

POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Biomedica

Tesi di Laurea Magistrale

Sviluppo di un algoritmo automatico basato su deep learning per la classificazione del carcinoma prostatico in immagini istologiche



Relatori

Prof. Filippo Molinari
Ing. Massimo Salvi

Candidata

Sara Vanessa Morciano

Dicembre 2021

Sommario

Il carcinoma prostatico è una delle neoplasie che affligge maggiormente la popolazione maschile in tutto il mondo, con 1.4 milioni di casi diagnosticati e 375 mila morti nel 2020. Questo tipo di tumore nella maggior parte dei casi si presenta in maniera asintomatica e cresce molto lentamente, mentre in altri casi è molto più aggressivo. Dunque è fondamentale stimare in maniera corretta il grado del tumore per poter attuare una strategia terapeutica adatta, che può essere a scopo radicale (prostatectomia radicale o radioterapia) oppure di sorveglianza attiva, a seconda dell'aggressività.

Attualmente, la diagnosi e la valutazione del grado del tumore viene effettuata da un patologo tramite l'analisi istologica dei campioni di tessuto ottenuti tramite agobiopsia, a cui viene attribuito un punteggio di Gleason (Gleason score) a seconda del livello di differenziazione delle ghiandole del tessuto prostatico.

La valutazione del Gleason score è molto dispendiosa in termini di tempo e porta ad una grande variabilità intra e inter-operatore, dovuta anche alla natura qualitativa di questa stima, che porta inevitabilmente ad una prognosi variabile e un trattamento non ottimale della patologia.

L'obiettivo di questo lavoro di tesi è lo sviluppo di un sistema automatico che identifichi il tessuto canceroso in immagini istologiche di prostata. Il sistema si basa sull'utilizzo di reti neurali convoluzionali (CNN) per la classificazione delle immagini istologiche.

Indice

Elenco delle tabelle	V
Elenco delle figure	VI
1 Introduzione	1
1.1 Anatomia e fisiologia della Prostata	1
1.2 Carcinoma prostatico	3
1.2.1 Incidenza e fattori di rischio	3
1.2.2 Diagnosi	5
1.2.3 Gleason score	8
1.2.4 Trattamento	11
1.3 Digital pathology	14
1.4 Uso del Deep Learning in digital pathology	17
1.4.1 Introduzione al Deep Learning	17
1.4.2 Deep learning per la classificazione di immagini istologiche di carcinoma prostatico	20
2 Materiali e metodi	23
2.1 Dataset	23
2.2 Architettura dell'algoritmo	24
2.3 Pre-processing	25
2.3.1 Rifinitura manuale delle annotazioni	25
2.3.2 Creazione maschere	26
2.3.3 Divisione dataset	28
2.3.4 Patch extraction	28
2.4 Allenamento CNN	29

2.4.1	Allenamento Alexnet	30
2.4.2	Allenamento DenseNet	31
2.5	Heatmap	32
2.6	Metriche di validazione	32
2.6.1	Confusion matrix	32
2.6.2	Curva ROC e AUC	35
3	Risultati	37
3.1	Risultati Alexnet	37
3.1.1	Risultati allenamento	37
3.1.2	Risultati test	39
3.2	Risultati DenseNet	40
3.2.1	Risultati allenamento	40
3.2.2	Risultati test	41
3.3	Risultati heatmap	42
4	Conclusioni e sviluppi futuri	43
4.1	Conclusioni	43
4.2	Sviluppi futuri	44
	Bibliografia	47

Elenco delle tabelle

1.1	Corrispondenza tra Gleason Score e Grade Group. . . .	10
2.1	Composizione del dataset utilizzato.	23
2.2	Distribuzione delle patch sane e tumorali estratte all'interno di training set, validation set e test set.	29
3.1	Tabella riassuntiva risultati ottenuti per la rete Alexnet.	38
3.2	Tabella risuntiva dei risultati ottenuti per la DenseNet.	40

Elenco delle figure

1.1	Suddivisione della prostata nelle varie zone: in verde si nota lo stroma fibro-muscolare anteriore, in blu la zona periferica, in viola la zona centrale e in rosso la zona di transizione.	2
1.2	Evoluzione del Gleason score. A) classificazione originale dei pattern del Gleason score. B) modifica del Gleason score proposta dall'International Society of Urological Pathology (ISUP) nel 2005. C) modifica del Gleason score proposta dall'ISUP nel 2014.	9
1.3	WSI di tessuto prostatico colorate ematossilina e eosina, ma che presentano una colorazione diversa. Ciò può essere dovuto ad una diversa concentrazione dei coloranti, ad un tempo di tintura differente o all'utilizzo di scanner diversi per la digitalizzazione.	16
1.4	Confronto tra Deep Learning e Machine Learning: la feature extraction nel Deep Learning viene effettuata automaticamente dall'algoritmo, mentre nel caso del Machine Learning le features devono essere estratte da parte di un operatore.	18
1.5	Struttura di una rete neurale convoluzionale.	19
2.1	Algoritmo ad alto livello.	24
2.2	Rifinitura annotazioni manuali dell'immagine P18_M appartenente al gruppo 2.	26

2.3	Maschere ottenute dalle annotazioni manuali dell'immagine P18_M appartenente al gruppo 2: b) frustolo annotato c) tessuto tumorale d) tessuto sano all'interno di tessuto tumorale.	27
2.4	Esempi patch estratte dalla stessa WSI (66-18-B1 del gruppo 3): (a) patch sana, (b) patch tumorale.	29
2.5	Architettura di Alexnet.	30
2.6	Architettura DenseNet. Si nota come ogni layer della DenseNet sia collegato con tutti gli altri layer e non solo con quelli direttamente successivi.	31
2.7	Esempio di confusion matrix per classificazione binaria.	33
2.8	Esempio di curva ROC.	35
3.1	Learning curve ottenuta dall'allenamento di Alexnet.	37
3.2	Confusion matrix ottenute da Alexnet per training set (a sinistra) e validation set (a destra).	38
3.3	Confusion matrix Alexnet per il test set.	39
3.4	Curve ROC del sistema con Alexnet ottenuta sul test set.	39
3.5	Confusion matrix DenseNet per training set (a sinistra) e validation set (a destra).	40
3.6	Confusion matrix ottenuta per il test set con DenseNet.	41
3.7	Curva ROC ottenuta per la DenseNet.	41
3.8	A sinistra, un campo della WSI, e a destra la heatmap corrispondente.	42

Capitolo 1

Introduzione

In questo capitolo vengono presentate le principali caratteristiche anatomiche e fisiologiche della prostata, a cui farà seguito la descrizione del carcinoma prostatico, indicandone epidemiologia, fattori di rischio, iter diagnostico, valutazione del grading tramite *Gleason score* e possibile trattamento della patologia. In seguito viene illustrato il campo della *digital pathology*, delle sfide e delle opportunità che questo offre nell'ambito dell'analisi di immagini istologiche. Infine si espone lo stato dell'arte riguardante l'uso del Deep Learning in ambito della classificazione di immagine mediche di tipo istologico.

1.1 Anatomia e fisiologia della Prostata

La prostata è una ghiandola esocrina che fa parte dell'apparato genitale maschile, situata nel compartimento sub-peritoneale, tra il diaframma pelvico e la cavità peritoneale, posteriormente alla sinfisi pubica, anteriormente al retto e inferiormente alla vescica urinaria. La posizione di questa ghiandola fa sì che si possa esaminare facilmente tramite palpazione digitale [1].

In un uomo adulto, in condizioni normali, ha dimensioni e forma assimilabili a una castagna, avente la base adiacente alla superficie inferiore della vescica, e l'apice rivolto verso il basso, e ha un peso compreso tra i 10 e i 20 g.

La prostata è divisa in 4 zone (vedi Figura 1.1), aventi diverse caratteristiche e origine embriologica :

- **Zona Centrale (CZ)** : area conica presente nell'individuo adulto, avente l'apice del cono posta nel collicolo seminale, ovvero alla confluenza tra dotti eiaculatori e uretra prostatica.
- **Zona di Transizione (TZ)** : zona interessata perlopiù da iperplasia benigna, in questa zona si sviluppano solo il 25 % dei carcinomi prostatici.
- **Zona Periferica (PZ)** : zona composta dal tessuto ghiandolare prostatico situato all'apice e nella porzione posteriore della prostata, in cui si sviluppano il 70 % dei carcinomi prostatici.
- **Stroma fibro-muscolare anteriore (AFMS)** : zona composta da tessuto fibro-muscolare in cui la parte distale è importante nella funzione di sfintere volontario mentre la parte prossimale gioca un ruolo centrale nelle funzioni di sfintere involontario.

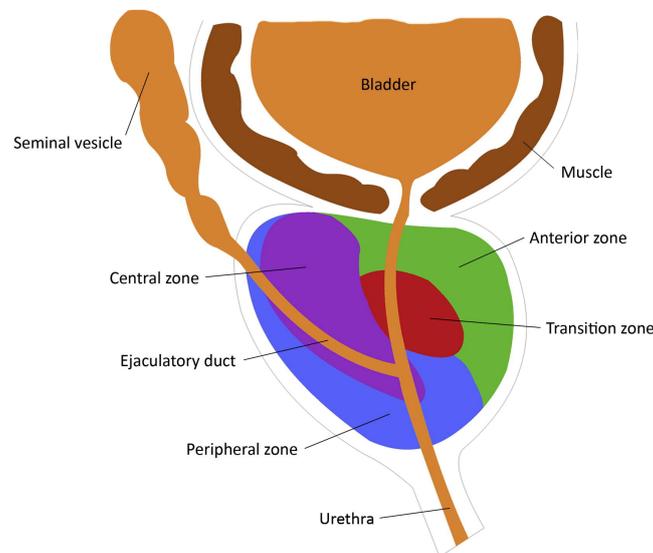


Figura 1.1: Suddivisione della prostata nelle varie zone: in verde si nota lo stroma fibro-muscolare anteriore, in blu la zona periferica, in viola la zona centrale e in rosso la zona di transizione. Immagine tratta da [2].

La prostata è composta da elementi ghiandolari e stromali, fusi insieme in una pseudocapsula composta da uno strato interno di tessuto muscolare liscio con uno strato esterno ricoperto da collagene [1].

Durante l'eiaculazione, le ghiandole prostatiche sono compresse per azione della muscolatura liscia adiacente, e le secrezioni prodotte vengono sospinte all'interno dell'uretra. Queste secrezioni prostatiche comprendono citrato, che gli spermatozoi utilizzano come fonte di energia, e molti enzimi tra cui l'antigene specifico della prostata (PSA). Il PSA agevola lo scorrimento del seme all'interno del tratto riproduttivo femminile, inibendone la coagulazione. La produzione di PSA aumenta all'aumentare delle dimensioni della prostata, che a loro volta possono aumentare in seguito a varie patologie che interessano la ghiandola. Per questo motivo, un importante screening iniziale per la diagnosi di carcinoma prostatico consiste nel dosaggio ematico del PSA: se il livello di PSA è elevato, è necessario approfondire con ulteriori indagini, in maniera da confermare o confutare il sospetto diagnostico. Infatti bisogna ricordare che solo il 30 % degli uomini che presenta alti livelli di PSA manifesta il cancro [3].

1.2 Carcinoma prostatico

L'adenocarcinoma prostatico è una delle patologie più frequenti che interessa la prostata, insieme all'iperplasia prostatica benigna. Generalmente si presenta in maniera asintomatica, soprattutto nelle fasi iniziali, infatti è spesso diagnosticato dopo il riscontro di un livello elevato di PSA. Ciò nonostante, a volte causa dei sintomi simili a quelli dell'iperplasia prostatica benigna, che comprendono sia problemi urinari che problemi legati alla funzione sessuale, quali la difficoltà a raggiungere l'erezione e l'eiaculazione dolorosa [4].

1.2.1 Incidenza e fattori di rischio

Il carcinoma prostatico, con una stima di quasi 1.4 milioni di nuovi casi e 375 mila morti in tutto il mondo, è la seconda neoplasia per frequenza e la quinta per mortalità tra la popolazione maschile nel 2020.

L'incidenza di questa patologia è tre volte superiore nei paesi sviluppati rispetto a quelli in via di sviluppo, nonostante la mortalità non presenti una variabilità così ampia. Inoltre, è il tumore diagnosticato più frequentemente nella popolazione maschile in oltre metà dei paesi del mondo, nonché la principale causa di morte per cancro negli uomini in 48 paesi.

