



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

A.a. 2023/2024

Sessione di Laurea dicembre 2024

Sviluppo di un algoritmo di AI per la detection e la caratterizzazione di tumori prostatici

Relatori:

Rosati Samanta
Giannini Valentina

Candidato:

Spinola Francesco

Indice

Abstract.....	1
Capitolo 1.....	4
1.1 Anatomia della prostata.....	4
1.2 Cancro alla prostata.....	6
1.2.1 Epidemiologia del tumore alla prostata.....	6
1.2.2 Diagnosi del tumore.....	8
1.2.3 Trattamento del tumore.....	11
1.3 MRI per il cancro alla prostata.....	13
Capitolo 2.....	15
2.1 Intelligenza artificiale.....	15
2.2 Il neurone.....	17
2.2.1 Le funzioni di trasferimento.....	19
2.3 Architetture delle reti neurali.....	25
2.3.1 Le Feed Forward Network (FNN).....	25
2.3.2 Convolutional Neural Network (CNN).....	27
2.3.3 Recurrent neural networks (RNN).....	30
2.3.4 Generative adversarial networks (GAN).....	30
2.4 AI nel medical imaging.....	32

Capitolo 3	35
3.1 Dataset	35
3.2 Pre-processing	36
3.3 Costruzione del classificatore	37
3.4 Ottimizzazioni	38
3.5 Metriche di valutazione	40
3.5.1 Confusion Matrix.....	40
3.5.2 Curva ROC	42
Risultati	44
Conclusioni	53
Bibliografia	54
Ringraziamenti	60

Abstract

Lo scopo di questa tesi è quello di sviluppare un algoritmo deep learning che aiuti il radiologo nella classificazione dell'aggressività di un tumore alla prostata. Ad oggi il tumore viene classificato tramite una biopsia prelevando campioni di tessuto canceroso e determinandone l'aggressività mediante l'utilizzo del cosiddetto Gleason Grade (GG). Il progetto prevede di sviluppare un algoritmo deep learning basato su immagini di Risonanza Magnetica che sfrutta le reti neurali convoluzionali su un set di immagini di 142 pazienti. Per ogni paziente sono state acquisite le immagini tramite MRI con 3 diverse tecniche: T2, ADC e DWI.

L'algoritmo è stato realizzato per distinguere il set di pazienti in due classi $GG = 2$ (classe 0) e $GG > 2$ (classe 1) per distinguere i tumori clinicamente significativi, i quali necessitano di più attenzioni da quelli meno rischiosi per la salute del paziente. Le immagini mediche sono state prelevate con macchinari diversi; per questo motivo sono state pre-processate uniformandole e normalizzandole per rendere l'apprendimento della rete più efficiente. Il set è stato diviso in due gruppi, uno di training formato da 120 pazienti e uno di test che contiene il resto di 22 pazienti; per ogni rete sono state eseguite 5 prove utilizzando la procedura del *cross_validation*. Essa consiste nel dividere il training set in 5 parti e per ogni addestramento sono state prese 4 parti per il training set e una parte per il validation set, per ogni prova eseguita il validation set cambiava. Tale tecnica è stata impiegata per garantire l'affidabilità del modello. Inoltre, a causa del basso numero di elementi del set è stata applicata anche la tecnica di *data augmentation*: il training set già suddiviso con il validation è stato quadruplicato applicando leggere modifiche alle immagini iniziali, questo per rendere il modello più robusto e ridurre il problema dell'overfitting.

Tra le varianti testate, la rete che ha ottenuto i migliori risultati sul test set ha dimostrato un'accuratezza del 68,18%, una sensibilità dell'81,82% una specificità del 54,55%. Questi risultati evidenziano una preferenza per l'identificazione dei tumori più gravi, riducendo al minimo i falsi negativi. I risultati ottenuti dimostrano che l'utilizzo di reti neurali convoluzionali applicate a immagini di risonanza magnetica rappresenta un approccio promettente per supportare i radiologi nella classificazione dell'aggressività

del tumore alla prostata. Nonostante la complessità intrinseca dei dati e il numero limitato di pazienti nel dataset, i risultati ottenuti suggeriscono che il modello sviluppato è particolarmente efficace nell'identificazione dei tumori clinicamente significativi, con un buon compromesso tra sensibilità e specificità.

Capitolo 1

1.1 Anatomia della prostata

La ghiandola prostatica è un organo a forma di piramide del sistema riproduttivo maschile, situata nel bacino; essa si trova sotto la vescica urinaria e davanti al retto. La prostata di un adulto sano pesa circa 15–20 g e misura circa 3-4 cm di lunghezza e la sua funzione principale è quella di produrre un fluido che insieme agli spermatozoi e ai fluidi degli altri organi del sistema riproduttivo, forma il liquido seminale.

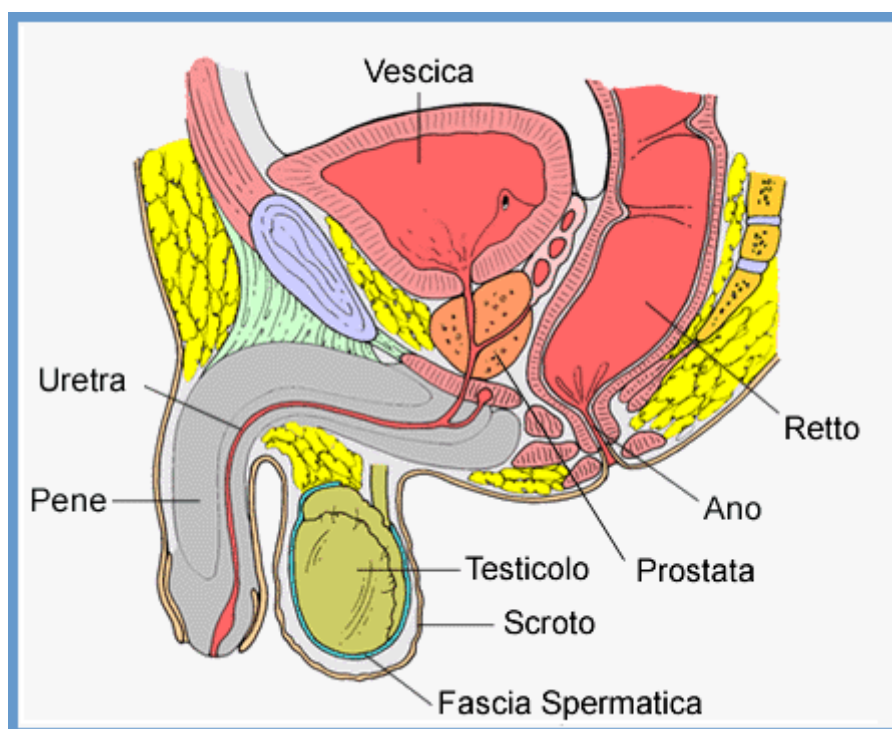


Figura 1.1: Rappresentazione della prostata e degli organi circostanti.

La struttura interna della prostata può essere descritta in termini di lobi e zone, ciascuna con caratteristiche anatomiche e funzionali distinte:

- Nell'anatomia macroscopica, la prostata è classificata in cinque lobi: il lobo anteriore (situato anteriormente), il lobo posteriore (posteriormente), due lobi laterali (sui lati) e un lobo mediano (al centro) ^[1].
- L'anatomia della prostata è anche descritta secondo le sue aree anatomiche, che hanno un significato clinico: zone periferica, di transizione e centrale ^[2].
 - La **zona periferica** si trova nella parte posteriore e inferiore della prostata, avvolge parte dell'uretra e contiene la zona centrale della ghiandola. È l'area più vicina al retto ed è la zona più grande della ghiandola, costituendo circa il 70% del volume prostatico negli uomini adulti. Questa area è il principale sito di prostatite e carcinoma della prostata (PCa) ^[3];
 - La **zona centrale** anteriormente si rapporta con l'uretra prostatica, contiene il dotto eiaculatorio e rappresenta circa il 20% della prostata ^[4];
 - La **zona di transizione** si trova vicino all'uretra prostatica rappresenta circa il 5% della prostata, per questo motivo nei giovani è poco visibile ^[2], essa costituisce il sito in cui si verifica l'iperplasia prostatica benigna (BPH) ^[3,4].

Inoltre, la prostata è circondata da uno stroma fibromuscolare che gioca un ruolo fondamentale nel mantenere la struttura e la funzione della ghiandola ^[4].

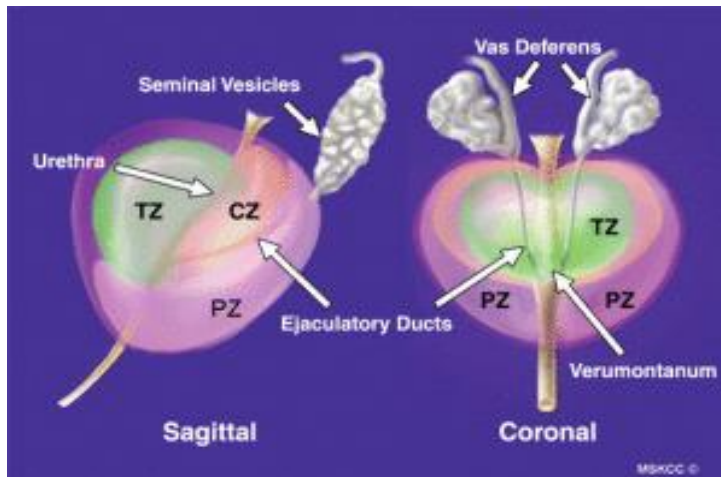


Figura 1.2: L'immagine mostra una rappresentazione anatomica della prostata in due sezioni, sagittale e coronale, si vedono le principali aree anatomiche della prostata: la zona centrale (CZ), la zona transizionale (TZ) e la zona periferica (PZ).

1.2 Il Cancro alla prostata

1.2.1 Epidemiologia del tumore alla prostata

Il cancro alla prostata è il quarto tumore più comune a livello mondiale, con un totale di 1.467.854 nuovi casi diagnosticati nel 2022 e un tasso di incidenza standardizzato per età (ASR) di 29,4 per 100.000 persone. Ha anche un tasso di mortalità standardizzato per età (ASR) di 7,3 per 100.000 persone, rendendolo l'ottavo tumore più letale al mondo ^[5]. Esso può presentarsi senza sintomi nella fase iniziale della malattia e in alcuni casi anche durante tutto il decorso della stessa, richiedendo un trattamento lieve se non nullo. Ad ogni modo i sintomi maggiormente ricorrenti comprendono la difficoltà urinaria, l'aumento della frequenza e la nicturia ^[6].

