle diverse lesioni è fortemente condizionata dalla quantità e dalla qualità dei dati utilizzati durante la fase di allenamento. A tal scopo si è valutato l'utilizzo di tecniche di pre-processing ed estrazione smart delle patches per migliorare la robustezza dei dati di allenamento e garantire che il sistema apprenda a pieno le caratteristiche utili alla discriminazione dei diversi pattern. Il sistema sviluppato mostra delle buone prestazioni dal punto di vista della classificazione delle patches con dei valori di accuratezza sul test set pari al 78,4%.

L'algoritmo sviluppato consente l'identificazione del Gleason Score di immagini istologiche tramite la valutazione delle mappe di probabilità ottenute attraverso la classificazione di patches estratte dall'immagine stessa. I risultati ottenuti dal sistema nella valutazione del Grade Group sono stati confrontati con le conclusioni di due patologi esperti mostrando un tasso di concordanza, k_{esperto1} uguale a 0.69 e k_{esperto2} uguale a 0.49. I due valori di kappa di Cohen ottenuti sono comparabili al valore inter-patologo di 0.48, ciò dimostra la corretta capacità di generalizzazione del problema ottenuta dalla rete convoluzionale.

Nonostante le dimensioni ridotte del dataset i risultati ottenuti confermano le grandi potenzialità nell'utilizzo di sistemi basati su machine learning come strumenti di supporto per la Digital Pathology.

5.2 Sviluppi futuri

Nonostante le buone prestazioni raggiunte dal sistema permane il limite legato alla soggettività nella valutazione del Gleason Score da parte del patologo che ha annotato il training set. Per superare i limiti rappresentati dalla soggettività è auspicabile costruire il dataset originale basandosi su annotazioni che siano condivise da più patologi. Per migliorare ulteriormente le prestazioni della rete neurale convoluzionale è consigliabile: aumentare la quantità di dati di allenamento e utilizzare soglie più selettive per l'algoritmo di estrazione delle patches.

Si suggerisce inoltre la partecipazione di un patologo esperto ad ogni fase di sviluppo del sistema che possa fornire dei suggerimenti utili a migliorare le performance e a rendere il sistema più funzionale per gli stessi patologi.

In base alle prestazioni registrate si può ipotizzare l'utilizzo di questo sistema come punto di partenza per l'implementazione futura della tecnica anche nel campo dell'analisi di immagini WSI riducendo drasticamente il tempo di analisi istologiche da parte del patologo.

Bibliografia

- [1] Eirini Arvaniti et al. *Replication Data for: Automated Gleason grading of prostate cancer tissue microarrays via deep learning*. Ver. V1. 2018. DOI: 10.7910/DVN/OCYCMP. URL: <https://doi.org/10.7910/DVN/OCYCMP>.
- [2] Christine H Lee, Oluyemi Akin-Olugbade e Alexander Kirschenbaum. “Overview of prostate anatomy, histology, and pathology”. In: *Endocrinology and Metabolism Clinics* 40.3 (2011), pp. 565–575.
- [3] URL: www.studiourologicoallo.it/prostata.
- [4] URL: <https://teachmesurgery.com/urology/prostate/prostatitis/>.
- [5] John E McNeal. “Normal histology of the prostate”. In: *The American journal of surgical pathology* 12.8 (1988), pp. 619–633.
- [6] DF Gleason. “Histologic grading and staging of prostatic carcinoma”. In: *The American Journal of Surgical Pathology* 5.2 (1981), p. 193.
- [7] C.L. Stanfield. *Fisiologia*. Edises, 2012. ISBN: 9788879597142. URL: <https://books.google.it/books?id=2i0juQAACAAJ>.
- [8] R. A. S. Hemat. *Principles of Orthomolecularism*. Urotext, 2004. ISBN: 9781903737057. URL: https://books.google.it/books?id=ED%5C_xI-CEzFYC.
- [9] URL: <http://www.phytolab.it/iperplasia-prostatica-benigna-ipb-possibili-cause-e-principali-sintomi/>.
- [10] URL: <http://www.mpuh.org/centreforroboticsurgery/treatments-related-to-various-stages-of-prostate-cancer/>.
- [11] James E Montie. “Observations on the epidemiology and natural history of prostate cancer”. In: *Urology* 44.6 (1994), pp. 2–8.
- [12] Claire H Pernar et al. “The epidemiology of prostate cancer”. In: *Cold Spring Harbor perspectives in medicine* 8.12 (2018), a030361.
- [13] P Kheirandish e F Chinegwundoh. “Ethnic differences in prostate cancer”. In: *British journal of cancer* 105.4 (2011), pp. 481–485.
- [14] Fritz H Schröder et al. “Screening and prostate-cancer mortality in a randomized European study”. In: *New England journal of medicine* 360.13 (2009), pp. 1320–1328.
- [15] Fritz H Schröder et al. “Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up”. In: *The Lancet* 384.9959 (2014), pp. 2027–2035.
- [16] AIRO AIOM, SIURO SIU e SIRM. *Linee guida Carcinoma della Prostata*. 2019.