Sebbene sia una tipologia di tumore così diffusa, la sua eziologia rimane dubbia, ma nonostante ciò sono stati riconosciuti alcuni fattori di rischio [5], quali:

- età avanzata (maggiore di 50 anni);
- storia familiare di carcinoma prostatico;
- mutazioni genetiche (ad esempio BRCA1 and BRCA2);
- alcune patologie, come la sindrome di Lynch;
- ascendenza dall'Africa occidentale: infatti i tassi più elevati di carcinoma prostatico sono presenti nella popolazione nera maschile degli Stati Uniti d'America e dei Caraibi.
- fumo;
- peso corporeo eccessivo;

Inoltre ha la più alta prevalenza a 5 anni tra tutti i tipi di cancro, rappresentando il 25 % di tutti i tumori prevalenti. Questo ha importanti implicazioni per quanto riguarda la decisione di sottoporre gli uomini ad un trattamento o ad un regime di sorveglianza per questa malattia [6].

Data l'enorme diffusione di questa neoplasia nella popolazione sopra i 50 anni di età e le ottime possibilità di eradicazione durante le fasi iniziali, risulta evidente l'importanza che ricopre la diagnosi precoce di questa patologia.

1.2.2 Diagnosi

La diagnosi di carcinoma prostatico si basa essenzialmente su 4 tipologie di indagini: esplorazione rettale, dosaggio del PSA, tecniche di imaging e agobiopsia prostatica.

Esplorazione rettale

Il primo approccio diagnostico che viene offerto a coloro i quali presentano dei sintomi che potrebbero essere riconducibili ad una patologia prostatica consiste nell'esplorazione rettale. Tramite questa tecnica è possibile rilevare facilmente i noduli neoplastici dovuti al carcinoma prostatico, dato che oltre il 70 % di questi tumori insorgono nella zona periferica della prostata. Infatti, circa il 18 % delle neoplasie prostatiche sono rilevate grazie alla sola palpazione digitale, indipendentemente dai valori ematici del PSA [7]. Però, nonostante questa tecnica diagnostica risulti fondamentale per la valutazione del paziente, essa non può essere l'unica metodica diagnostica utilizzata, perché presenta bassi livelli di sensibilità e specificità [8].

Dosaggio del PSA

Un'altra indagine importante che consenta la diagnosi del carcinoma prostatico consiste nel dosaggio del livello di PSA nel sangue. Infatti, in condizioni fisiologiche, solo minime quantità di questo enzima raggiungono il circolo ematico, ma in alcune condizioni si può verificare un innalzamento del livello di PSA che è dovuto alla modifica della normale architettura del tessuto prostatico, come per esempio in presenza di lesioni maligne o benigne. Perciò il dosaggio del livello di PSA è un utile indicatore per la presenza di una possibile patologia legata alla prostata. Però bisogna ricordare che il PSA potrebbe essere elevato anche in condizioni fisiologiche, o in caso di tumore benigno, oppure anche dopo l'esecuzione di alcune manovre diagnostiche, tra cui la cistoscopia, ovvero l'endoscopia della vescica urinaria attraverso l'uretra, o la biopsia prostatica. Dunque, la decisione sull'effettuare o meno una biopsia non si deve basare solo sul risultato ottenuto dal

dosaggio del PSA, ma bisogna integrare questa informazione i risultati dell'esplorazione rettale e con la valutazione di eventuali fattori di rischio e informazioni cliniche relative al paziente. Quindi un singolo valore anomalo di PSA non deve indurre a decisioni cliniche, ma bisogna confermare il sospetto tramite altre rilevazioni nelle settimane successive.

Attualmente, dato che non si sa molto sulle cause di questa patologia, non si può intervenire sulla prevenzione primaria per ridurre il tasso di incidenza del carcinoma prostatico. Invece, si potrebbe intervenire sulla prevenzione secondaria, ovvero sulla diagnosi precoce, per ridurre la letalità, e in particolare svolgendo uno screening della popolazione interessata. Il test di screening che è apparso potenzialmente più confacente allo scopo, per considerazioni complessive di costi, convenienza e accuratezza diagnostica, è il dosaggio periodico del PSA. Questo metodo di screening da un lato ha portato ad una riduzione della mortalità dovuta al cancro di circa il 20 %, ma allo stesso tempo è anche aumentato il rischio di sovradiagnosi, e di conseguenza di sovratrattamento, associato al dosaggio di PSA. Infatti, nell'ultimo ventennio nei paesi occidentali è stato registrato un drastico aumento dell'incidenza di carcinoma prostatico, il che coincide con il periodo in cui si è diffusa la pratica dello screening del PSA. Infatti, in questo modo, sono stati portati alla luce molti casi che altrimenti sarebbero passati inosservati e verosimilmente non avrebbero causato problemi al paziente. Una conferma di ciò è data dal fatto che il tasso di mortalità legato alla patologia registrato nello stesso periodo di tempo è rimasto sostanzialmente stabile. Dunque, uno screening con il PSA può essere offerto agli uomini asintomatici di età compresa tra i 50 e i 75 anni che lo richiedano, oppure agli uomini di età compresa tra i 40 e i 50 anni con fattori di rischio, purché essi vengano adeguatamente informati sui benefici e sui rischi conseguenti. Bisogna infine ricordare che nella valutazione del PSA si riscontra un tasso di falsi negativi nel 20-25 % dei tumori clinicamente significativi iniziali, e che i tumori più indifferenziati possono esordire con valori di PSA basso. Pertanto va tenuto in considerazione il rischio di una possibile falsa rassicurazione data da un risultato negativo del test [8].

Tecniche di imaging

Per definire in maniera più precisa le lesioni prostatiche sospette, spesso viene utilizzata la risonanza magnetica multi-parametrica, che permette di ottenere importanti informazioni sia anatomiche che funzionali [9]. Questa tecnica di imaging è particolarmente indicata per la guida del campionamento bioptico in pazienti con sospetto clinico di carcinoma prostatico e con precedenti biopsie negative [10], perché precedendo la biopsia con una risonanza magnetica, si riescono ad individuare un maggior numero di tumori clinicamente significativi rispetto alla biopsia standard [11]. Infatti in questo modo si riduce l'individuazione di tumori prostatici clinicamente indolenti e si evita la biopsia in metà dei pazienti sottoposti a risonanza magnetica [12].

Un'altra tecnica di imaging utile per la diagnosi del carcinoma prostatico è l'ecografia prostatica trans-rettale, che consiste nell'inserimento di una sonda nel retto a contatto diretto con la prostata, a scopo di ottenere informazioni esatte sul volume dell'organo e la presenza di eventuali noduli dovuti a calcoli prostatici. Questa tecnica viene effettuata soprattutto per visualizzare la prostata durante una biopsia, mentre ha una bassa affidabilità diagnostica nell'identificare le neoplasie.

Agobiopsia prostatica

L'agobiopsia prostatica è una procedura indispensabile per confermare o escludere la diagnosi di carcinoma prostatico, che consiste nel prelievo di tessuto sospetto e dell'esame istologico dei campioni prelevati, in maniera da determinare la posizione e il grado di un eventuale tumore. L'indicazione all'esecuzione di questo esame può derivare dal risultato anomalo dei test diagnostici presentati precedentemente, quindi dal riscontro di una consistenza insolita della ghiandola all'esplorazione rettale, dal rilievo di aree sospette all'ecografia trans-rettale o alla risonanza magnetica o da un'elevazione dei valori sierici di PSA.

Attualmente il gold standard per questa tecnica diagnostica consiste nella biopsia prostatica eco-guidata, che può essere eseguita per via trans-rettale o trans-peritoneale. Le percentuali diagnostiche tra i

due approcci sono comparabili, tuttavia alcune evidenze suggeriscono un minor rischio di infezioni per l'approccio trans-perineale [8]. La procedura avviene in anestesia locale e consiste nel prelievo per mezzo di un ago sottile di frustoli di tessuto prostatico, provenienti da diverse aree sospette della ghiandola evidenziate dall'esplorazione rettale o dalle tecniche di imaging precedentemente effettuate.

Si ottengono dei campioni di tessuto che vengono trattati e colorati con tinture apposite per poi essere analizzati al microscopio da un patologo esperto che per ricercare le eventuali cellule neoplastiche e stabilire il grado del tumore valutando il *Gleason score*.

1.2.3 Gleason score

Il Gleason score è un sistema per la valutazione del grado del carcinoma prostatico che si basa sul grado di differenziazione del tessuto ghiandolare. Viene assegnato un punteggio da 1 a 5 ai due aspetti strutturali più rappresentati all'interno del tessuto in esame: quello prevalente viene definito primario e il secondo pattern più presente viene definito secondario. Questo punteggio aumenta al crescere della malignità del tumore, dunque maggiore è il Gleason score, maggiore sarà l'aggressività della neoplasia e la probabilità che essa si diffonda in altre zone del corpo.

Nella formulazione originaria, sviluppata nel 1966, il Gleason score prevedeva la somma del punteggio del grado primario e del grado secondario, che veniva definito come il pattern più rappresentato dopo il quello primario. Tuttavia, nel 2005, l'International Society of Urological Pathology (ISUP) propose una modifica a questo sistema per correggerne alcune imprecisioni. Infatti, secondo la classificazione ISUP del 2005, il pattern secondario non sarà più il secondo più rappresentato, ma viene definito come il pattern peggiore presente, ad esclusione del Gleason primario. Quando non esiste un grado secondario, si deve raddoppiare il grado primario per ottenere il Gleason score.[8]

I gradi di Gleason (pattern) attribuiti al tessuto neoplastico secondo la classificazione ISUP 2005 [14] sono:

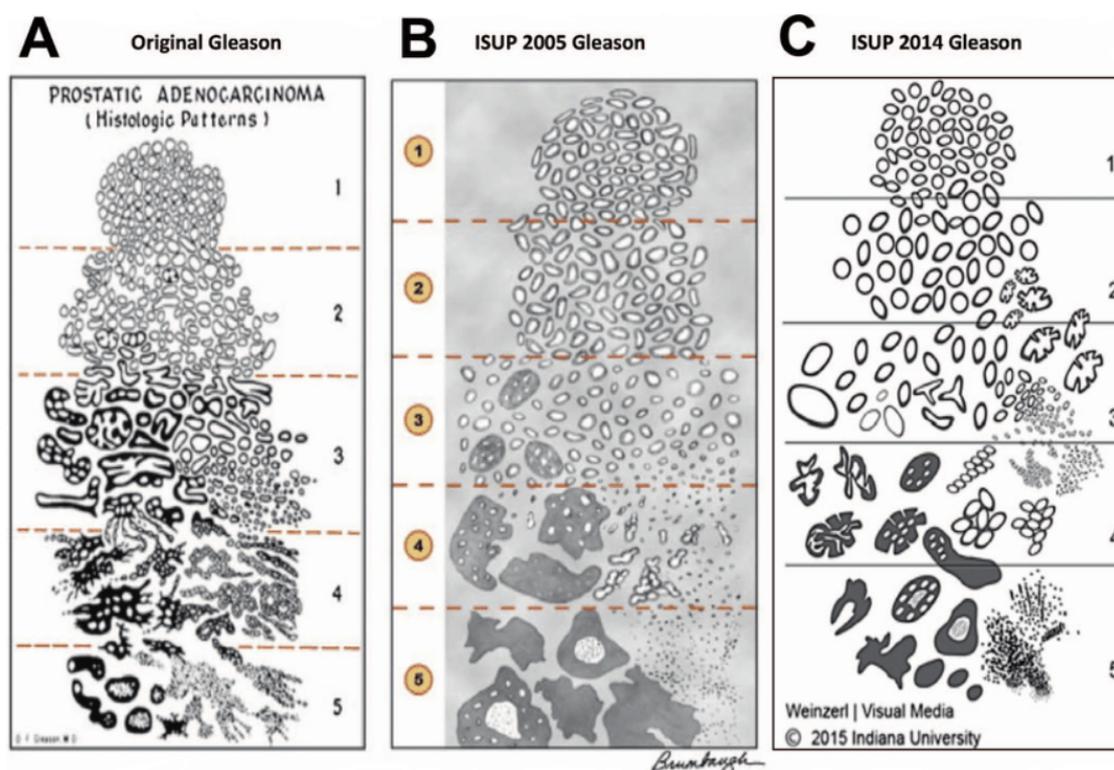


Figura 1.2: Evoluzione del Gleason score. A) classificazione originale dei pattern del Gleason score. B) modifica del Gleason score proposta dall'International Society of Urological Pathology (ISUP) nel 2005. C) modifica del Gleason score proposta dall'ISUP nel 2014. Immagine tratta da [13].