I principali fattori di rischio per il cancro alla prostata possono essere suddivisi in fattori irreversibili e reversibili.

Tra i **fattori irreversibili**, il più importante è l'età: le probabilità di sviluppare la malattia sono basse prima dei 40 anni, ma aumentano significativamente dopo i 50 anni, e circa due casi su tre vengono diagnosticati in persone di età superiore ai 65 anni. In secondo luogo, per coloro che hanno un parente di sangue (padre, fratello,

ecc.) con diagnosi di cancro alla prostata, il rischio di sviluppare la malattia è il doppio rispetto agli individui senza casi familiari. Anche l'etnia influisce sulla probabilità di sviluppare un tumore alla prostata; infatti, si è notato che l'incidenza di un carcinoma prostatico è maggiore nella popolazione caraibica e in quella nera nordamericana. Inoltre, mutazioni in geni specifici come BRCA1 e BRCA2, associati anche al cancro al seno e alle ovaie, così come la sindrome di Lynch (cancro ereditario del colon-retto non poliposico, HNPCC), possono aumentare il rischio di sviluppare il cancro alla prostata [6,7].

I **fattori reversibili** sono legati allo stile di vita e includono una dieta ricca di grassi saturi, obesità e mancanza di attività fisica [6,7], che sono purtroppo caratteristici nei paesi più sviluppati.

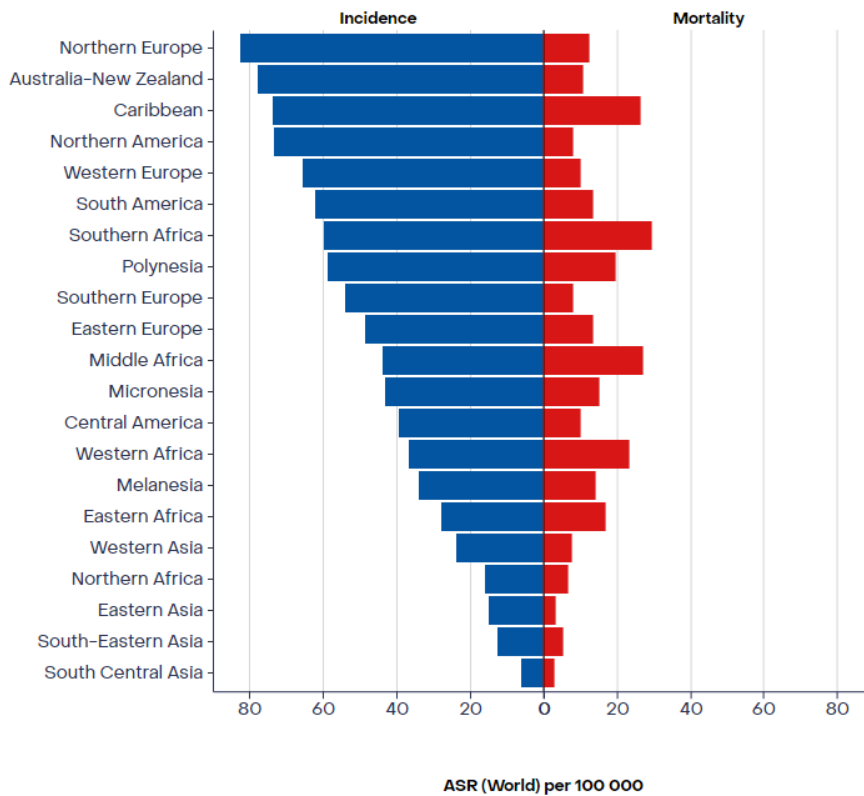


Figura 1.3: Age-Standardized Rate (World) per 100 000, incidenza e mortalità, entrambi i sessi, nel 2022 [5]

1.2.2 Diagnosi del tumore

La diagnosi del tumore alla prostata può essere effettuata con diversi esami.

Esplorazione rettale (DRE)

La più immediata viene eseguita mediante un'**esplorazione rettale** attraverso cui il medico evidenzia digitalmente se sono presenti o meno anomalie. È la principale tecnica solitamente esperita quando il paziente presenta una sintomatologia che possa essere riconducibile alla presenza di un tumore e permette di individuare mediante la palpazione eventuali malformazioni, consentendo una diagnosi precoce della malattia. Tuttavia, alcuni studi hanno dimostrato che essa non garantisce un'accuratezza adeguata all'individuazione del tumore, per tale ragione è importante che il paziente venga sottoposto anche ad ulteriori esami dotati di maggiore sensibilità e specificità [8,9].

PSA

Un altro esame è quello di monitorare l'antigene specifico della prostata che è il **PSA**. Esso è una glicoproteina prodotta dalla prostata che si può trovare sia nel liquido seminale sia nel sangue e in base ai valori del PSA si può stabilire se ci sono o meno problemi. Infatti, se i livelli di PSA sono inferiori a 4 ng/l significa che non sussiste alcuna anomalia, se i valori di PSA salgono tra i 4-10 ng/l le probabilità di avere un tumore alla prostata sono del 25%; invece se si supera il valore di 10ng/l le probabilità superano il 50%. Tuttavia, variazioni di PSA nel sangue non sempre indicano un tumore alla prostata ma possono derivare anche da altre patologie come BPH o prostatiti [10].

Biopsia

L'unico esame che è in grado di individuare con certezza la presenza di un tumore alla prostata è la **biopsia**, che viene effettuata prelevando piccoli campioni dalla ghiandola che poi verranno esaminati al microscopio [7,10].

La classificazione del tumore alla prostata viene fatto studiando i prelievi della biopsia in base al **modello di Gleason**, i cui criteri sono stati successivamente aggiornati dall'ISUP (International Society of Urological Pathology). Questa classificazione prevede un punteggio da 1 a 5 e permette di stabilire quanto le cellule tumorali differiscono dalle cellule normali dell'organo. In particolare, i gradi 1-2 individuano delle cellule tumorali nodulari separate e ben formate; il grado 3 invece presenta cellule tumorali molto più numerose; in questo grado e in quelli successivi possono essere presenti, altresì, infiltrazioni nello stroma. Nel grado 4 le cellule tumorali cominciano a fondersi tra loro formando dei pattern e possono assumere una struttura cribriforme o infiltrarsi nelle zone adiacenti. Infine, il grado 5 presenta delle cellule tumorali che formano delle masse ben definite di forma tonda o cilindrica con zone necrotiche al suo interno ^[11,12]. Una volta definito il Gleason Grade sui campioni si definisce il Gleason Score che è la somma del primo e del secondo punteggio più comune; invece, se è presente solo un pattern il valore viene duplicato. Il GS viene scritto come somma dei due numeri ed i tumori che presentano un **GS ≤ 6** sono poco aggressivi; pertanto, vi è una bassa probabilità di essere un cancro clinicamente significativo. I tumori che presentano un **GS = 7 (3+4 o 4+3)** cominciano ad essere clinicamente significativi e la pericolosità aumenta in base al grado del pattern più comune. Infine, i tumori che hanno un **GS ≥ 8** sono tumori che presentano maggior rischio per il paziente perché più aggressivi e maggiormente predisposti a formare metastasi ^[12,13].

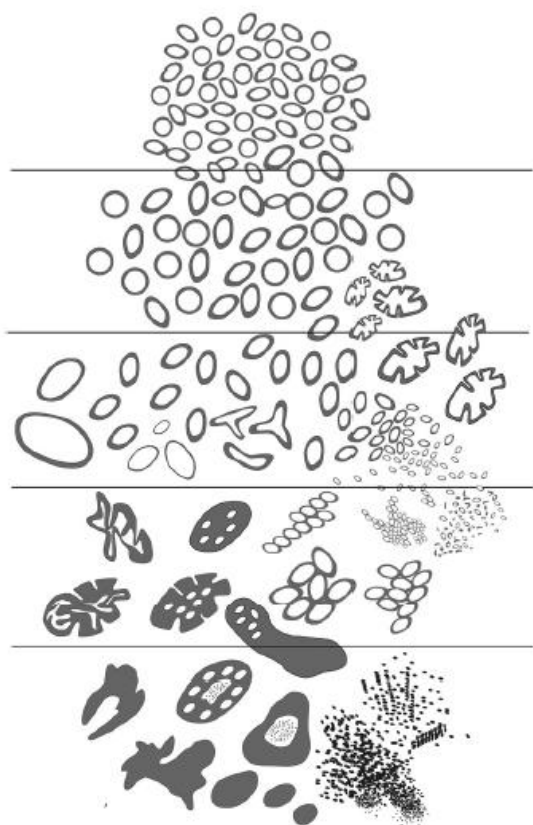


Figura 1.4: Gradi del cancro alla prostata. Diagramma schematico della classificazione modificata di Gleason del 2015 secondo la Società Internazionale di Patologia Urologica (ISUP). Schemi di grado dal 1 (in alto) al 5 (in basso) [11].

Nel 2014 l'ISUP ha aggiornato il sistema di valutazione poiché due tumori con stesso GS (3+4 o 4+3) contengono una sostanziale differenza nella pericolosità e nel trattamento; quindi, sono stati introdotti 5 gruppi in base al GS [12]:

- $GS \leq 6$ = Grade Group 1;
- $GS = 7 (3+4)$ = Grade Group 2;
- $GS = 7 (4+3)$ = Grade Group 3;
- $GS = 8$ = Grade Group 4;
- $GS > 8$ = Grade Group 5.

1.2.3 Trattamento del tumore

Il trattamento del tumore alla prostata si articola su più livelli in base allo stadio della malattia, al rischio di progressione e alla salute generale del paziente. Nei tumori localizzati a basso rischio, la sorveglianza attiva rimane la scelta più conveniente, specialmente per pazienti più anziani o con comorbidità. Questo approccio evita interventi invasivi, monitorando il cancro tramite esami regolari, inclusi i livelli di PSA, biopsie e risonanze magnetiche. Nei tumori localizzati ma a rischio più elevato, si preferisce intervenire con terapie definitive come la prostatectomia radicale o la radioterapia a fasci esterni ^[14,15].