- [17] URL: <http://www.cardiec.com/old/immagine/7neo.htm>.
- [18] Dirk Beyersdorff et al. "MRI of prostate cancer at 1.5 and 3.0 T: comparison of image quality in tumor detection and staging". In: *American Journal of Roentgenology* 185.5 (2005), pp. 1214–1220.
- [19] Gladell P Paner et al. "Essential updates in grading, morphotyping, reporting, and staging of prostate carcinoma for general surgical pathologists". In: *Archives of pathology & laboratory medicine* 143.5 (2019), pp. 550–564.
- [20] Anna Bill-Axelson et al. "Radical prostatectomy or watchful waiting in early prostate cancer". In: *New England Journal of Medicine* 370.10 (2014), pp. 932–942.
- [21] Freddie C Hamdy et al. "10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer". In: *N Engl J Med* 375 (2016), pp. 1415–1424.
- [22] Stacy Loeb et al. "Overdiagnosis and overtreatment of prostate cancer". In: *European urology* 65.6 (2014), pp. 1046–1055.
- [23] URL: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- [24] Jayanth Koushik. "Understanding convolutional neural networks". In: *arXiv preprint arXiv:1605.09081* (2016).
- [25] Kunal Nagpal et al. "Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer". In: *NPJ digital medicine* 2.1 (2019), pp. 1–10.
- [26] Wouter Bulten et al. "Automated gleason grading of prostate biopsies using deep learning". In: *arXiv preprint arXiv:1907.07980* (2019).
- [27] Olaf Ronneberger, Philipp Fischer e Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [28] Eirini Arvaniti et al. "Automated Gleason grading of prostate cancer tissue microarrays via deep learning". In: *Scientific reports* 8.1 (2018), pp. 1–11.
- [29] Gabriele Campanella, Vitor Werneck Krauss Silva e Thomas J. Fuchs. "Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology". In: *CoRR* abs/1805.06983 (2018). arXiv: 1805.06983. URL: <http://arxiv.org/abs/1805.06983>.
- [30] Gabriele Campanella et al. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images". In: *Nature Medicine* 25 (ago. 2019), p. 1. DOI: 10.1038/s41591-019-0508-1.
- [31] Marc Macenko et al. "A method for normalizing histology slides for quantitative analysis". In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE. 2009, pp. 1107–1110.
- [32] Geert Litjens et al. "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis". In: *Scientific reports* 6 (2016), p. 26286.

- [33] Quy Dinh Duong et al. “Scale embedding shared neural networks for multiscale histological analysis of prostate cancer”. In: *Medical Imaging 2019: Digital Pathology*. Vol. 10956. International Society for Optics e Photonics. 2019, p. 1095606.
- [34] Thomas CW Landgrebe e Robert PW Duin. “Approximating the multiclass ROC by pairwise analysis”. In: *Pattern recognition letters* 28.13 (2007), pp. 1747–1758.
- [35] Mikel Galar et al. “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes”. In: *Pattern Recognition* 44.8 (2011), pp. 1761–1776.