- **grado 1:** nodulo circoscritto di acini fitti ma distinti, uniformi, rotondi o ovali, di medie dimensioni (ghiandole più grandi del pattern 3);
- **grado 2:** nodulo relativamente circoscritto, ma ai margini possono essere presenti minime infiltrazioni. Le ghiandole sono disposte in modo meno serrato e uniforme rispetto al pattern 1;
- **grado 3:** unità ghiandolari discrete; in genere le ghiandole sono più piccole di quanto visto nel modello 1 e 2. infiltrati sono presenti tra gli acini non neoplastici. notevole variabilità in forma e dimensione, talora con aspetti cribriformi;

- **grado 4:** ghiandole micro-acinari fuse, mal definibili, con lume ghiandolare scarsamente formato; ghiandole cribriformi, anche con bordi irregolari; talora aspetti ipernefromatoidi;
- **grado 5:** relativa assenza di differenziazione ghiandolare; cordoni solidi composti o singole cellule; comedocarcinoma con necrosi centrale circondata da masse papillari, cribriformi o solide;

Tuttavia, alcuni aspetti critici del sistema Gleason originale persistevano nonostante la modifica effettuata. Infatti, nonostante il Gleason score potesse assumere valori compresi tra 2 e 10, il punteggio più basso predominante consisteva nel Gleason $3+3=6$, il che poteva essere interpretato come un tumore mediamente aggressivo ed essere causa di ansia non necessaria nei pazienti. Inoltre il sistema non permetteva di distinguere adeguatamente un Gleason $3+4=7$ da un $4+3=7$, con quest'ultimo che ha una prognosi decisamente peggiore del precedente. Per questo motivo il sistema è stato nuovamente revisionato nel 2014.

L'Organizzazione Mondiale della Sanità raccomanda di riportare sempre la percentuale di grado 4 sia nelle agobiopsie che nelle prostatectomie radicali, dato l'importante valore prognostico rappresentato da questo tipo di pattern di Gleason.

A questo proposito sono stati sviluppati 5 diversi gruppi di grado (Grade Group) che hanno incontrato il consenso di specialisti patologi ed urologi alla conferenza ISUP del 2014 [15] definiti in base al Gleason score. I gruppi sono illustrati nella tabella 1.1.

GRADE GROUP	GLEASON SCORE
Grade Group 1	≤ 6
Grade Group 2	$3+4 = 7$
Grade Group 3	$4+3 = 7$
Grade Group 4	$4+4=8, 3+5=8, 5+3=8$
Grade Group 5	9 - 10

Tabella 1.1: Corrispondenza tra Gleason Score e Grade Group.

Attualmente si raccomanda di riportare sempre il gruppo di grado (Grade Group) insieme al Gleason score modificato secondo ISUP 2014.

1.2.4 Trattamento

Il trattamento offerto in caso di tumore alla prostata prende in considerazione vari fattori, tra cui l'estensione anatomica e l'aggressività della malattia, ma anche le attese del paziente e la sua speranza di vita associata alla presenza di comorbidità. Dunque, in base a questi elementi si decide se sottoporre il paziente ad una politica di vigile attesa o sorveglianza attiva, oppure se sottoporlo a trattamento con scopo radicale, quali la prostatectomia radicale o la radioterapia.

Vigile attesa

La vigile attesa è una politica di sorveglianza offerta ai pazienti nei quali il trattamento immediato del tumore non è in grado di impattare positivamente sulla loro speranza di vita. Infatti la vigile attesa è generalmente indicata per pazienti con una speranza di vita inferiore a 10 anni, a causa dell'età avanzata o della presenza di comorbidità con più elevata letalità. Pertanto, in questi casi, eventuali terapie possono essere dilazionate alla comparsa dei sintomi, con finalità pressoché esclusivamente palliative.

Sorveglianza attiva

La sorveglianza attiva è una strategia di trattamento differito che viene offerta ai pazienti con una diagnosi a basso rischio. A questi pazienti, invece del trattamento immediato, viene offerto uno stretto monitoraggio al fine di rilevare tempestivamente l'eventuale progressione della malattia. Inoltre, questo periodo di osservazione iniziale mira ad evitare trattamenti non necessari dei tumori innocui e le relative complicanze, pur fornendo un intervento tempestivo per gli uomini che ne hanno bisogno.

Ciò che contraddistingue maggiormente la sorveglianza attiva rispetto alla vigile attesa è la capacità di correggere il possibile errore di stadiazione iniziale, attraverso le procedure di monitoraggio, prima fra tutte la ripetizione della biopsia prostatica.

La sorveglianza attiva implica regolari test di follow-up per monitorare la progressione del tumore alla prostata: analisi del sangue, esami rettali e biopsie. Quando le prove dimostrano che la malattia sta progredendo, si può optare per un trattamento, come la chirurgia o la radioterapia.

Prostatectomia radicale

Con il termine di prostatectomia radicale (PR) si intende la rimozione chirurgica della prostata, delle ampolle deferenziali e delle vescicole seminali, insieme ad una porzione del tessuto circostante sufficiente per ottenere margini chirurgici negativi. L'obiettivo di questa procedura chirurgica è l'eradicazione della patologia e, quando possibile, la preservazione della continenza e della funzionalità erettile, elementi che restano però subordinati al fine oncologico [8].

La prostatectomia radicale può essere eseguita tramite intervento a cielo aperto o attraverso tecniche mini-invasive, come la PR laparoscopica o la PR laparoscopica robot-assistita. Le tecniche mini-invasive offrono alcuni vantaggi, quali la riduzione delle perdite ematiche intraoperatorie, diminuzione del rischio di trasfusioni perioperatorie, riduzione delle giornate di degenza, minor consumo di analgesici e un miglior risultato estetico, anche se a discapito di un aumento dei tempi operatori e dei costi. Per quanto riguarda i risultati oncologici, non vi è evidenza scientifica che i risultati ottenuti con la chirurgia mini-invasiva siano superiori a quelli ottenibili con le tecniche chirurgiche convenzionali [16].

Gli studi che hanno confrontato i risultati della prostatectomia radicale con quelli della vigile attesa, hanno evidenziato un vantaggio in termini di sopravvivenza e di riduzione del rischio di progressione locale e di diffusione a distanza a favore dei pazienti trattati con chirurgia. Tale vantaggio si è rivelato evidente in particolare nei pazienti giovani, ovvero con un'età inferiore a 65 anni, e a rischio intermedio [17]. Emerge quindi l'indicazione a prendere in considerazione l'opzione chirurgica quando indicata soprattutto nei pazienti con una speranza di vita adeguata e in assenza di comorbidità o evidenti controindicazioni

di tipo chirurgico. Tuttavia bisogna ricordare che il beneficio clinico ottenuto con il trattamento chirurgico immediato è bilanciato da una maggiore incidenza di effetti collaterali. Dunque, la prostatectomia radicale sembra presentare un vantaggio in sopravvivenza globale e cancro-specifica rispetto alla vigile attesa almeno per alcuni gruppi di pazienti. Per questo motivo l'uso della vigile attesa nella pratica clinica è limitato a pazienti selezionati, ovvero pazienti anziani, con breve aspettativa di vita e non adatti per trattamenti locali con fini di radicalità [8].

Radioterapia

Un'altra opzione di trattamento a scopo radicale del carcinoma prostatico localizzato consiste nella radioterapia, e in particolare nella radioterapia a fasci esterni e nella brachiterapia, ovvero la collocazione di una sorgente di radiazioni all'interno o vicino alla zona da trattare.

Lo studio PROTECT ha dimostrato che nei pazienti a rischio intermedio o inferiore, la prostatectomia radicale, la radioterapia a fasci esterni e la brachiterapia producono risultati a 10 anni sovrapponibili in termini di sopravvivenza globale, di ricaduta biochimica e di ricaduta a distanza della malattia [18].

Parametri importanti da valutare in radioterapia oncologica sono:

- **dose totale** da somministrare, che abitualmente è ≥ 76 Gy per ottenere una buona riuscita del trattamento in tutte le classi di rischio;
- **dose somministrata ad ogni seduta**, che può essere convenzionale, ovvero intorno a 2 Gy, oppure ipofrazionata, cioè a dosaggio più alto ma nell'arco di meno sedute, permettendo uguale efficacia e tossicità ma riducendo la durata complessiva del trattamento;
- **volume del trattamento**, ovvero il volume che riceve il trattamento radioterapico, che varia in base alla classe di rischio.

La brachiterapia prevede l'acquisizione dei dati anatomici del paziente con ecografia e la determinazione della geometria d'impianto

ottimale dal punto di vista dosimetrico tramite software appositi. La procedura chirurgica d'impianto è attuata per via perineale, in anestesia generale, mediante guida ecografica e fluoroscopica. Gli isotopi inseriti sono Iodio o Palladio, sigillati in capsule di titanio.

1.3 Digital pathology

La patologia digitale è la pratica dell'acquisire, condividere e interpretare materiale istopatologico in un ambiente digitale. Per fare ciò bisogna ottenere delle immagini istologiche digitali tramite un processo di *Whole Slide Imaging (WSI)*, che prevede la scansione dei preparati istologici per mezzo di scanner appositi che forniscono un'immagine digitale ad alta risoluzione. Le operazioni da eseguire per la preparazione dei vetrini istologici e la successiva digitalizzazione sono [19]:

1. **Prelievo dei tessuti e fissazione:** processo fondamentale per mantenere intatto l'aspetto del tessuto istologico ed evitare che subisca alterazioni di tipo chimico o fisico. Consiste nel trattare il campione con un agente fissante, la cui scelta dipende dalla tipologia di tessuto del campione in esame e dal tipo di analisi che si vuole portare avanti.
2. **Deidratazione:** procedimento che consiste nella rimozione dell'acqua dal tessuto. Nel caso in cui quest'operazione non venisse eseguita adeguatamente, alcune regioni del campione potrebbero risultare opache a causa della presenza di gocce d'acqua. In questo caso si andrebbero a perdere dettagli microscopici del vetrino e si potrebbero ottenere anche cambiamenti inaspettati nella colorazione delle strutture del tessuto.
3. **Inclusione in paraffina:** incorporazione del campione in paraffina riscaldata, con lo scopo di far evaporare il solvente usato nelle operazioni precedenti e di riempire gli spazi contenuti nel tessuto.
4. **Microtomia:** sezionamento del blocco di paraffina, in modo da ottenere delle fette di tessuto aventi uno spessore tale da poter

identificare sia l'architettura del tessuto che la morfologia cellulare. Lo spessore ideale a tale scopo è di 5 μm . Tuttavia durante questa operazione è necessario evitare di posizionare in maniera imprecisa il campione sul vetrino, perché questo potrebbe causare il ripiegamento del tessuto su sé stesso.

5. **Colorazione:** tintura del tessuto istologico con lo scopo di evidenziarne le strutture cellulari. La colorazione finale della sezione di tessuto dipende da vari fattori, quali il pH, la concentrazione di colorante e il tempo di tintura. Generalmente nel caso di immagini prostatiche si utilizza una colorazione con ematossilina-eosina.
6. **Montaggio:** processo di inclusione delle sezioni di tessuto all'interno di un vetrino, al fine di proteggere il tessuto da agenti esterni. Gli artefatti in cui si può incorrere durante questa operazione sono dovuti alla presenza di polvere, di bolle d'aria o contaminazione del tessuto con microrganismi.
7. **Digitalizzazione:** scansione del vetrino istologico tramite scanner appositi. La tipologia di scanner utilizzato può variare l'aspetto dell'immagine ottenuta, che può risultare diversa scansionando con scanner diversi la medesima sezione istologica. Inoltre, per evitare di ottenere immagini sfocate, è necessario allineare attentamente il campione con il fuoco dello scanner.

L'utilizzo di immagini digitali per l'analisi istologica presenta numerosi vantaggi rispetto alla classica analisi microscopica, e consente di superarne alcuni limiti. Infatti la digital pathology permette di trasferire con facilità i dati tra luoghi distanti senza aver bisogno spedire fisicamente i preparati istologici, consentendo il consulto tra patologi in remoto e favorendo una collaborazione coerente e trasparente, migliorando l'efficienza e la produttività [20]. Inoltre si riduce il rischio di rottura o smarrimento dei vetrini, che non hanno più bisogno di essere conservati in loco ma sono resi accessibili rapidamente e da remoto tramite un sistema di archiviazione centrale.