La **brachiterapia** (radioterapia interna) è una tecnica che prevede l'impianto di piccoli semi radioattivi direttamente nella prostata e viene utilizzata principalmente per pazienti con malattia a basso o intermedio rischio ^[15]. La **prostatectomia radicale**, come per la brachiterapia, si utilizza in presenza di tumori localizzati e con un rischio per il paziente basso/intermedio. Essa consiste nell'asportazione completa della prostata e del tumore, riducendo la probabilità di sviluppare metastasi ma comporta una maggiore incidenza di effetti collaterali come incontinenza e disfunzione erettile, rispetto alla radioterapia ^[14].

Nei pazienti con malattia avanzata o metastatica, l'approccio terapeutico si focalizza sull'arresto della produzione di androgeni, ormoni che favoriscono la crescita del tumore. La **terapia di deprivazione androgenica (ADT)**, che può essere realizzata tramite farmaci che bloccano la produzione di testosterone o la sua azione sulle cellule tumorali oppure mediante castrazione chirurgica, consiste nella rimozione totale dei testicoli, che sono gli addetti principali alla produzione di testosterone. Tuttavia, questa tecnica è sempre meno utilizzata a causa degli effetti irreversibili. L'utilizzo della terapia ADT è spesso combinato con altre terapie per ottimizzare i risultati. Nei casi di tumore resistente alla castrazione, quando la malattia progredisce nonostante livelli ridotti di androgeni, si ricorre a nuove terapie ormonali come abiraterone ed enzalutamide, che hanno dimostrato di prolungare significativamente la sopravvivenza nei pazienti ^[16].

La chemioterapia gioca un ruolo importante nei casi avanzati. Il docetaxel, ad esempio, è stato il primo farmaco chemioterapico a dimostrare un beneficio in termini di sopravvivenza nei pazienti con cancro alla prostata resistente alla castrazione metastatico, ed è considerato uno standard di cura in combinazione con prednisone. Successivamente, altri agenti come cabazitaxel, studiato in pazienti resistenti al docetaxel, hanno fornito ulteriori opzioni terapeutiche per prolungare la vita dei pazienti [16].

Le immunoterapie, un campo emergente, offrono un nuovo approccio alla cura del cancro alla prostata avanzato. Sipuleucel-T, un'immunoterapia cellulare autologa, ha dimostrato di migliorare la sopravvivenza nei pazienti con mCRPC asintomatico o minimamente sintomatico e consiste nello stimolare il sistema immunitario a riconoscere e attaccare le cellule tumorali [16]. Oltre a ciò, trattamenti innovativi come gli inibitori della PARP (ad esempio, olaparib) sono particolarmente promettenti per i pazienti con mutazioni nei geni di riparazione del DNA, come BRCA1/2, aumentando le possibilità di risposta positiva nei pazienti con tumori aggressivi [16].

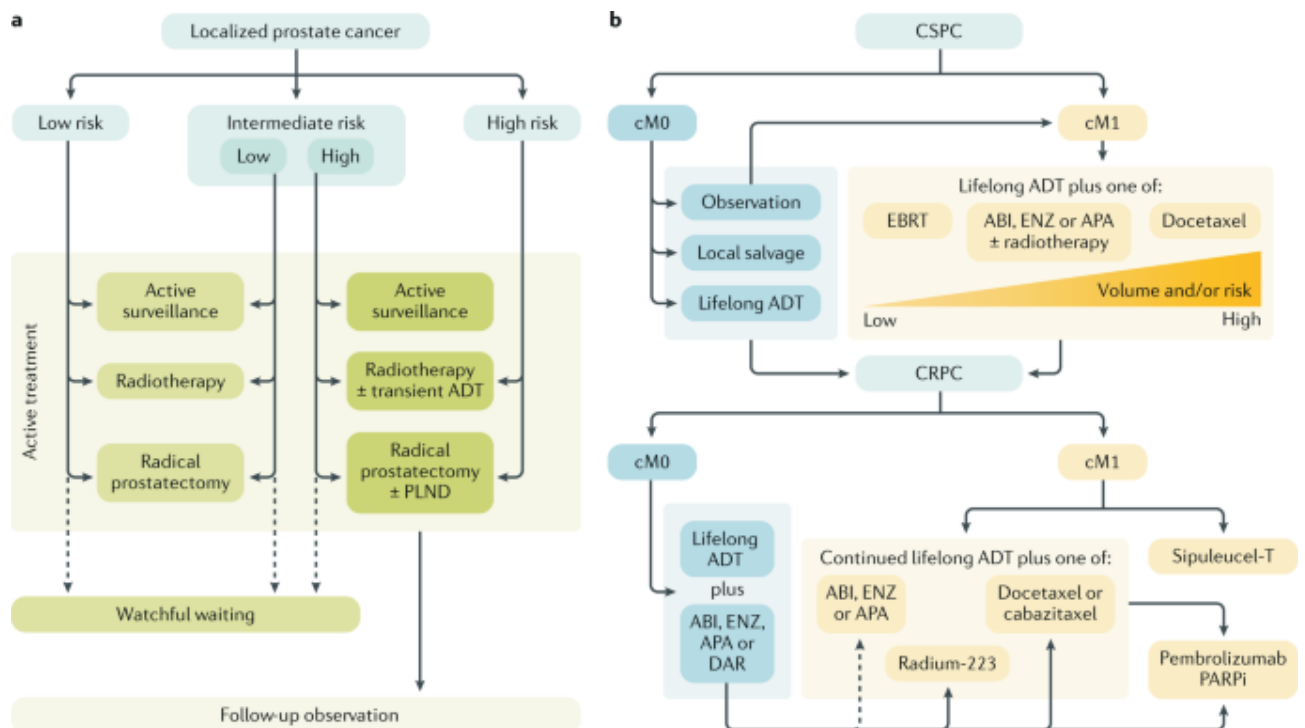


Figura 1.5: Overviev della gestione di un tumore alla prostata

1.3 MRI per il cancro alla prostata

La **risonanza magnetica (MRI)** svolge un ruolo fondamentale per il rilevamento e la diagnosi della lesione, in quanto permette di ottenere una rappresentazione tridimensionale della prostata ^[17].

Quando si studiano i tumori alla prostata si utilizzano due tecniche di acquisizione dell'immagine: bi-parametricMRI (bpMRI) e multi-parametricMRI (mpMRI). Le due tecniche prelevano le 2 sequenze di immagini, T2-weighted images (T2WI) e Diffusion weighted imaging (DWI), però la mpMRI differisce dalla bpMRI perché utilizza un'ulteriore sequenza di immagini che è la Dynamic Contrast Enhancement (DCE).

- **T2-weighted images (T2WI):** misura il contenuto d'acqua nei tessuti e permette di distinguere chiaramente le diverse zone anatomiche della prostata, come la zona di transizione e quella periferica. Quest'ultima appare come un'area bianca, mentre la zona di transizione presenta una combinazione di aree ad alto e basso segnale. Le aree in cui è presente il cancro vengono rilevate come regioni con segnale leggermente più basso ^[18,20].
- **Diffusion-weighted imaging (DWI),** misura la mobilità delle molecole d'acqua presenti nei tessuti. La maggiore densità cellulare nel tessuto canceroso ostacola il movimento delle molecole di acqua; quindi, è possibile individuare la differenza tra tessuto sano e tessuto canceroso all'interno dell'immagine poiché il tessuto canceroso appare come una zona più chiara rispetto alle altre. Da più acquisizioni a diverse intensità di pesatura della diffusione (indica quanto l'acquisizione è sensibile alla diffusione delle molecole d'acqua) si possono ricavare le immagini ADC che sono una misurazione quantitativa della diffusione dell'acqua, e nelle mappe ADC i tumori e i tessuti con bassa diffusione sono scuri ^[19]
- **Dynamic Intravenous contrast-Enhanced imaging (DCE),** queste immagini sono ottenute effettuando un'acquisizione subito dopo aver iniettato un mezzo di contrasto al paziente. Questa tecnica permette di evidenziare la vascolarizzazione dell'organo e in presenza di tumore questa vascolarizzazione sarà anormale rispetto al resto, rendendolo individuabile ^[18,21].

Inoltre, per avere una standardizzazione dell'acquisizione, interpretazione e reportistica delle immagini mpMRI sono state introdotte delle linee guida dal *The International Prostate MRI Working Group* che ha introdotto il **PI-RADS** (Prostate Imaging Reporting and Data System).

La prima versione è stata introdotta nel 2012, PI-RADS v1, la quale, tuttavia, presentava problemi di variabilità nell'applicazione e interpretazione. Successivamente è stata introdotta una seconda versione PI-RADS v2 che definisce tutti i protocolli da seguire per l'imaging di un tumore alla prostata e inoltre fornisce una scala di valori che va dall'uno al cinque per indicare la probabilità di un tumore clinicamente significativo ($GS \geq 7$): più è bassa la valutazione e più sarà bassa la probabilità di trovare un tumore clinicamente significativo ^[22].

I risultati delle due tecniche di imaging sono comparabili per quanto riguarda il rilevamento di un tumore alla prostata; tuttavia, la mpMRI è maggiormente accurata nella diagnosi in quanto fornisce dettagli più accurati sull'aggressività del tumore ^[23].

Capitolo 2

2.1 Intelligenza Artificiale

L'Intelligenza Artificiale (IA) è una disciplina che si concentra sulla creazione di sistemi informatici in grado di compiere attività che, in condizioni normali, richiederebbero l'intelligenza umana. Queste attività spaziano dal riconoscimento di oggetti alla comprensione del linguaggio naturale e alla risoluzione di problemi complessi. Quando si sviluppa un sistema Machine Learning, a differenza dei software tradizionali a cui vengono assegnate istruzioni statiche, si procede con la creazione di un algoritmo in grado di apprendere i dati forniti e di fare una predizione o prendere una decisione da essi [24].

In base al problema specifico e ai dati disponibili, si distinguono tre tipi di ML: apprendimento supervisionato, non supervisionato e per rinforzo [25,26].