D'altra parte la digitalizzazione dei preparati istologici permette l'applicazione di algoritmi automatici per l'analisi delle immagini, che

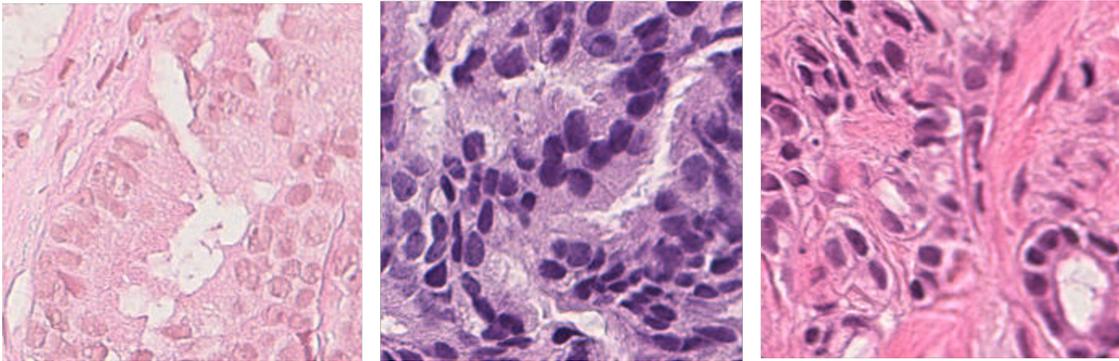


Figura 1.3: WSI di tessuto prostatico colorate ematossilina e eosina, ma che presentano una colorazione diversa. Ciò può essere dovuto ad una diversa concentrazione dei coloranti, ad un tempo di tintura differente o all'utilizzo di scanner diversi per la digitalizzazione.

hanno il vantaggio di essere più precisi, veloci e obiettivi rispetto alla microscopia, e di poter ricavare informazioni difficilmente reperibili altrimenti. Questo processo di automazione permette di anche ridurre il carico di lavoro a carico dei patologi, consentendo orari lavorativi più flessibili. Inoltre bisogna ricordare che le valutazioni dei patologi nell'analisi di immagini istologiche sono affette da grande variabilità intra-operatore e inter-operatore, ma applicando dei metodi automatici che forniscono una seconda opinione si può ridurre questa variabilità.

Gli algoritmi per l'analisi delle immagini tradizionali applicati alla digital pathology devono tenere in conto le grandi variazioni che ci possono essere a livello di morfologia, tessitura e colorazione del tessuto (vedi Figura 1.3), ma senza risultare troppo generale né troppo specifico, perché ciò potrebbe portare ad errori di classificazione [21]. Per questo motivo spesso in quest'ambito trovano grande applicazione metodi di Deep Learning, che riescono a tenere conto di questa grande variabilità dato che questi algoritmi apprendono automaticamente le variabili necessarie per compiere il task utilizzando solo il dataset messo a disposizione.

1.4 Uso del Deep Learning in digital pathology

L'applicazione di metodi basati sull'intelligenza artificiale (AI), e in particolare sul Deep Learning, per l'analisi delle immagini istologiche ha cambiato completamente il campo della digital pathology.

Questa tecnologia può essere utilizzata per effettuare diverse tipologie di task, come per esempio la classificazione, l'individuazione e la segmentazione di elementi all'interno di questo tipo di immagini. L'utilizzo dell'intelligenza artificiale permette inoltre di sviluppare dei sistemi automatici che possono essere usati come sistemi CAD (Computer Aided Diagnosis), ovvero dei sistemi di aiuto alla decisione che supporterebbero il patologo nella diagnosi fornendo una seconda opinione.

1.4.1 Introduzione al Deep Learning

Il campo dell'intelligenza artificiale è molto ampio, e comprende tutti quei metodi che cercano di svolgere un certo compito replicando e imitando l'intelligenza umana. In particolare alcuni di questi algoritmi, che utilizzano metodi di apprendimento automatico (Machine Learning), hanno la capacità di apprendere una propria base di conoscenza dai dati grezzi forniti come input, tramite la valutazione di una serie di variabili (features) e migliorare l'algoritmo stesso in base alla conoscenza raccolta.

Il Deep Learning è una branca del Machine Learning che comprende algoritmi in cui l'estrazione delle features da prendere in considerazione per la risoluzione di un dato task è effettuata dall'algoritmo stesso unicamente a partire dal dataset messo a disposizione, e non deve essere eseguita manualmente dall'operatore. Tuttavia una condizione necessaria per usare questo metodo è di avere a disposizione una grande quantità di dati da analizzare, in maniera da tenere conto in modo adeguata della variabilità presente tra gli elementi che compongono il dataset.

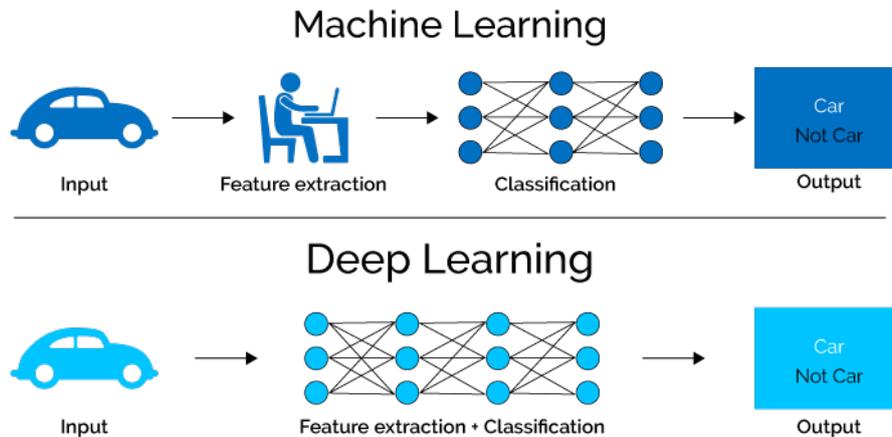


Figura 1.4: Confronto tra Deep Learning e Machine Learning: la feature extraction nel Deep Learning viene effettuata automaticamente dall'algoritmo, mentre nel caso del Machine Learning le features devono essere estratte da parte di un operatore. Immagine tratta da [22].

Il Deep Learning trova particolare applicazione nel campo dell'analisi delle immagini e nella digital pathology, a causa della mole di dati da analizzare e della notevole variabilità a livello di colorazione, e morfologia del tessuto presente in un dataset composto da WSI.

Le reti per la classificazione di immagini si basano su un tipo di apprendimento supervisionato, ovvero in cui l'allenamento viene svolto su un *training set* di cui si conosce la classificazione, con il fine di minimizzare l'errore presente tra la predizione del sistema e la classe reale di appartenenza.

Nello specifico, per la classificazione di immagini istologiche, generalmente si usano le reti neurali convoluzionali (CNN) che sono costituite da diversi strati (layer), come illustrato nella Figura 1.5:

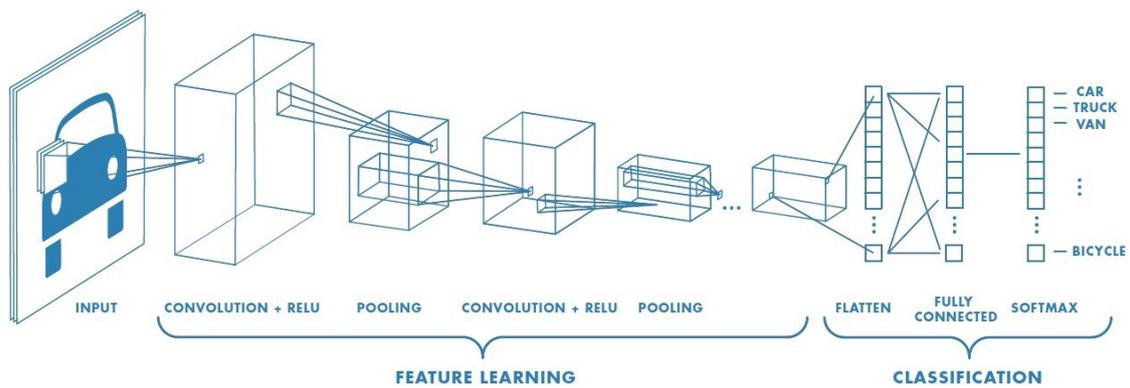


Figura 1.5: Struttura di una rete neurale convoluzionale. Immagine tratta da [23]

- **input layer:** riceve l'informazione da analizzare;
- **hidden layers:** svolgono varie operazioni al fine di estrarre le features dall'immagine che serviranno successivamente per la classificazione:
 - *layer convoluzionale:* effettua un'operazione di convoluzione tra l'input del layer e un kernel, o filtro, che scorre su di esso, ottenendo così un output a cui viene aggiunta non-linearità tramite l'applicazione di una funzione di attivazione. Le funzioni di attivazione sono molteplici, la più utilizzata è la ReLU;
 - *pooling layer:* fornisce una statistica riepilogativa dell'output vicino [24], ad esempio fornendo il valore massimo in una certa zona della matrice (max pooling) o facendo la media di questi valori (average pooling);
- **output layer:** blocco in cui avviene la classificazione, è composto da uno o più fully connected layer.

In ambito di digital pathology è molto utilizzato il transfer learning, ovvero l'utilizzo di reti pre-allenate per effettuare un certo task che vengono utilizzate in un ambito diverso, allenando solamente i layer adibiti alla classificazione, in maniera da rendere l'allenamento più veloce e efficiente rispetto ad allenare tutti i parametri della rete. Un

altro approccio molto utile in queste applicazioni consiste nella data augmentation, che prevede l'applicazione di trasformazioni casuali alle immagini di input per la rete neurale, in maniera da aumentare artificialmente la numerosità del dataset, in modo da rendere il modello più robusto ed evitare l'overfitting.

1.4.2 Deep learning per la classificazione di immagini istologiche di carcinoma prostatico

In letteratura sono presenti diversi lavori riguardanti la costruzione di un sistema automatico per la classificazione e il grading del tessuto istologico prostatico.

Un primo studio è quello svolto da Källén [25], in cui si è sviluppato un algoritmo per assistere i patologi nella determinazione del Gleason score utilizzando una CNN pre-allenata a cui sono stati rimossi i layer adibiti alla classificazione, che sarà invece effettuata tramite altre tecniche di Machine Learning. Il dataset utilizzato è composto da sole WSI con un Gleason score omogeneo, cosa che generalmente nella realtà non accade vista l'eterogeneità della composizione del tessuto. Quindi la parte di feature extraction a partire dalle patch viene effettuata con la rete convoluzionale pre-allenata OverFeat, mentre la classificazione delle patch avviene con 2 diversi classificatori, basati su Random Forest oppure Support Vector Machine. Invece la classificazione dell'immagine intera viene effettuata tramite un voto di maggioranza di tutte le patch che compongono un'immagine. L'accuratezza raggiunta da questo sistema è di 81.1 % nella classificazione delle patch e di 89.2 % nella classificazione dell'immagine intera.

Nel lavoro di Litjens [26] viene usato un dataset formato 225 WSI ottenute da biopsie scansionate con un ingrandimento 40x annotate manualmente da patologi, da cui sono state estratte delle patch in maniera casuale per effettuare l'allenamento della CNN. La dimensione ottimale delle patch si è rivelata essere 128x128, perché è un compromesso ideale tra accuratezza a livello di patch e memoria impegnata dalla GPU. Dopo l'allenamento, la CNN è stata convertita in una fully

convolutional network in maniera da poter essere applicata a l'immagine intera per ottenere una mappa di probabilità della presenza di tumore.

Nello studio di Zhou [27] ci si focalizza sulla distinzione tra Gleason pattern 3 e Gleason pattern 4, che costituisce un parametro importante per l'indicazione della prognosi e la determinazione di una diagnosi tumore con Gleason group 2 (3+4) o Gleason group 3 (4+3). Le ghiandole contenute nella WSI sono segmentate tramite k-means nello spazio colore lab. Successivamente viene applicata una color deconvolution e viene allenata una CNN per la classificazione. Questo approccio fornisce una valutazione del grado del tumore per i Gleason pattern intermedi senza aver bisogno di annotazioni manuali come ground truth.

Arvaniti [28] sfrutta un dataset composto da microarrays di tessuto prostatico per sviluppare un sistema di annotazione per il Gleason score che sfrutti le reti neurali convoluzionali. La CNN è stata sia allenata che valutata sulla base di annotazioni manuali effettuate da esperti, creando innanzitutto un classificatore di patch patch estratte dal dataset che è stato poi convertito in un annotatore a livello dei pixel, generando una una mappa di probabilità per ogni Gleason score applicando una sliding window. Inoltre la performance della rete è stata valutata in fase di testing comparando i risultati con quelli di una coorte di esperti, ottenendo buoni risultati e errori di classificazione soprattutto in caso di Gleason score vicini. In più i risultati della rete sono stati confrontati con i dati sulla sopravvivenza dei pazienti, in maniera da avere un ground truth più oggettivo rispetto alle annotazioni manuali che invece sono basate su valutazioni soggettive, dimostrando come il sistema sviluppato effettui una stratificazione delle diagnosi allo stesso livello di quella di patologi esperti.