- **Apprendimento supervisionato:** mediante tale tecnica si utilizza un dataset che fornisce sia elementi di input sia elementi etichettati di output e da questi il modello viene poi addestrato. La performance dell'algoritmo viene misurata attraverso la funzione di perdita; infatti, tanto minore è la funzione, quanto minore sarà l'errore commesso dalla rete. Una volta addestrato l'algoritmo esso potrà essere utilizzato per predire altri target partendo da dati di input non ancora visti. Inoltre, l'apprendimento supervisionato può risolvere due tipi di problemi: **regressione** e **classificazione**. Nella regressione l'algoritmo deve trovare una correlazione tra variabili di input e variabili di output, mentre nella classificazione l'algoritmo deve attribuire una classe ad ogni elemento del set di input;
- **Apprendimento non supervisionato:** a differenza dell'apprendimento supervisionato l'apprendimento non supervisionato non ha un dataset con variabili etichettate, per questo motivo il modello deve trovare proprietà comuni all'interno del set di input (**clustering**);

- **Apprendimento per rinforzo**, consiste nel fornire al modello un input iniziale e un obiettivo da raggiungere imponendo vincoli e azioni consentite. Lo scopo di questo modello è quello trovare la strada migliore per minimizzare la perdita.

Detto questo, occorre evidenziare che il Deep Learning è un sottoinsieme del Machine Learning, il quale sfrutta reti neurali profonde per analizzare grandi quantità di dati e riconoscere pattern complessi. A differenza degli algoritmi Machine Learning tradizionali, che richiedono la progettazione manuale delle caratteristiche da analizzare, il deep learning consente al sistema di apprendere direttamente dai dati grezzi. Poiché i dati non vengono etichettati, per avere un output ottimale dalla macchina bisogna fornire un gran volume di dati. È necessario, inoltre, che essi siano molto rappresentativi per evitare il rischio che l'algoritmo non sia in grado di trovare un pattern e quindi non riesca ad addestrarsi.

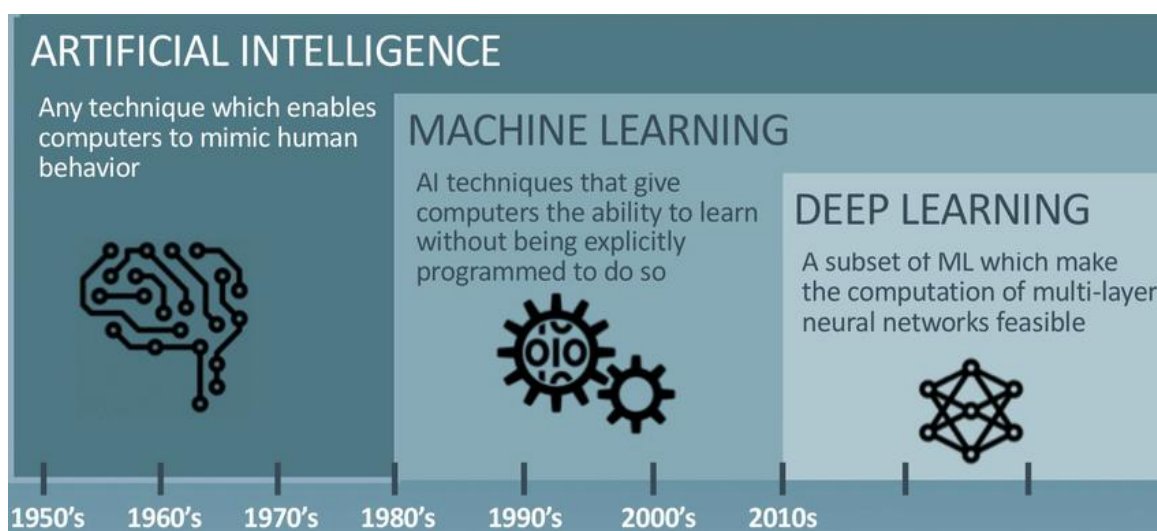


Figura 2.1: Relazione tra IA, ML e DL.

Tratta da https://www.researchgate.net/figure/Timeline-of-the-advances-in-artificial-intelligence-and-the-rise-of-machine-learning_fig4_350372354

Le reti neurali profonde trovano applicazione in molti ambiti. Grazie alla notevole capacità di riconoscere le immagini il deep learning si è fatto strada nel mondo sanitario. Infatti, ad esempio esso viene utilizzato per l'analisi di immagini mediche e per la diagnostica assistita dall'IA. In oncologia, il deep learning supporta i medici nella diagnosi, nel riconoscere il tipo, l'aggressività e l'avanzamento di tumori e nella previsione del decorso della malattia [27].

2.2 Il neurone

Una rete neurale artificiale è un modello che si ispira al cervello umano, e al modo in cui vengono captate elaborate e trasmesse le informazioni. L'unità fondamentale della rete neurale è il neurone o nodo.

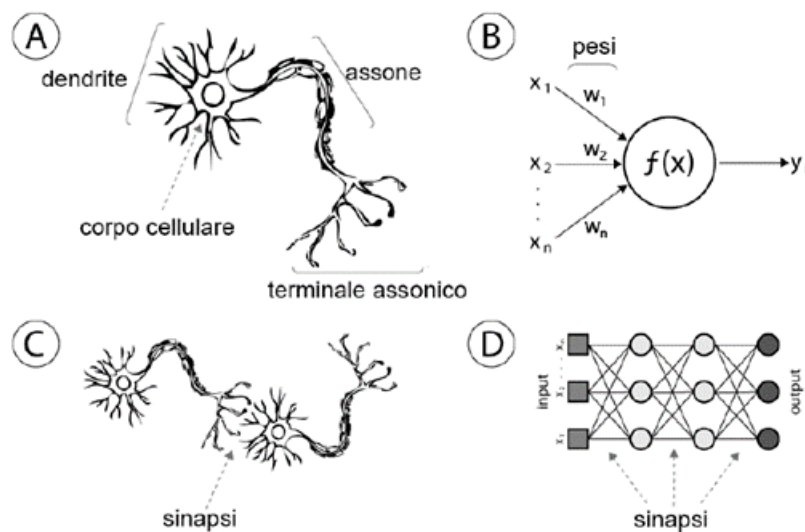


Figura 2.2: Confronto tra neurone biologico e neurone artificiale.
Tratta da: https://www.researchgate.net/figure/Diagram-to-correlate-between-Biological-and-Artificial-Neural-Network_fig3_269035338

I passaggi principali di un funzionamento del **neurone** sono:

- **Ricezione dell'input:** il neurone riceve diversi input da altri neuroni $x(n)$, la connessione dei diversi nodi è unidirezionale;
- **Pesatura dell'input:** ad ogni input viene associato il suo peso denominato w_{kj} , il quale determina l'importanza della connessione tra unità k e unità j . Il valore di w_{kj} può essere sia negativo che positivo, se il valore è negativo vuol dire che la connessione tra i due neuroni deve essere inibita; invece, se è positivo deve essere rafforzata;

- **Somma pesata:** tutti gli input moltiplicati ai rispettivi pesi vengono sommati ottenendo così uno scalare; a questa somma si può aggiungere un valore di bias, chiamato b_k , che serve a correggere il valore di input;

$$y_k = \sum_{j=1}^n w_{kj} * x_j + b_k$$

- **Funzione di trasferimento:** il valore della somma pesata viene fatto passare per una funzione detta funzione di trasferimento, il risultato che uscirà da questa funzione sarà l'output del neurone. Spesso la funzione di trasferimento è una funzione non-lineare poiché con quelle lineari l'output è semplicemente proporzionale all'input. Le funzioni di trasferimento non-lineari aiutano molto di più la rete ad apprendere modelli complessi ^[28,29,30].

2.2.1 Le funzioni di trasferimento

Le funzioni di trasferimento più utilizzate sono:

Sigmoide: è una delle prime funzioni utilizzate per aggiungere la non-linearità alla rete, la formula è

$$f(x) = \frac{1}{1 + e^{-x}}$$

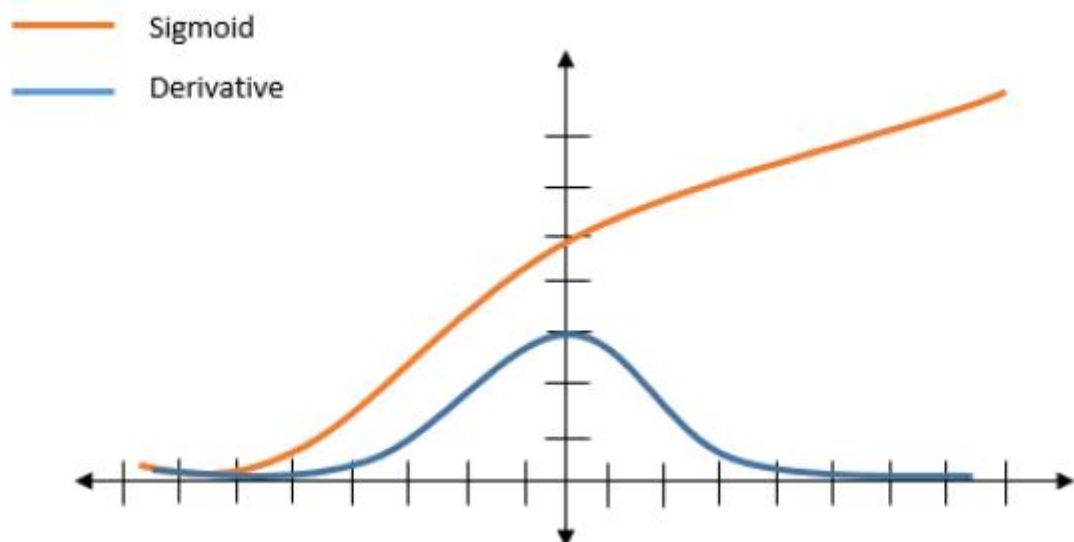


Figura 2.3: grafico della Sigmoide (in arancio) e della sua derivata (in blu). [33]

Questa funzione trasforma il valore in ingresso in un valore compreso tra 0 e 1, rendendola ideale nelle classificazioni binarie. Il difetto principale di questa funzione sta nel fatto che satura per i valori vicini allo 0 e all'1 dando origine al problema del “*vanishing gradient*” rallentando l'apprendimento della rete. Un altro problema consiste nel fatto che la funzione non è centrata nello 0 il che rende più lenta la

convergenza del modello [30,31,32]. A causa di questi difetti ultimamente viene utilizzata sempre meno, ad eccezione degli ultimi layer per le classificazioni binarie [31,32,33].