Nello studio di Campanella [29], invece di avere a disposizione un dataset manualmente annotato dai patologi, viene usato un approccio di Deep Learning che usa come classi per il training solo le diagnosi riportate, quindi evitando le annotazioni manuali dei patologi, che sono costose sia in termini di tempo che monetario. Il dataset raccolto era composto da 24859 WSI ottenuti da biopsia prostatica, che rappresentano l'equivalente delle immagini istologiche generate in un laboratorio

di patologia nell'arco di un anno. Successivamente a questo modello è stato aggiunto un modello basato su RNN che integrano l'informazione sulla classificazione finale.

Nagpal [30] ha elaborato un sistema basato su Deep Learning che effettui una valutazione del Gleason score su campioni ottenuti da prostatectomia, che è stato validato da una coorte di patologi esperti. Il metodo è stato sviluppato in 2 fasi: CNN per la classificazione dei pattern di Gleason e k-NN (k-nearest neighbors) per la classificazione dell'intero vetrino istologico e l'assegnazione del Gleason grade group. Le annotazioni manuali fornite dai patologi sono servite per fornire la classe dei patch estratti dalle immagini, che a loro volta sono state usate per allenare la CNN.

Capitolo 2

Materiali e metodi

2.1 Dataset

Il dataset utilizzato in questo lavoro di tesi è composto da 906 WSI ottenute da biopsie prostatiche colorate tramite tintura di Ematossilina e Eosina, precedentemente annotate da uno o più patologi esperti. La suddivisione del dataset nei corrispettivi gruppi di grado è riassunta nella tabella 2.1, ed è la seguente: 100 WSI appartenenti al Gleason group 1, 100 al Gleason group 2, 100 al Gleason group 3, 100 al Gleason group 4, 100 al Gleason group 5, 204 WSI attribuite alla categoria "no group", che consiste in WSI annotate ma con Gleason group assente o discordante tra i patologi, e un 202 di WSI aventi solo tessuto sano.

Group 1	100
Group 2	100
Group 3	100
Group 4	100
Group 5	100
No Group	204
Healthy	202
totale	906

Tabella 2.1: Composizione del dataset utilizzato.

2.2 Architettura dell'algoritmo

Uno schermo riassuntivo degli step principali dell'algoritmo è mostrato nella Figura 2.1.

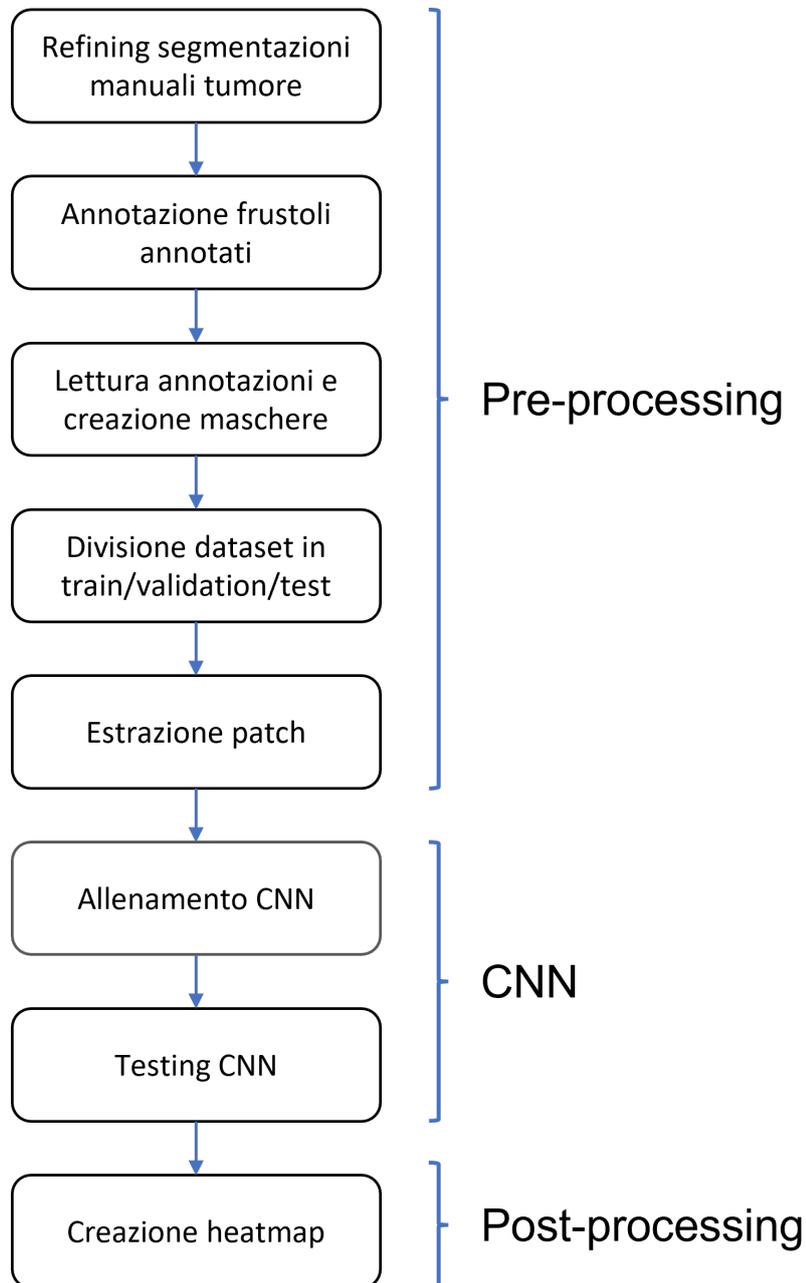


Figura 2.1: Algoritmo ad alto livello.

2.3 Pre-processing

Di seguito vengono indicate le strategie utilizzate per preparare le immagini prima dell'allenamento della rete neurale. Questa parte del processo è fondamentale al fine di ottenere dei buoni risultati nella classificazione, perché in assenza di pre-processing potrebbero venire forniti alla rete dati non informativi che potrebbero intaccarne la performance [31].

2.3.1 Rifinitura manuale delle annotazioni

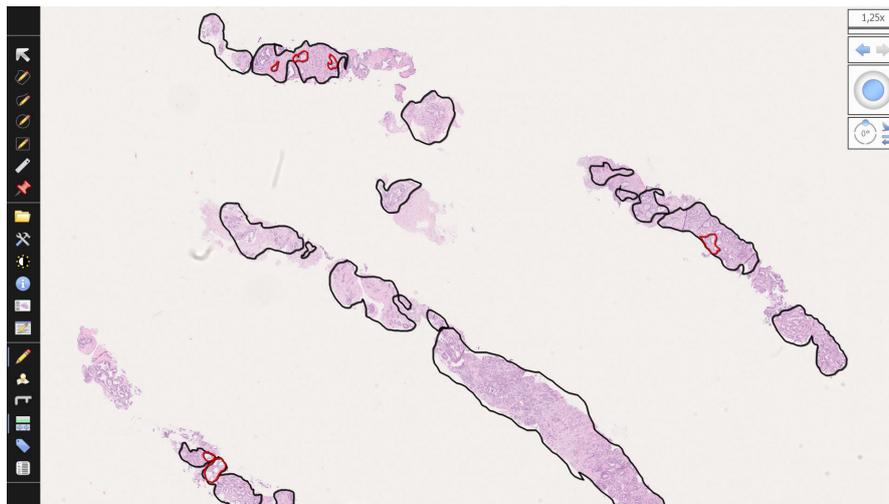
Le annotazioni del tumore fornite dai patologi sono state rifinite manualmente in modo da escludere dall'annotazione le regioni che chiaramente non contenevano tessuto ghiandolare, come illustrato nella Figura 2.2.

Per svolgere questa operazione sono stati adoperati due software differenti a seconda del formato dell'immagine: per le WSI in formato .ndpi è stato impiegato *NDP.view2* di Hamamatsu, mentre per quelle in formato .svs si è utilizzato *Aperio ImageScope* di Leica.

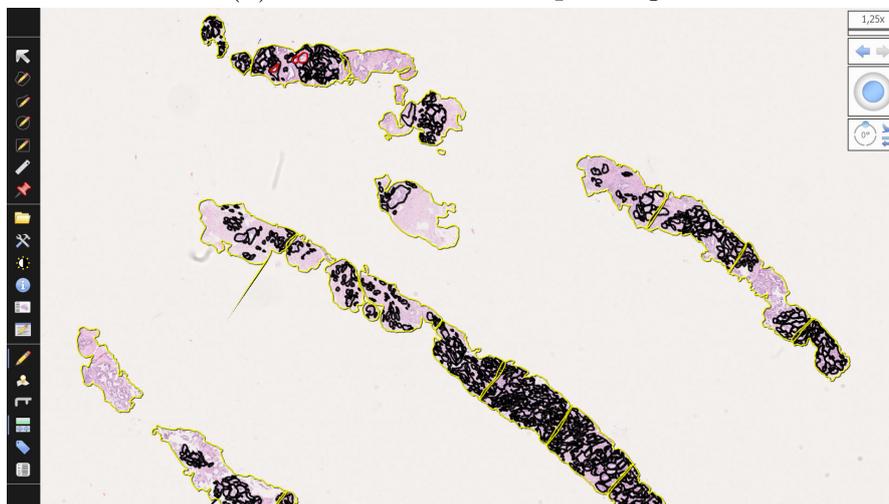
In seguito, dato che generalmente in ogni WSI erano contenuti più frustoli successivi di cui solo uno contenente annotazioni, si è passato ad annotare manualmente i frustoli annotati dal patologo, adottando un codice colore diverso rispetto a quello utilizzato per l'annotazione dell'area tumorale.

Il codice colore utilizzato è il seguente:

- **nero** : regioni contenenti tessuto tumorale;
- **rosso** : regioni di tessuto sano all'interno delle zone tumorali;
- **giallo** : frustolo annotato.



(a) WSI annotata dal patologo



(b) WSI con annotazioni rifinite e annotazione del frustolo

Figura 2.2: Rifinitura annotazioni manuali dell'immagine P18_M appartenente al gruppo 2.

2.3.2 Creazione maschere

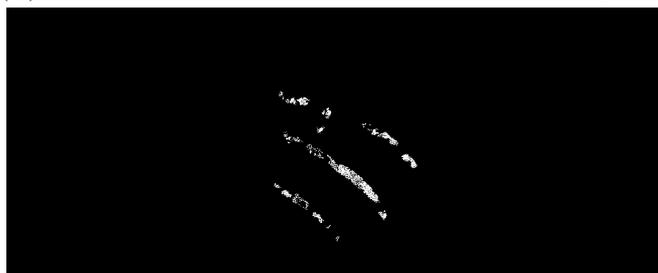
Le annotazioni rifinite al punto precedente sono state utilizzate per ricavare delle maschere per ogni tipo di annotazione, come è mostrato nella Figura 2.3.



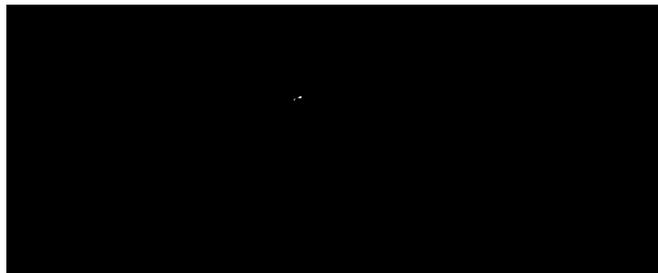
(a) WSI senza annotazioni



(b) maschera relativa alle annotazioni in giallo



(c) maschera relativa alle annotazioni in nero



(d) maschera relativa alle annotazioni in rosso

Figura 2.3: Maschere ottenute dalle annotazioni manuali dell'immagine P18_M appartenente al gruppo 2: b) frustolo annotato c) tessuto tumorale d) tessuto sano all'interno di tessuto tumorale.

2.3.3 Divisione dataset

Per proseguire con l'estrazione delle patch e l'allenamento della rete, le WSI del dataset sono state divise in maniera casuale in training set, validation set e test set, seguendo le seguenti proporzioni:

- **training set:** 75 % del dataset, 75 WSI per ogni Gleason Group;
- **validation set:** 15 % del dataset, 15 WSI per ogni Gleason Group;
- **test set:** 10 % del dataset, 10 WSI per ogni Gleason Group.

In questo modo ci si è assicurati che le patch usate per l'allenamento della rete provenissero da WSI diverse rispetto a quelle utilizzate per il testing della rete.

2.3.4 Patch extraction

Per l'estrazione delle patch da adoperare per l'allenamento della CNN è stato utilizzato un algoritmo che estrae un numero massimo di patch sane e tumorali da ogni WSI, rispettando il criterio di sovrapposizione con le rispettive maschere. Una patch è ritenuta valida se contiene almeno una certa percentuale del tessuto interessato: nello specifico, una patch sana è ritenuta valida se contiene almeno il 50 % di tessuto sano e lo 0 % di tessuto tumorale; una patch tumorale è ritenuta valida se contiene almeno il 70 % di tessuto tumorale. Il numero massimo di patch estratte per il tessuto sano è di 20, mentre per il tessuto tumorale è di 30.

La dimensione scelta per le patch è di 224x224, poiché scegliendo delle patch di dimensioni maggiori (512x512), in 30 WSI non si riusciva ad ottenere alcuna patch tumorale, a causa del fatto che le dimensioni del tumore sono significativamente minori della dimensione della patch adoperata. Diminuendo le dimensioni delle patch questo problema è comunque presente, ma è più limitato, essendo presente solo in 10 WSI.

Estraendo le patch in questo modo, si ottengono un totale di 10002 patch di tessuto sano e 12616 patch di tessuto tumorale, distribuite in training set, validation set e test set come mostrato in tabella

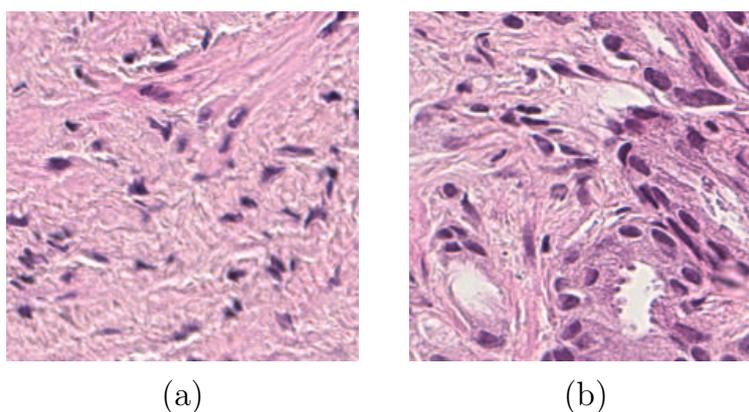


Figura 2.4: Esempi patch estratte dalla stessa WSI (66-18-B1 del gruppo 3): (a) patch sana, (b) patch tumorale.

2.2, che verranno successivamente usate come input della rete di Deep Learning.

	healthy	tumor	totale
training set	7501	9464	16965
validation set	1504	1921	3425
test set	997	1231	2228
totale	10002	12616	22618

Tabella 2.2: Distribuzione delle patch sane e tumorali estratte all'interno di training set, validation set e test set.

2.4 Allenamento CNN

L'algoritmo per l'allenamento della rete neurale convoluzionale è stato sviluppato in linguaggio Python usando PyTorch, una libreria open source utilizzata in applicazioni di *computer vision*.

In particolare le tipologie di architetture impiegate sono le seguenti:

1. Alexnet [32];
2. DenseNet121 [33].

Queste reti neurali sono state adoperate per effettuare una classificazione delle patch di tipo binario, con le seguenti classi:

- *classe 0*: tessuto sano;
- *classe 1*: tessuto tumorale.

2.4.1 Allenamento Alexnet

In un primo momento la rete neurale scelta per l'allenamento è stata Alexnet [32], che è una CNN composta da 5 layer convoluzionali e 3 fully connected layer, come si può vedere in Figura 2.5.

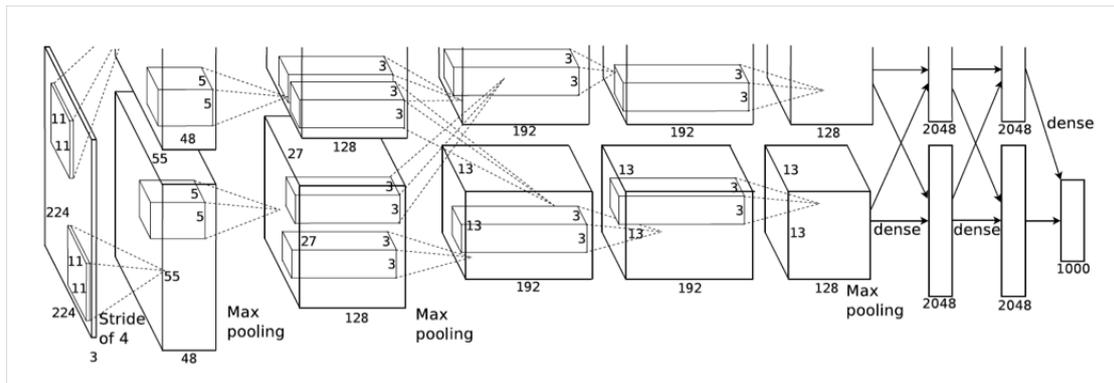


Figura 2.5: Architetture di Alexnet. Immagine tratta da [32].

La rete è stata allenata per effettuare la classificazione binaria delle patch, dunque è stata modificata la softmax finale in modo da avere solo 2 neuroni e non 1000 come nell'architettura originale dell'Alexnet.

La rete è stata allenata per 150 epoche utilizzando un batch di dimensione 32 e un learning rate iniziale di 0.001, che dopo ogni 30 epoche diminuisce di un fattore 0.1. La rete è stata ottimizzata tramite il metodo Adam. Inoltre sono stati allenati tutti i parametri della rete e non solo quelli adibiti alla classificazione, per permettere di effettuare la feature extraction in maniera ottimale.

2.4.2 Allenamento DenseNet

L'altra rete neurale che è stata adoperata per la classificazione delle patch è stata una DenseNet [33], nello specifico una DenseNet121. A differenza delle CNN tradizionali, in cui i layer sono collegati solo con quelli immediatamente successivi, ogni layer della DenseNet è collegato con tutti gli altri layer, come si può osservare nella figura 2.6.

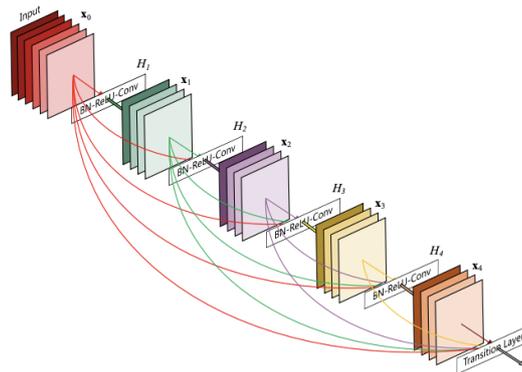


Figura 2.6: Architettura DenseNet. Si nota come ogni layer della DenseNet sia collegato con tutti gli altri layer e non solo con quelli direttamente successivi. Immagine tratta da [33].

Come per Alexnet, la rete è stata allenata per 150 epoche usando un batch di dimensione 32 e un learning rate iniziale di 0.001, che viene diminuito di un fattore 0.1 ogni 30 epoche. Anche in questo caso l'ottimizzazione è stata effettuata tramite il metodo Adam.

2.5 Heatmap

Per le immagini del test set sono state create delle heatmap utilizzando una *sliding window* delle stesse dimensioni delle patch con cui si è allenata la rete, ovvero 224x224.

Innanzitutto, per eliminare le discontinuità lungo i bordi, è stato effettuato un padding dell'immagine, aggiungendo 84 zeri lungo tutti i bordi dell'immagine. In seguito tutta l'immagine è stata scansionata tramite una finestra di dimensioni 224x224 che è stata fatta scorrere con un passo di 56. Le porzioni di immagine così ottenute sono state date in input alla rete precedentemente allenata, ottenendo la softmax che è stata salvata in una maschera apposita in una zona di dimensioni 56x56, a cui sarà successivamente applicata una scala colore.

In questo modo vengono evidenziate nella heatmap le zone che hanno maggiore probabilità di contenere tessuto tumorale secondo il sistema.

2.6 Metriche di validazione

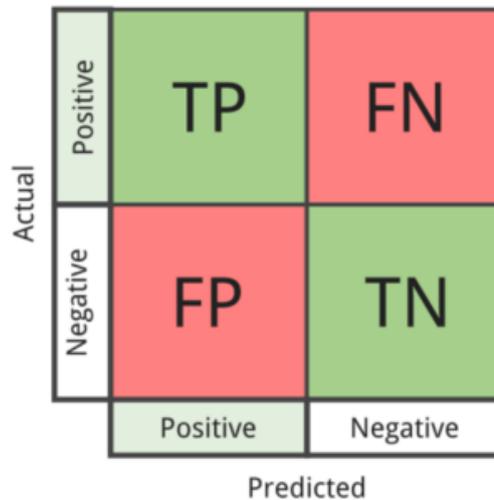
Di seguito vengono descritte le varie metriche che sono state utilizzate per valutare i risultati e le performance del sistema.

2.6.1 Confusion matrix

La confusion matrix è uno strumento molto utile per comprendere la correttezza di un sistema di classificazione. Può essere applicata sia a problemi di tipo binario, in cui le classi possibili sono 2, generalmente positivo o negativo, oppure anche nel caso di classificazione multi-classe. Ogni riga della matrice rappresenta il valore effettivo di una data classe, e ogni riga rappresenta il valore predetto dal sistema preso in esame.

Nel caso di confusion matrix binaria, un elemento può essere classificato in 4 modi diversi, come illustrato nella Figura 2.7:

- **Vero positivo (TP)**: la classe reale è positiva e anche la classe predetta è positiva;



Actual	Positive	TP	FN
	Negative	FP	TN
		Positive	Negative
		Predicted	

Figura 2.7: Esempio di confusion matrix per classificazione binaria. Immagine tratta da [34].

- **False negativo (FN)**: la classe reale è positiva ma la classe predetta è negativa: si tratta di un errore del primo tipo;
- **Falso positivo (FP)**: la classe reale è negativa ma la classe predetta è positiva: si tratta di un errore del secondo tipo;
- **Vero negativo (TN)**: la classe reale è negativa e anche la classe predetta è negativa;

A partire da questi valori ottenuti dalla confusion matrix, si possono ottenere altre metriche per valutare la performance dell'algoritmo.

Al fine di mantenere il significato delle metriche che si stanno per illustrare, nelle valutazioni del sistema contenute nel capitolo 3, sono stati considerati veri positivi gli elementi tumorali che sono stati classificati come tali.

Sensibilità o recall

La sensibilità o recall è una metrica che misura il tasso di veri positivi, ovvero quanti tra gli elementi positivi sono stati predetti come tali. Un

valore alto di sensibilità indica un numero basso di falsi positivi.

$$\text{Sensibilità} = \frac{TP}{TP + FN}$$

Specificità

La specificità misura il tasso di veri negativi, ovvero indica quanti tra gli elementi predetti come negativi lo sono effettivamente. Un valore elevato di specificità indica un numero basso di falsi negativi.

$$\text{Specificità} = \frac{TN}{TN + FP}$$

Precision

La precision, detta anche valore predittivo positivo, indica quanti tra gli elementi che hanno ricevuto una predizione di positività sono effettivamente positivi.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Accuratezza

Indica il valore di elementi correttamente classificati dal sistema.

$$\text{Accuratezza} = \frac{TP + TN}{TP + TN + FP + FN}$$

F₁ score o coefficiente di similarità di Dice

Si tratta di una media armonica di precision e recall. É anche detto coefficiente di Sørensen–Dice, e serve per valutare la similarità tra gli elementi.

$$F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$$

2.6.2 Curva ROC e AUC

La Receiver Operating Characteristic (ROC) è una rappresentazione grafica di un classificatore. Sull'asse delle ascisse si trova il tasso di falsi positivi, ovvero $1 - \text{specificità}$, e sull'asse delle ordinate la sensibilità. Tramite questo grafico si intuisce facilmente qual è il rapporto tra il tasso di veri positivi e i falsi allarmi di un classificatore. Un altro

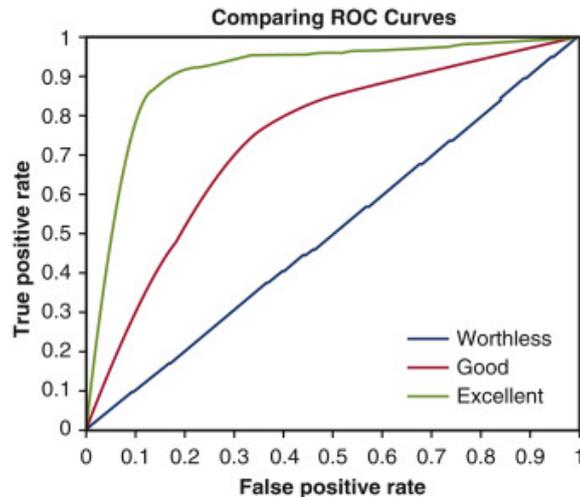


Figura 2.8: Esempio di curva ROC. Immagine tratta da [35].

parametro importante che si può ricavare dalla curva ROC è l'area sottesa dalla curva (AUC, Area Under Curve), che indica la capacità di un classificatore di distinguere adeguatamente i positivi dai negativi:

- $AUC < 0.5$: correlazione inversa tra predizione e valore reale;
- $AUC = 0.5$: classificatore casuale, quindi di fatto non classifica;
- $AUC > 0.5$: classificatore accettabile.