Tangente iperbolica (Tanh): molto simile alla funzione Sigmoide, il valore di output è compreso tra -1 e 1, la formula è

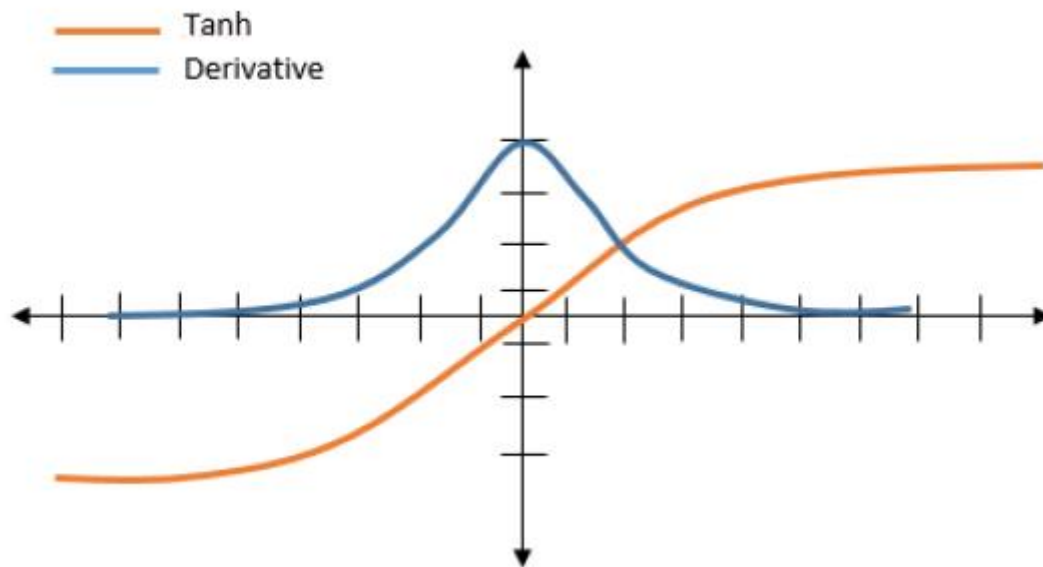


Figura 2.4: grafico della *Tanh* (in arancio) e della sua derivata (in blu). [33]

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

A differenza della funzione Sigmoid essa è centrata in 0 e questo aiuta il modello a convergere. Tuttavia, anch'essa soffre del *vanishing gradient* ed ha un costo computazionale elevato a causa dei calcoli esponenziali. Per via di questi difetti è utilizzata solo in modelli semplici mentre per le reti neurali profonde viene poco utilizzata [31].

ReLU (*rectified linear units*): è la funzione più utilizzata nelle reti profonde ed è definita come segue

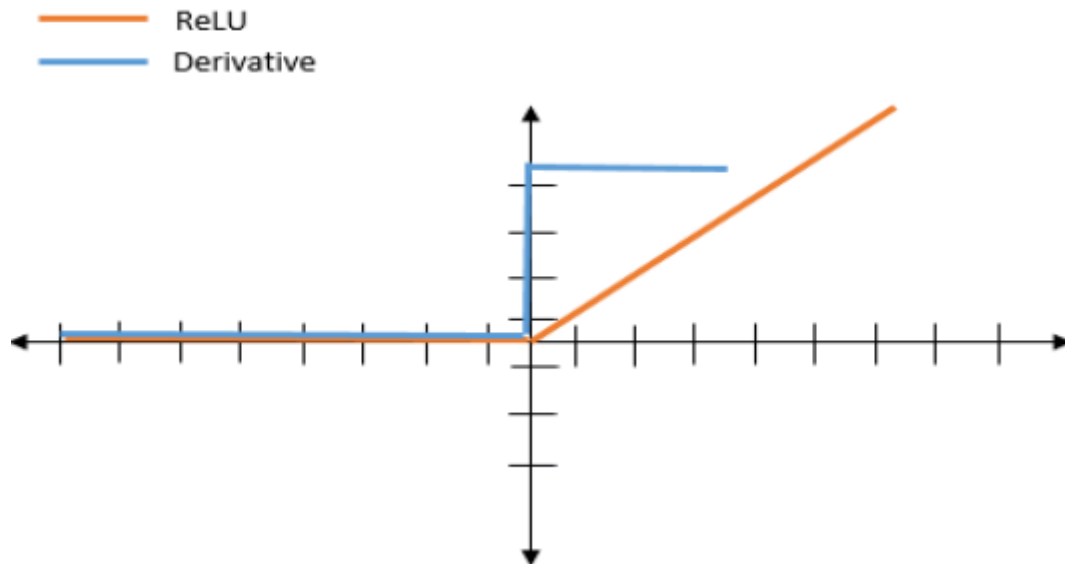


Figura 2.5: grafico della ReLU (in arancio) e della sua derivata (in blu). [33]

$$f(x) = \max(0, x)$$

La funzione ReLU restituisce semplicemente 0 per valori negativi di x e lascia invariati i valori positivi. Queste caratteristiche rendono questa funzione computazionalmente efficiente in quanto consiste in un confronto e non richiede calcoli esponenziali e/o divisioni da fare. Inoltre, risolve il problema del *vanishing gradient*. Questa funzione presenta però il difetto dei neuroni morti, i quali hanno valori negativi e vengono mappati a zero, rendendoli inefficaci per l'apprendimento [32,33].

Leaky ReLU: è una variante della ReLU, la formula è

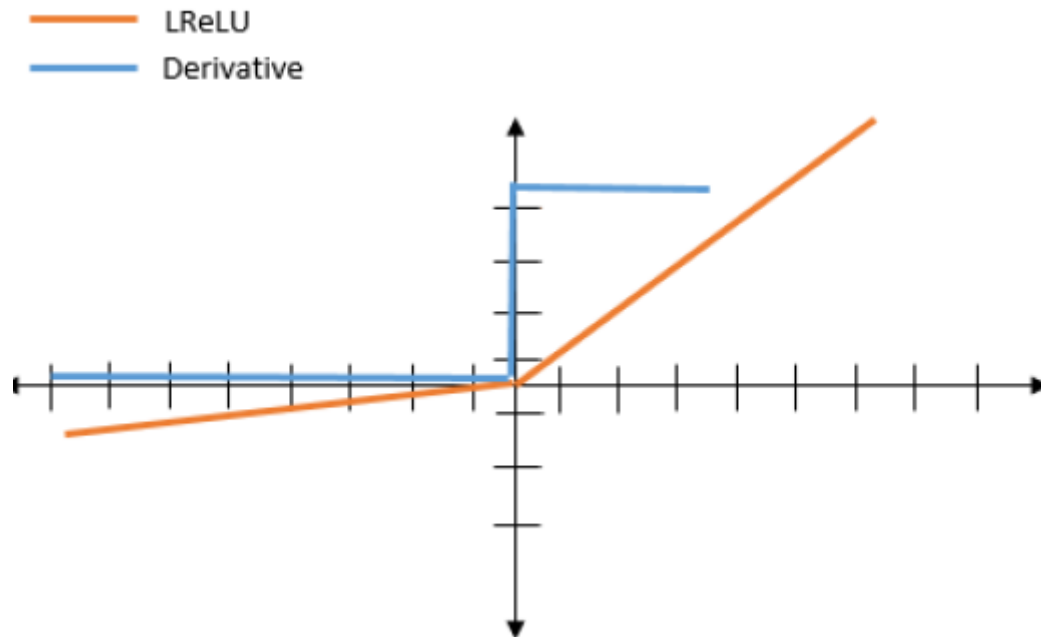


Figura 2.6: grafico della LReLU (in arancio) e della sua derivata (in blu). [33]

$$f(x) = \begin{cases} x, & x > 0 \\ ax, & x \leq 0 \end{cases}$$

Dove a è un parametro che definisce la pendenza dei valori negativi, spesso è impostato come 0.01

Questa variante elimina il problema dei neuroni morti mantenendo un gradiente anche per i neuroni morti. Tuttavia, la scelta di a può influire sulle prestazioni limitandone l'efficacia della funzione ^[32,33].

Exponential Linear Unit (ELU): anche questa è una variante della funzione ReLu ed è definita come segue

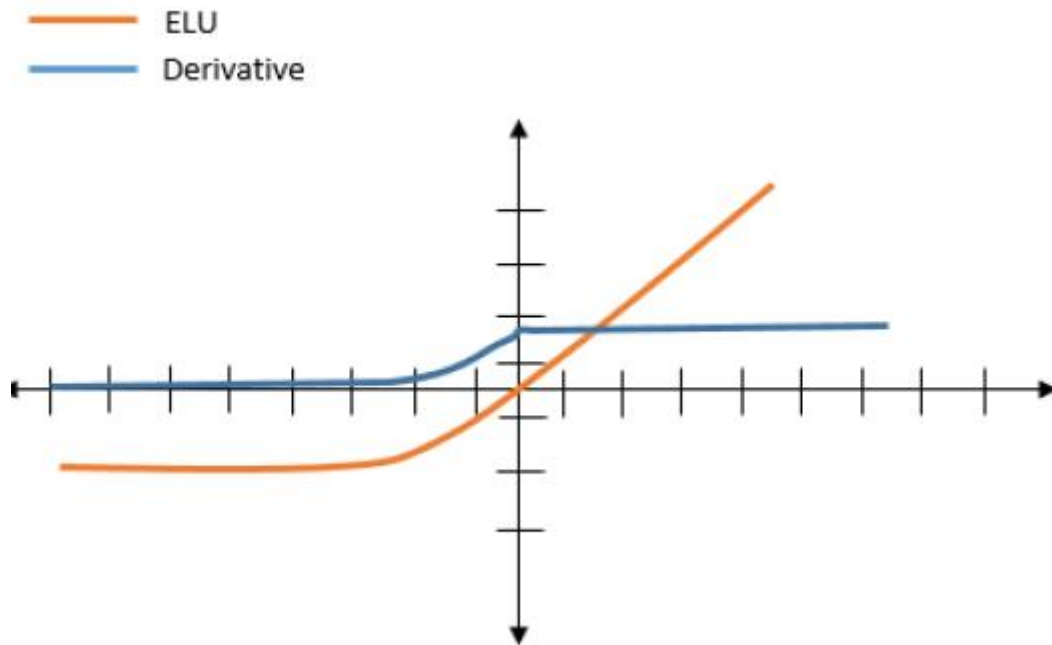


Figura 2.7: grafico della ELU (in arancio) e della sua derivata (in blu). [33]

$$f(x) = \begin{cases} x, & x > 0 \\ a(e^x - 1), & x \leq 0 \end{cases}$$

Dove a è un parametro che determina l'output dei valori positivi. Come per LReLU anche questa funzione risolve il problema dei neuroni morti e favorisce la convergenza del modello ma ha un maggiore costo computazionale ed è influenzata dal parametro a [32,33].

Swish: è una funzione introdotta da Google e combina le proprietà delle funzioni ReLu e sigmoide, la formula è

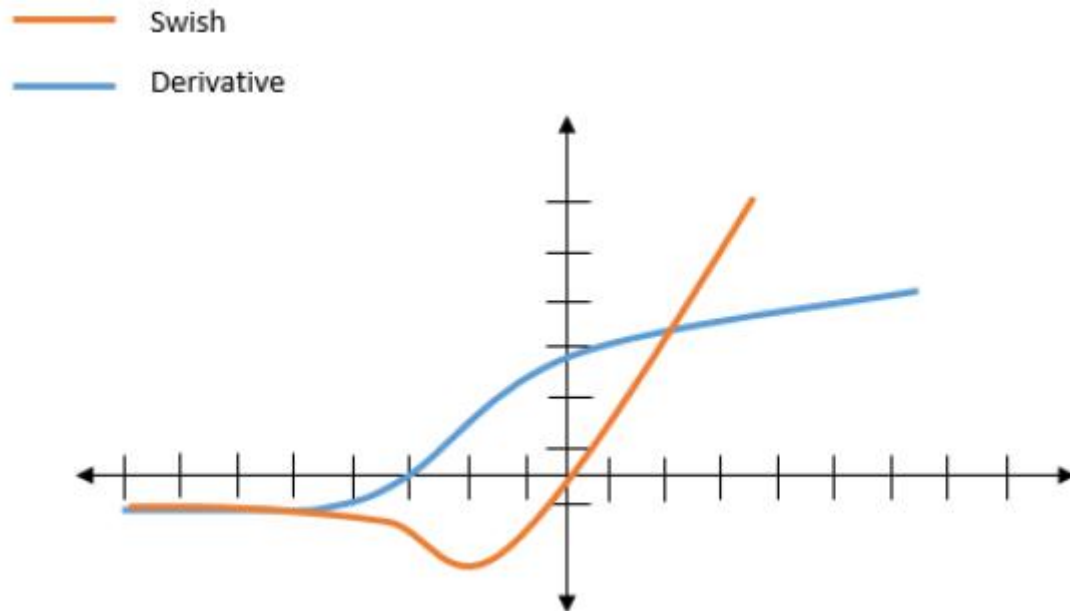


Figura 2.8: grafico della Swish (in arancio) e della sua derivata (in blu). [33]

$$f(x) = x * sigmoid(x) = \frac{x}{1 + e^{-x}}$$

Questa funzione elimina entrambi i problemi presentati dalle singole funzioni concernenti i neuroni morti e il *vanishing gradient*. Per tale ragione essa è una funzione molto performante soprattutto nelle reti profonde. Un fattore negativo è costituito dall'elevato costo computazionale in quanto deve risolvere sia la sigmoide che la moltiplicazione [31,32,33].

Softmax: è una funzione molto utilizzata nei problemi di classificazione multiclasse, non viene applicata mai in layer nascosti ma solo nell'ultimo, la sua formula è

$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

x_i rappresenta il valore di input del singolo nodo. La funzione calcola l'esponenziale di x_i e lo normalizza dividendolo per la sommatoria di tutti gli esponenziali; il risultato sarà una probabilità garantendo che la somma delle singole probabilità per ogni classe sia 1 [32,33].

2.3 Architetture delle reti neurali

Le reti neurali presentano diverse architetture a seconda del problema da risolvere. Le principali architetture sono: Feed-Forward Networks, Convolutional neural networks, Recurrent neural networks, Generative adversarial networks, Transformer neural networks.

2.3.1 Feed-Forward Networks (FFN)

La **rete FFN** è uno dei modelli più semplici tra le architetture presenti. La caratteristica principale di questa rete sta nel fatto che l'informazione si propaga attraverso la rete in una sola direzione, senza cicli. Ogni rete presenta un layer di input, uno o più *hidden layers* e un *output layer*; ogni layer rappresenta un insieme di neuroni che processano e trasferiscono l'informazione. L'input layer acquisisce i dati grezzi e li trasferisce ai layer successivi, la dimensione di questo layer dipende dalla dimensione dei dati di input da dare alla rete. Gli hidden layer sono il cuore della rete: questi strati elaborano le informazioni di input attraverso una funzione di attivazione e trasferiscono l'informazione ai neuroni successivi, maggiore sarà la profondità e maggiore sarà la

capacità della rete ad apprendere pattern più complessi. L'output layer rappresenta l'ultimo strato dell'architettura e fornisce il risultato finale dell'elaborazione, che sarà una classe nei problemi di classificazione o una previsione nei problemi di regressione. Nei problemi di classificazione il numero di neuroni di questo strato sarà uguale al numero delle classi. Ogni neurone del singolo strato è connesso a tutti i neuroni dello strato successivo, con un determinato peso che determina l'apprendimento della rete [34,35].

L'addestramento della rete si può dividere in due step, la Forward Propagation e la Back Propagation [34,36].

Nella Forward Propagation il segnale parte dallo strato di input e si trasferisce negli strati nascosti dove in ogni neurone viene calcolata la somma pesata e successivamente fatta passare per la funzione di attivazione. Si ripetono questi passaggi finché l'informazione non arriva allo strato di output in cui viene elaborato il risultato.

Una volta calcolati gli output si calcola l'errore globale della rete con la seguente formula

$$E = \frac{1}{2} \sum_o (y_o - t_o)^2$$

y_o sono i valori di output mentre t_o sono le etichette reali.

Successivamente si passa alla fase di Back Propagation, l'errore viene propagato a ritroso nella rete; viene calcolato il gradiente di E rispetto a ciascun peso w_{ij} , aggiornandolo secondo questa equazione

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

η è il learning rate. Esso consiste in un parametro che assicura che il modello converga senza oscillazioni [36,37].

Le reti FFN vengono oggi utilizzate come blocchi principali per altre architetture più complesse in quanto l'efficienza di queste reti è molto dipendente dalla loro struttura;

infatti, la capacità della rete di performare dipende molto dal numero di neuroni e dal numero di strati nascosti utilizzati [35,38].

2.3.2 Convolutional Neural Network (CNN)

Un importante impulso alla risoluzione di problemi legati al riconoscimento di pattern e classificazione delle immagini è stato fornito dalle Convolutional Neural Networks. Questa tipologia di rete neurale è stata sviluppata partendo dalle ANN, le quali negli ultimi anni hanno rinnovato la materia relativa alla visione artificiale e elaborazione di immagini, poiché sono in grado di apprendere autonomamente gli elementi distintivi dei dati non strutturati. La differenza delle CNN con le ANN consiste nel fatto che viene aggiunta un'altra dimensione ai vari layer della rete; infatti, mentre in precedenza erano presenti solo due dimensioni (altezza*larghezza) nelle CNN è stata aggiunta la profondità, che non si riferisce alla dimensione totale della rete, ma solo al volume di attivazione. Oltre a ciò, i neuroni non sono tutti connessi tra loro come nelle ANN ma i neuroni di un layer si connettono solo ad una piccola regione del layer precedente [39].

Esistono diverse varianti di CNN ciascuna con caratteristiche specifiche, ma tutte presentano 3 blocchi principali:

1. **Strati di convoluzione:** è il blocco principale delle CNN. È in questo strato che le informazioni rilevanti provenienti dall'input vengono elaborate; l'elaborazione avviene tramite il processo di convoluzione, che consiste nel far passare un filtro (o kernel) sull'immagine per produrre una **mappa di attivazione** o **feature map**. Ogni filtro è progettato per rilevare una determinata caratteristica visiva, come bordi, texture o forme.

Il kernel è sostanzialmente una matrice definita da tre variabili [39,40];

La **dimensione**, solitamente impostata su 3x3, 5x5 o 7x7, determina l'area locale sulla quale il filtro agisce.

Lo **stride**, determina di quanti pixel il filtro si deve spostare sull'immagine. normalmente viene impostato su 1, ma a volte si può optare per valori più grandi nel caso si volesse una riduzione della dimensionalità delle feature maps.

Il **padding** è utilizzato per aggiustare indirettamente la dimensione delle feature maps e a volte capita che la dimensione del filtro non è congruente con la dimensione dell'immagine elaborata. Esistono diversi tipi di padding: il valid padding, il quale consiste nel rimuovere l'ultima convoluzione non congruente alla dimensione del filtro, riducendo la dimensione dell'output; il same padding, che aggiunge zeri ai bordi mantenendo la dimensione dell'output uguale a quella di input; infine, il full padding aggiunge zeri ai bordi aumentando volutamente la dimensione dell'output ^[41].

Dopo ogni operazione di convoluzione viene applicata una funzione di attivazione per aggiungere non-linearità al modello, la più utilizzata è ReLu e le sue varianti ^[42].

- 2. Strato di Pooling:** questo strato è utilizzato per ridurre la dimensionalità delle feature maps, riducendo quindi i dati da elaborare e di conseguenza il costo computazionale dell'addestramento, inoltre riduce anche il rischio di overfitting (il modello si adatta eccessivamente al set di addestramento e quindi inefficace per altri set). Anche lo strato di pooling, come in quello di convoluzione utilizza un filtro di una certa dimensione e con un determinato stride da far passare sull'immagine, però in questo caso non ha pesi. Esistono due tecniche principali per implementare lo strato di pooling che sono: max pooling e average pooling ^[41]. Nel max pooling il filtro che elabora l'immagine prende solo il pixel con valore più alto scartando il resto, mentre nell'average pooling il filtro che si sposta lungo l'immagine calcola la media dei pixel. Poiché il pooling riduce la quantità di informazioni elaborabili dalla rete è consigliato elaborare la stessa zona più volte, **overlapping**, per non perdere dettagli importanti e non utilizzare filtri di dimensione superiori al 3 o le prestazioni del modello potrebbero diminuire di molto ^[39].
- 3. Strato completamente connesso:** questo blocco è costituito da uno o più layer di neuroni completamente connessi, per far rendere questo strato al meglio è necessario che l'output dei blocchi precedenti sia di una sola dimensione. Grazie alla struttura di questo blocco i neuroni sono in grado di elaborare l'immagine nel suo insieme, raccogliendo tutte le caratteristiche rilevate negli strati precedenti. Solitamente l'ultimo blocco di questo strato avrà lo stesso numero di neuroni tante quanto sono le classi, e la funzione di attivazione maggiormente utilizzata è softmax.