Maggiore è il valore della AUC, migliore sarà il classificatore, perché vorrà dire che riuscirà ad identificare un numero elevato di veri positivi causando un numero molto ristretto di falsi allarmi.

Capitolo 3

Risultati

3.1 Risultati Alexnet

3.1.1 Risultati allenamento

I risultati ottenuti durante l'allenamento e la validazione dell'Alexnet sono rappresentati graficamente in Figura 3.1, in cui sono raffigurati l'andamento della *loss function* e dell'accuratezza, sia per la fase di training che per quella di validation.

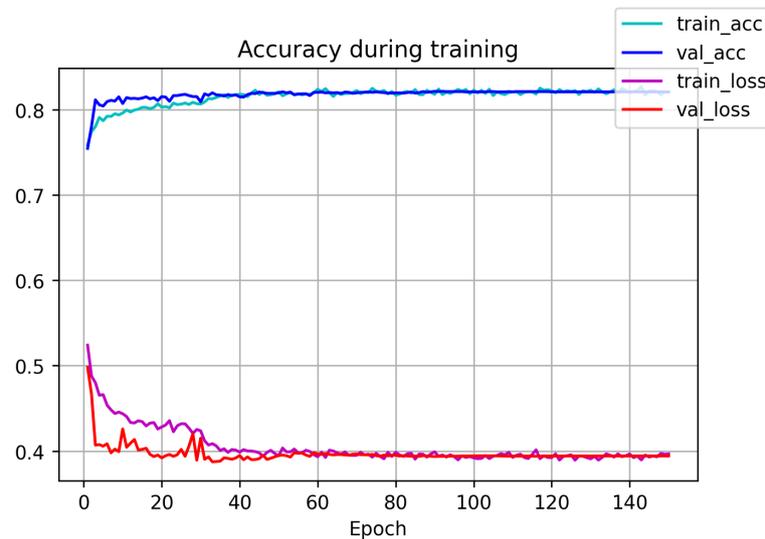


Figura 3.1: Learning curve ottenuta dall'allenamento di Alexnet.

L'accuratezza massima sul validation set è di circa 82 % e viene raggiunta intorno all'epoca 60, dopo la quale non vi è alcun ulteriore miglioramento.

Le confusion matrix che si ottengono sul training set e sul validation set sono illustrate in figura 3.2, dove la classe 0 corrisponde al tessuto sano e la classe 1 corrisponde al tessuto tumorale.

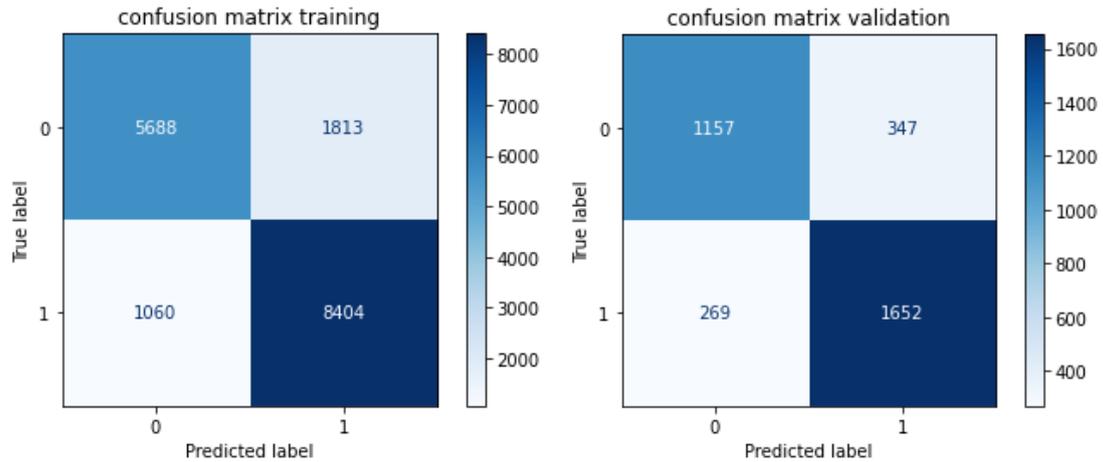


Figura 3.2: Confusion matrix ottenute da Alexnet per training set (a sinistra) e validation set (a destra).

Dalle confusion matrix vengono in seguito estrapolate tutte le metriche utili alla valutazione delle performance del sistema, che sono illustrate nella tabella 3.1.

	<i>Alexnet</i>		
	train	val	test
accuratezza	83,19%	82,01%	82,32%
sensibilità	89,06%	86,00%	86,92%
specificità	75,79%	76,93%	76,63%
precision	82,27%	82,64%	82,12%
F-score	85,53%	84,29%	84,45%

Tabella 3.1: Tabella riassuntiva risultati ottenuti per la rete Alexnet.

3.1.2 Risultati test

La confusion matrix ottenuta testando la rete sulle immagini appartenenti al test set è rappresentata nella figura 3.3, e le metriche estratte per la valutazione dei risultati sono stati inseriti nella tabella 3.1.

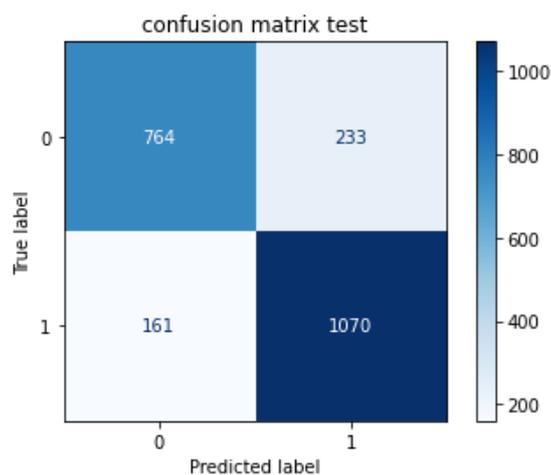


Figura 3.3: Confusion matrix Alexnet per il test set.

Inoltre è stata ricavata anche la curva ROC del sistema, che presenta una AUC di 0.82, come si può notare dalla figura 3.4.

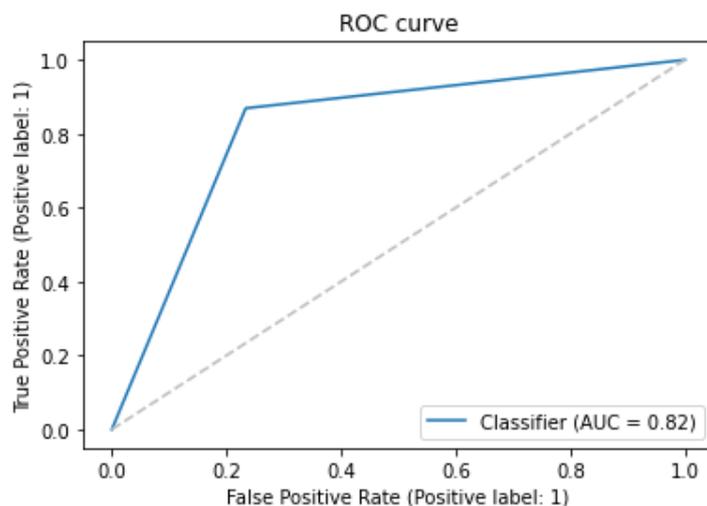


Figura 3.4: Curve ROC del sistema con Alexnet ottenuta sul test set.

3.2 Risultati DenseNet

3.2.1 Risultati allenamento

I risultati ottenuti dall'allenamento di questo tipo di rete sono in linea con i risultati ottenuti precedentemente con Alexnet: anche in questo caso le performance del sistema giungono ad un punto di saturazione oltre il quale non vi è ulteriore miglioramento. Le confusion matrix ottenute per il training set e il validation set sono mostrate in Figura 3.5, e le metriche da esse ottenute sono sintetizzate nella tabella 3.2.

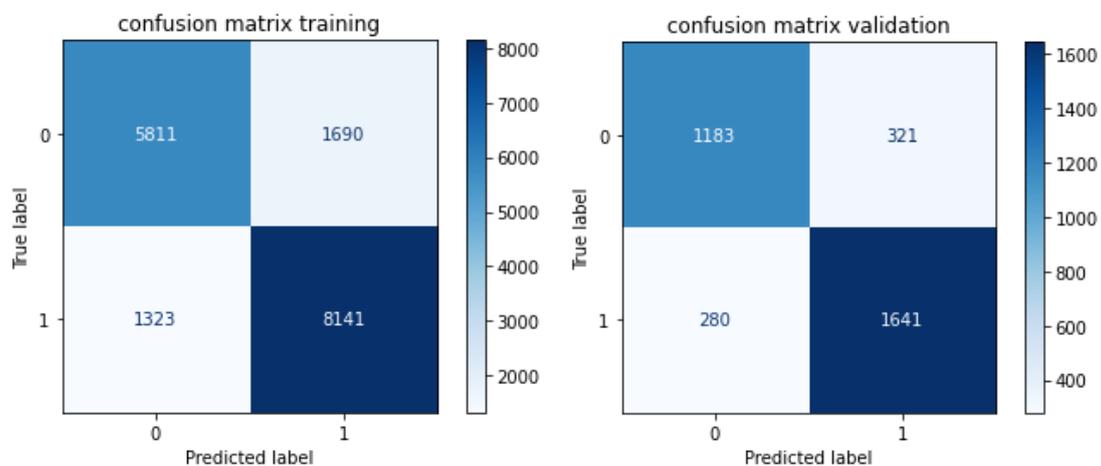


Figura 3.5: Confusion matrix DenseNet per training set (a sinistra) e validation set (a destra).

<i>DenseNet</i>			
	train	val	test
accuratezza	82,24%	82,45%	82,72%
sensibilità	86,02%	85,42%	86,52%
specificità	77,47%	78,66%	78,03%
precision	82,81%	83,64%	82,94%
F-score	84,38%	84,52%	84,69%

Tabella 3.2: Tabella risuntiva dei risultati ottenuti per la DenseNet.

3.2.2 Risultati test

La confusion matrix ottenuta sul test set è rappresentata in figura 3.6. La curva ROC ottenuta per il modello è rappresentata in Figura 3.7.

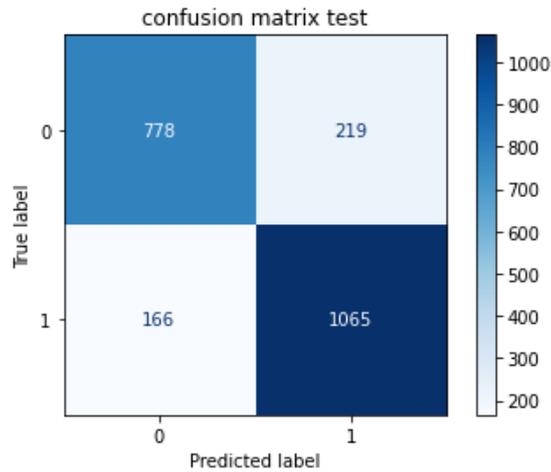


Figura 3.6: Confusion matrix ottenuta per il test set con DenseNet.

Come si può notare la AUC ottenuta è la medesima di quella ottenuta con Alexnet.

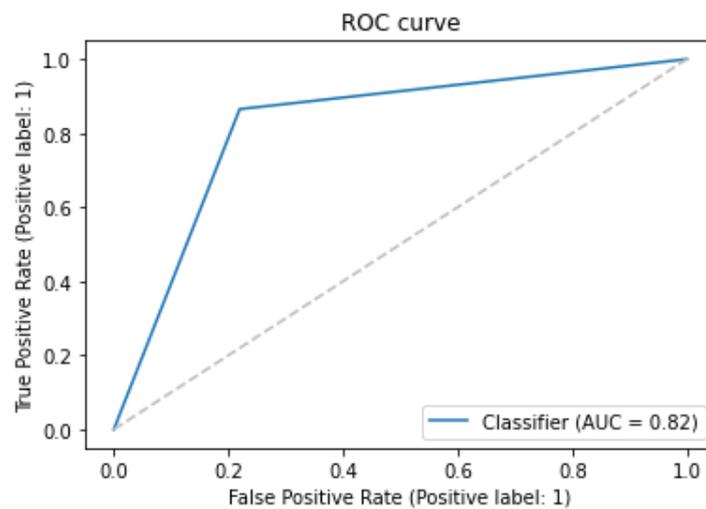


Figura 3.7: Curva ROC ottenuta per la DenseNet.