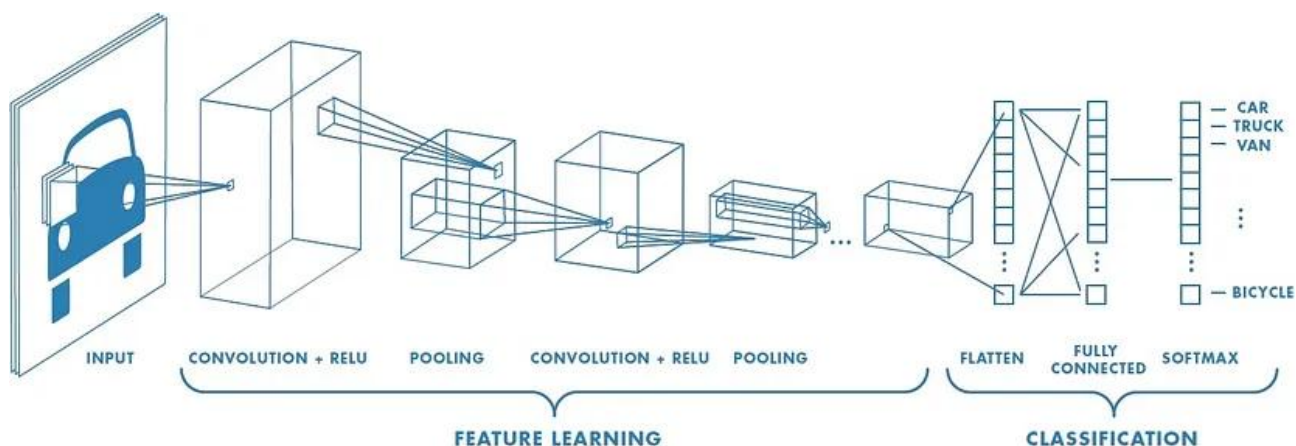


Figura 2.9: Schematizzazione di una rete CNN, da sinistra a destra si possono notare i layer di convoluzione, di pooling e infine i layer densi. Sumit Saha. A comprehensive guide to convolutional neural networks - the eli5 way, Dec 2018.

Il rischio di overfitting in questa tipologia di rete neurale è molto alto, la regolarizzazione è l'arma principale per combatterlo e migliorare la generalizzazione della rete. Le tecniche principali per ridurre l'overfitting sono:

- **Dropout:** è una tecnica che spegne neuroni casuali, prevenendo l'eccessiva dipendenza della rete da certe connessioni e migliorando la robustezza;
- **Batch Normalization:** è un metodo utilizzato per la normalizzazione degli input e accelerare la convergenza della rete;
- **Data Augmentation:** aumenta il numero di campioni del set di dati generando varianti casuali delle immagini (rotazioni, traslazioni, scaling) e migliorando la capacità di generalizzazione delle CNN.

L'utilizzo di queste tecniche, combinate con architetture ottimizzate, consente di migliorare le prestazioni delle CNN anche in contesti con set di dati limitati ^[40].

2.3.3 Recurrent neural networks (RNN)

Le **Reti Neurali Ricorrenti** sono un'altra tipologia di ANN, la quale viene utilizzata per riconoscere e processare dati sequenziali come testo e audio. Le RNN riescono a fare ciò perché a differenza delle architetture descritte in precedenza ha una “memoria” degli stati precedenti proveniente dalle connessioni ricorrenti presenti nel modello ^[43].

La struttura di una RNN è molto simile a quella di una FNN e la particolarità che li distingue sta nel fatto che nei layer nascosti si trova anche un flusso ricorrente o a ciclo automatico, che funge da memoria a breve termine. In questa maniera la rete può utilizzare per la sua previsione sia l'input che sta ricevendo sia quello che ha in archivio. Il problema principale di questa architettura è il vanishing gradient che limita l'apprendimento della rete soprattutto per problemi con dipendenza a lungo termine e per affrontare questo problema è stata introdotta una variante della RNN che è la Long Short-Term Memory (LSTM) e successivamente la Gated Recurrent Unit (GRU) ^[44,45].

Le LSTM introducono dei blocchi di memoria speciale nel layer nascosto: questi blocchi vengono chiamate celle, ogni cella presenta 3 gate, uno in ingresso uno in uscita e uno per “dimenticare”, questi gate gestiscono quali informazioni conservare o scartare rendendo la rete più performante.

Le GRU sono molto simili alle LSTM con la differenza che non utilizzano più le celle come blocco di memoria ma il layer nascosto e presentano solo due gate uno di reset e uno di aggiornamento, rendendole quindi molto più snelle ed efficienti.

2.3.4 Generative adversarial networks (GAN)

Le GAN sono una tipologia di rete neurali sviluppata negli ultimi anni, le quali sfruttano l'apprendimento semi-supervisionato e non supervisionato. Queste reti sono utilizzate soprattutto per la modellazione generativa, cioè la capacità della rete di individuare automaticamente pattern presenti nei dati di input e generare un output plausibile che potrebbe fare parte del set di input ^[46,47].

L'architettura di base di una rete GAN è costituita principalmente da due reti neurali:

- **Generatore (G):** questo blocco riceve un vettore di dati casuali da input, non ha accesso infatti al set reale e cerca di trasformare l'input in immagine sempre più vicina a quelle del set originale;
- **Discriminatore (D):** funziona come un classificatore binario a cui vengono assegnate sia le immagini reali sia le immagini generate dal generatore e gli assegna un'etichetta (reale o falso) ^[46,48];

L'addestramento consiste nel far competere le due reti; l'obiettivo del generatore è quello di ingannare il discriminatore generando immagini sempre più vicine alla realtà, il discriminatore invece cerca di minimizzare l'errore. La rete smette di addestrarsi quando si raggiunge l'equilibrio, noto come **equilibrio di Nash** ^[48,49]; una volta raggiunto il discriminatore non è più in grado di distinguere i dati reali con quelli generati arrivando all'obiettivo della rete.

L'addestramento della GAN si basa sulla seguente formula matematica

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))]$$

- x rappresenta i dati provenienti dal set reale;
- z è un vettore di rumore casuale, solitamente campionato da una distribuzione semplice come una distribuzione normale o uniforme;
- $G(z)$ è il campione generato dal generatore utilizzando il rumore z ;
- $D(x)$ è la probabilità calcolata dal discriminatore che un dato x provenga dalla distribuzione reale, e non da quella generata.

Il Discriminatore cerca di massimizzare la funzione $V(D, G)$, cercando di massimizzare $D(x)$ facendolo tendere a 1, e cercando di minimizzare $D(G(z))$ facendolo tendere a 0. Al contrario l'obiettivo del Generatore è quello di massimizzare la probabilità che il generatore sbagli massimizzando $D(G(z))$ ^[47].

Le GAN hanno avuto un impatto significativo su una vasta gamma di applicazioni, le principali sono:

- Generazione di immagini: le GAN ad oggi sono in grado di generare immagini fotorealistiche anche di volti umani e per questo vengono utilizzate spesso nella creazione di contenuti sintetici per il cinema, i videogiochi e la realtà virtuale [46];
- Super risoluzione delle immagini: questa tecnica prevede l'utilizzo delle GAN per aumentare artificialmente la risoluzione di un'immagine [48];
- Traduzione delle immagini: le GAN sono in grado di trasformare le immagini, ad esempio possono essere utilizzate per cambiare una da diurna a notturna o trasformare una foto realistica a cartone animato [49];
- Sintesi del parlato e produzione di musica: oltre alle immagini le GAN sono anche in grado di riprodurre realisticamente la voce umana e anche di produrre musica [47].

2.4 AI nel medical imaging

Il deep learning ha migliorato notevolmente anche la diagnostica delle immagini mediche. Infatti, prima dell'utilizzo di questi algoritmi era il solo radiologo che stabiliva se ci fosse o meno una determinata patologia e quanto questa si trovasse in uno stadio avanzato o meno. Negli ultimi anni lo sviluppo del deep learning ha portato benefici anche in questo ambito, introducendo sistemi che consentono di aiutare il radiologo nell'individuazione di anomalie, **computer-aided detection (CADE)**, e nella diagnosi di patologie, **computer-aided diagnosis (CADx)** [50].

Solitamente quando vengono utilizzati sistemi CADE il radiologo effettua una prima diagnosi esaminando le immagini; successivamente quest'ultime vengono elaborate dalla rete che individua eventuali anomalie e il risultato viene messo a disposizione del radiologo, il quale controlla se la diagnosi è confermata o se ci sono altre anomalie [51]. Questa doppia revisione aiuta a ridurre i tassi di falsi negativi e migliora l'accuratezza diagnostica complessiva. Invece i sistemi CADx si concentrano sulla diagnosi, quindi elaborano immagini in cui l'anomalia è stata già diagnosticata e segmentata; quindi, l'algoritmo effettua una previsione sull'eventuale malignità e su quale sia la sua aggressività [52].

Le fasi di implementazione dei CAD includono l'acquisizione dati, il pre-processing, la segmentazione, la rilevazione e la classificazione. L'acquisizione dati comprende la creazione di un set di immagini mediche per l'addestramento, incluse radiografie, TC,

MRI e ultrasuoni, nonché informazioni mediche e analisi di laboratorio ^[51,53]. La difficoltà nell'ottenere dati sufficienti e di alta qualità ha portato allo sviluppo di database pubblici che facilitano l'accesso alle immagini mediche ^[50].

Il pre-processing consiste nel classificare i dati di addestramento e nel normalizzarli per ridurre al minimo le differenze al fine di rendere la performance della rete migliore ^[50].

La detection è la fase in cui l'algoritmo identifica le zone anomale e ciò mediante il riconoscimento delle caratteristiche, come dimensione, forma, livelli di grigi etc... Infine, si esegue la classificazione in cui si può utilizzare un algoritmo di Machine Learning (Decision Tree, K-Nearest Neighbor (KNN)) oppure una rete Deep Learning ^[52].