3.3 Risultati heatmap

Il modello migliore ottenuto è stato utilizzato per creare le heatmap relative immagini di test, tramite il metodo illustrato nel capitolo 2. In figura 3.8 è rappresentata un esempio di heatmap ottenuta su una porzione di WSI appartenente al test set.

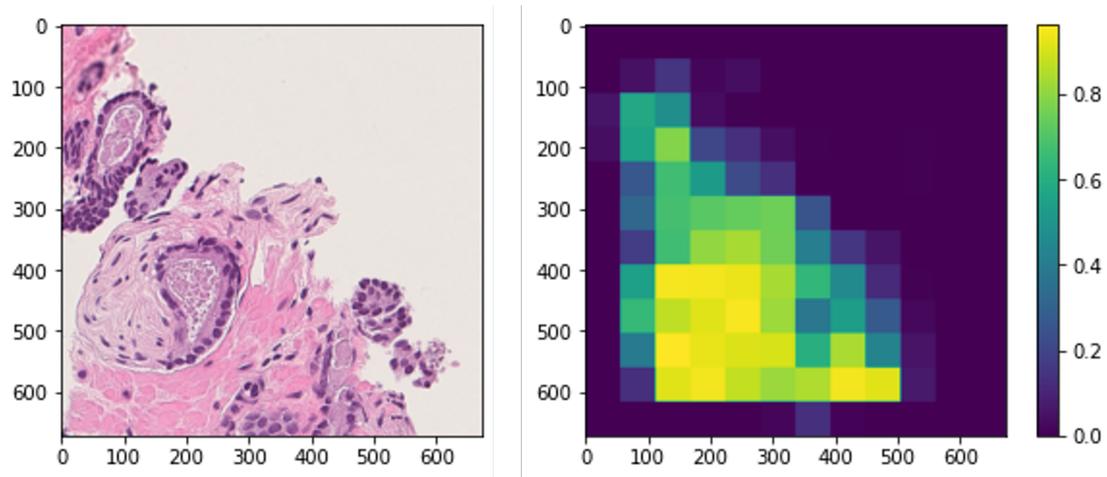


Figura 3.8: A sinistra, un campo della WSI, e a destra la heatmap corrispondente.

Capitolo 4

Conclusioni e sviluppi futuri

4.1 Conclusioni

In questo lavoro di tesi è stato sviluppato un sistema in grado di classificare e identificare il tessuto tumorale all'interno di immagini istologiche, in grado di distinguere abbastanza accuratamente porzioni di tessuto sano dal tessuto tumorale, raggiungendo un'accuratezza dell'82 % sul test set. Per questo motivo tale sistema potrebbe essere impiegato come sistema di supporto alla decisione dei patologi durante l'annotazione delle WSI, in maniera da fornire una seconda opinione e ridurre il carico di lavoro di questi operatori.

Indubbiamente l'utilizzo di metodi di Deep Learning è fondamentale per raggiungere tali risultati, che tuttavia sono molto influenzati dalla qualità dei dati forniti come input alla rete neurale convoluzionale. Perciò è auspicabile che tutte le immagini ricevano un pre-processing molto accurato prima di effettuare l'allenamento della CNN.

4.2 Sviluppi futuri

Per migliorare ulteriormente le performance del sistema si potrebbero implementare metodi di *stain normalization* per minimizzare la variazione di colore tra le varie WSI che compongono il dataset, in maniera da rendere il sistema più robusto di fronte a questa fonte di variabilità.

Inoltre, la scelta dell'architettura più appropriata per la CNN potrebbe contribuire all'ottenimento di un sistema migliore, in modo che riesca ad estrarre features più ad alto livello.

Si potrebbe anche implementare un classificatore multiclasse, in grado di distinguere i vari Gleason Grade Group, e quindi sviluppare un sistema che effettui anche il grading, oltre che l'individuazione del tessuto tumorale.

Ringraziamenti

Innanzitutto vorrei ringraziare il prof. Filippo Molinari per avermi dato l'opportunità di svolgere questo lavoro di tesi, che ha permesso di ampliare le mie conoscenze in un campo che mi interessa molto. Un importante ringraziamento va all'Ing. Massimo Salvi per la sua grande disponibilità durante questi mesi, per i consigli e per il supporto durante tutto lo svolgimento del lavoro.

Ringrazio tutti coloro che mi hanno sostenuto in questo percorso, primi fra tutti la mia famiglia che mi è sempre stata accanto in ogni momento.

Ringrazio i miei amici, quelli di lunga data con cui ho vissuto tante *epiche* avventure e che mi sono accanto nonostante la distanza. Un grazie anche a quelli che ho incontrato durante il mio percorso universitario, agli amici dell'aula studio che hanno assistito e partecipato a questo mio percorso, aiutandomi e supportandomi in ogni modo possibile.

Un grazie immenso va alle mie amiche e colleghe Margherita, Federica, Giulia R, Alessandra, Giusy, Giulia M, Laura e Michela, che mi hanno accompagnato in questo viaggio fin dall'inizio, aiutandomi e incoraggiandomi sempre, anche nelle situazioni più difficili.

Un grazie speciale va anche a Federico, che mi ha sempre aiutato, supportato e sopportato, soprattutto nell'ultimo periodo.

Infine ringrazio Paola, per essermi stata vicina in tutti questi anni, prima come amica e poi praticamente come sorella, per aver sempre creduto in me e per esserci sempre, nonostante tutto.

Bibliografia

- [1] Christine H Lee, Oluyemi Akin-Olugbade, and Alexander Kirshenbaum. Overview of prostate anatomy, histology, and pathology. *Endocrinology and metabolism clinics of North America*, 40(3):565–75, 2011.
- [2] M-C Franz, Pascale Anderle, Marc Bürzle, Y Suzuki, MR Freeman, MA Hediger, and Gergely Kovacs. Zinc transporters in prostate cancer. *Molecular aspects of medicine*, 34(2-3):735–741, 2013.
- [3] C.L. Stanfield. *Fisiologia*. Edises, 2012.
- [4] David C. Miller, Khaled S. Hafez, Andrew Stewart, James E. Montie, and John T. Wei. Prostate carcinoma presentation, diagnosis, and staging. *Cancer*, 98(6):1169–1178, 2003.
- [5] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.
- [6] Claire H Pernar, Ericka M Ebot, Kathryn M Wilson, and Lorelei A Mucci. The epidemiology of prostate cancer. *Cold Spring Harbor perspectives in medicine*, 8(12):a030361, 2018.
- [7] Jerome P Richie, William J Catalona, Frederick R Ahmann, A Hudson M’Liss, Peter T Scardino, Robert C Flanigan, Jean B

- Dekernion, Timothy L Ratliff, Louis R Kavoussi, Bruce L Dal-kin, et al. Effect of patient age on early detection of prostate cancer with serum prostate-specific antigen and digital rectal examination. *Urology*, 42(4):365–374, 1993.
- [8] AIOM. Linee guida carcinoma della prostata, 2020.
- [9] Pieter JL De Visschere, Alberto Briganti, Jurgen J Fütterer, Pirus Ghadjar, Hendrik Isbarn, Christophe Massard, Piet Ost, Prasanna Sooriakumaran, Cristian I Surcel, Massimo Valerio, et al. Role of multiparametric magnetic resonance imaging in early detection of prostate cancer. *Insights into imaging*, 7(2):205–214, 2016.
- [10] Sascha Kaufmann, Giorgio I Russo, Fabian Bamberg, Lorenz Löwe, Giuseppe Morgia, Konstantin Nikolaou, Arnulf Stenzl, Stephan Kruck, and Jens Bedke. Prostate cancer detection in patients with prior negative biopsy undergoing cognitive-, robotic-or in-bore mri target biopsy. *World journal of urology*, 36(5):761–768, 2018.
- [11] Veeru Kasivisvanathan, Antti S Rannikko, Marcelo Borghi, Valeria Panebianco, Lance A Mynderse, Markku H Vaarala, Alberto Briganti, Lars Budäus, Giles HELLAWELL, Richard G Hindley, et al. Mri-targeted or standard biopsy for prostate-cancer diagnosis. *New England Journal of Medicine*, 378(19):1767–1777, 2018.
- [12] Marloes van der Leest, Erik Cornel, Bas Israël, Rianne Hendriks, Anwar R Padhani, Martijn Hoogenboom, Patrik Zamecnik, Dirk Bakker, Anglita Yanti Setiasti, Jeroen Veltman, et al. Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naive men with elevated prostate-specific antigen: a large prospective multicenter clinical study. *European urology*, 75(4):570–578, 2019.
- [13] Gladell Paner, Jatin Gandhi, Bonnie Choy, and Mahul Amin. Essential updates in grading, morphotyping, reporting, and staging

- of prostate carcinoma for general surgical pathologists. *Archives of Pathology & Laboratory Medicine*, 143, 03 2019.
- [14] Jonathan I Epstein, William C Allsbrook Jr, Mahul B Amin, Lars L Egevad, ISUP Grading Committee, et al. The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 29(9):1228–1242, 2005.
- [15] Jonathan I Epstein, Lars Egevad, Mahul B Amin, Brett Delahunt, John R Srigley, and Peter A Humphrey. The 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 40(2):244–252, 2016.
- [16] Annika Herlemann, Janet E Cowan, Peter R Carroll, and Matthew R Cooperberg. Community-based outcomes of open versus robot-assisted radical prostatectomy. *European urology*, 73(2):215–223, 2018.
- [17] Anna Bill-Axelsson, Lars Holmberg, Hans Garmo, Kimmo Taari, Christer Busch, Stig Nordling, Michael Häggman, Swen-Olof Andersson, Ove Andrén, Gunnar Steineck, et al. Radical prostatectomy or watchful waiting in prostate cancer—29-year follow-up. *New England Journal of Medicine*, 379(24):2319–2329, 2018.
- [18] Freddie C Hamdy, Jenny L Donovan, J Lane, Malcolm Mason, Chris Metcalfe, Peter Holding, Michael Davis, Tim J Peters, Emma L Turner, Richard M Martin, et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med*, 375:1415–1424, 2016.
- [19] Massimo Salvi, U Rajendra Acharya, Filippo Molinari, and Kristen M Meiburger. The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine*, page 104129, 2020.

- [20] Colin Doolan, Olga Colgan, and Sherri Heffner. What is digital pathology, Jan 2019.
- [21] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7, 2016.
- [22] Vito Lavecchia. Caratteristiche e funzionamento del deep learning in informatica, Feb 2021.
- [23] Sumit Saha. A comprehensive guide to convolutional neural networks - the eli5 way, Dec 2018.
- [24] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [25] Hanna Källén, Jesper Molin, Anders Heyden, Claes Lundström, and Kalle Åström. Towards grading gleason score using generically trained deep convolutional neural networks. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1163–1167. IEEE, 2016.
- [26] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, 6(1):1–11, 2016.
- [27] Naiyun Zhou and Yi Gao. Optimized color decomposition of localized whole slide images and convolutional neural network for intermediate prostate cancer classification. In *Medical Imaging 2017: Digital Pathology*, volume 10140, page 101400W. International Society for Optics and Photonics, 2017.
- [28] Eirini Arvaniti, Kim S Fricker, Michael Moret, Niels Rupp, Thomas Hermanns, Christian Fankhauser, Norbert Wey, Peter J Wild, Jan H Rueschoff, and Manfred Claassen. Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific reports*, 8(1):1–11, 2018.

- [29] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Mirafior, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.
- [30] Kunal Nagpal, Davis Foote, Yun Liu, Po-Hsuan Cameron Chen, Ellery Wulczyn, Fraser Tan, Niels Olson, Jenny L Smith, Arash Mohtashamian, James H Wren, et al. Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer. *NPJ digital medicine*, 2(1):1–10, 2019.
- [31] Massimo Salvi, Filippo Molinari, U Rajendra Acharya, Luca Molinaro, and Kristen M Meiburger. Impact of stain normalization and patch selection on the performance of convolutional neural networks in histological breast and prostate cancer classification. *Computer Methods and Programs in Biomedicine Update*, 1:100004, 2021.
- [32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [33] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [34] MLee. Visual guide to the confusion matrix, Sep 2021.
- [35] Victor A Ferraris. Commentary: Should we rely on receiver operating characteristic curves? from submarines to medical tests, the answer is a definite maybe! *The Journal of thoracic and cardiovascular surgery*, 157(6):2354–2355, 2019.