I sistemi CAD vengono utilizzati principalmente nell'individuazioni di tumori, soprattutto al seno, ai polmoni, al colon e alla prostata, ma anche per patologie cardiovascolari o al cervello ^[53].

Nel contesto del tumore alla prostata, sono stati sviluppati numerosi algoritmi in grado di svolgere funzioni specifiche in base al compito richiesto.

Per la segmentazione di tale tumore vengono utilizzati diverse architetture CNN adattate al trattamento di immagini mediche. Una delle principali architetture è la rete U-Net ^[54], la quale presenta due percorsi principali, uno di encoder e uno di decoder. Il percorso di encoder è una semplice architettura CNN in cui sono presenti layers convoluzionali e consiste in un upsampling mediante il quale viene ricostruita l'immagine segmentata. Questa rete permette la segmentazione pixel-wise, in cui ogni pixel viene considerato come un'entità indipendente e classificato in base alle sue caratteristiche visive e al contesto fornito dai pixel vicini. La segmentazione produce una mappa di classificazione della stessa dimensione dell'immagine di input, in cui i valori rappresentano le classi assegnate a ciascun pixel ^[55,54]. Per migliorare le prestazioni della rete sono state anche apportate modifiche come nello studio di Ali Badiezadeha ^[56], in cui la rete U-net è stata potenziata sostituendo alcuni strati convoluzionali con moduli H-VSS. Questi moduli utilizzano operazioni di scansione selettiva bidimensionale (2D-selective-scan) per catturare dipendenze locali e globali, migliorando la comprensione delle relazioni spaziali nell'immagine. Per la diagnosi del tumore invece è molto utilizzata la ResNet e le sue varianti, la quale è in grado di

mitigare l'effetto del vanishing gradient grazie alla sua architettura. Ad esempio, nello studio di Talaat ^[57], viene utilizzata una Res-Net50 modificata, combinata con una Faster R-CNN, in cui vengono analizzate 11.000 immagini MRI per l'addestramento della rete, che offre dei risultati migliori rispetto alle reti presenti in letteratura raggiungendo il 97,40% di accuratezza. Un altro approccio è stato quello di Campanella ^[58] in cui è stato presentato un sistema di apprendimento profondo che utilizza solo le diagnosi riportate come etichette per l'allenamento, senza la necessità di annotazioni sul set da parte del patologo, riducendo quindi costi e tempo. Per questo studio è stata utilizzata una ResNet34 seguita da una RRN per combinare informazioni a livello di diapositiva, migliorando la sensibilità diagnostica.

Il deep learning è stato utilizzato negli anni anche per la predizione dell'aggressività del tumore alla prostata, andando a classificare il Gleason Grade della lesione. Un esempio è lo studio effettuato da D. Müller ^[59], molto recente, in cui sono stati analizzati un totale di 325 casi di carcinoma prostatico utilizzando diverse architetture di reti neurali quali DenseNet121, ResNeXt101, Xception, ViT short for Vision Transformer (variant B16), and ConvNeXt (variant Base). I risultati hanno mostrato prestazioni eccellenti per la rete ConvNeXt con un'area sotto la curva ROC (AUC) di 0,991 e un'accuratezza del 97,4%. Nello studio di Karimi ^[60], i ricercatori hanno combinato le predizioni di 3 diverse reti neurali CNN che analizzano porzioni di immagini a diverse dimensioni e successivamente le predizioni delle CNN sono poi combinate attraverso un modello di regressione logistica. Inoltre, hanno proposto nuove tecniche di data augmentation, inclusi l'uso di GAN per generare nuove immagini sintetiche e deformazioni elastiche per ampliare la diversità dei dati di addestramento. I risultati hanno mostrato che il loro metodo raggiunge una precisione del 92% nel distinguere tessuti cancerosi da benigni e dell'86% nel distinguere i gradi Gleason bassi (grado 3) da quelli alti (gradi 4 e 5).

Capitolo 3

3.1 Dataset

Il dataset utilizzato per questo progetto è composto da immagini mediche provenienti da 173 pazienti affetti da tumore alla prostata, acquisite tramite quattro diverse marche di scanner MRI: GE Medical System (101 pazienti), Philips (26 pazienti), Siemens (45 pazienti) e Toshiba (1 paziente). Il set presentava 3 macrogruppi, suddivisi in base al loro Gleason Grade, GG. In particolare, erano presenti 31 pazienti con GG=1, 70 con GG=2 e 72 con GG>2. Per garantire un'analisi robusta e clinicamente rilevante, sono stati esclusi i pazienti con GG = 1, in quanto le lesioni in questa categoria sono considerate non significative dal punto di vista clinico, mentre gli altri due insiemi sono stati etichettati rispettivamente con classe 0 e classe 1.

Dopo questa scrematura, il numero di pazienti utilizzati è sceso a 142, di cui:

- **102 pazienti** presentavano una singola lesione;
- **33 pazienti** avevano due lesioni;
- **6 pazienti** mostravano tre lesioni;
- **1 paziente** presentava ben quattro lesioni.

Le immagini associate a ciascun paziente includevano diverse sequenze, come la pesata in T2, la sequenza ADC e una o più sequenze DWI a diversi *b-value*. Per il progetto è stata utilizzata la sequenza con il più alto *b-value*, poiché è il più sensibile per la rilevazione di eventuali lesioni. A ciascuna serie di immagini è stata associata una maschera della prostata che evidenziava le tre zone anatomiche principali (centrale, transizionale e periferica) e una maschera del tumore che identificava le aree interessate da lesioni.

3.2 Pre-processing

Le immagini dei pazienti presentavano inizialmente una notevole disomogeneità a causa delle diverse tecniche di acquisizione e dei piccoli movimenti dei pazienti durante le scansioni. Pertanto, è stato necessario un pre-processing accurato per uniformare i dati e preparare il set per l'addestramento delle reti neurali convoluzionali.

Il primo passo del pre-processing è stato l'allineamento delle immagini, eseguito utilizzando la libreria Python Simple-ITK. La sequenza T2 è stata scelta come riferimento per il processo di co-registrazione grazie alla sua capacità di fornire dettagli chiari sulla morfologia della prostata e alla sua minore sensibilità agli artefatti di movimento. Le immagini ADC e DWI sono state quindi riallineate a questa sequenza di riferimento mediante una registrazione non rigida basata su una trasformazione B-spline, garantendo la coerenza spaziale tra le diverse tipologie di immagini. Le immagini registrate sono state quindi campionate e ritagliate per corrispondere alle dimensioni della sequenza di riferimento.

Successivamente, è stato implementato un algoritmo per la selezione delle slice sfruttando le maschere manuali della prostata e del tumore precedentemente annotate. In particolare, sono state analizzate entrambe le maschere per identificare le slice contenenti le informazioni rilevanti per il progetto, cioè quelle contenenti i pixel che identificano la presenza della prostata e/o del tumore. Fatto ciò, per ogni volume, sono state escluse tutte le slice completamente nere e quindi superflue. Inoltre, per uniformare la profondità dei volumi è stato scelto come riferimento il paziente con il maggior numero di slice pari a 24; le immagini con un numero inferiore sono state completate ripetendo le slice alle estremità. In questo modo si è assicurata l'omogeneità del set di dati, mantenendo centrati i dettagli anatomici importanti.

Le immagini sono state quindi ridimensionate a una risoluzione di 128x128 pixel per trovare un equilibrio tra qualità visiva e dimensioni di calcolo accettabili. Poiché i dati originali variavano anche in termini di rappresentazione numerica (ad esempio, float a 32 o 64 bit), è stato necessario uniformare i valori dei pixel in una scala di grigi con un range da 0 a 255, mantenendo il formato float 32-bit per garantire la compatibilità

con gli algoritmi di deep learning. Questi passaggi hanno reso il set di dati più omogeneo, favorendo un miglior addestramento delle reti neurali convoluzionali e un'analisi più accurata delle immagini di risonanza magnetica.

3.3 Costruzione del classificatore

Per questo progetto sono state utilizzate diverse architetture di CNN in cui varia sia il numero di neuroni che il numero di layer. Si è partiti da una rete contenente un input layer, seguito da un layer convoluzionale e da due layer completamente connessi, uno dei quali fornisce l'output di classificazione. A questa rete è stato prima incrementato il numero di neuroni e successivamente il numero di layer convoluzionali. Le configurazioni utilizzate sono descritte nelle seguenti tabelle.

1^ configurazione CNN con un solo layer convoluzionale

	Prova 1	Prova 2	Prova 3	Prova 4
Input layer				
Conv2D	8	16	24	32
Fully Connected	8	16	24	32
Fully Connected	2	2	2	2

2^ configurazione CNN con due layers convoluzionali

	Prova 5	Prova 6	Prova 7	Prova 8
Input layer				
Conv2D	8	16	24	32
Conv2D	8	16	24	32
Fully Connected	8	16	24	32
Fully Connected	2	2	2	2

3^ configurazione CNN con tre layers convoluzionali

	Prova 9	Prova 10	Prova 11	Prova 12
Input layer				
Conv2D	8	16	24	32
Conv2D	8	16	24	32
Conv2D	8	16	24	32
Fully Connected	8	16	24	32
Fully Connected	2	2	2	2

In aggiunta a queste configurazioni sono stati eseguiti altri due addestramenti, uno con architettura che presentava 4 layers convoluzionali e un'altra con 5 layers convoluzionali. Entrambe con 32 neuroni per ogni layer.

3.4 Ottimizzazioni

Per tutti gli addestramenti il set di dati è stato suddiviso in un training set di 120 pazienti e un test set di 22 pazienti. Per migliorare la capacità predittiva e ridurre il rischio di overfitting, il training set è stato ulteriormente suddiviso per la validazione incrociata (cross-folding) in cinque parti, utilizzando quattro parti per l'addestramento e la quinta per la validazione. Inoltre, il dataset è stato arricchito mediante tecniche di data augmentation, moltiplicando artificialmente il numero di immagini per quattro volte e applicando leggere modifiche quali rotazioni, traslazioni e zoom, allo scopo di aumentare la varietà dei dati e migliorare la generalizzazione dei modelli.

Per ogni layer convoluzionale utilizzato la funzione di attivazione impostata è stata la ReLu. Questa funzione è stata preferita per la sua capacità di introdurre non linearità, permettendo alla rete di apprendere relazioni complesse tra i dati. La ReLU ha il vantaggio di essere semplice da calcolare e aiuta a mitigare il problema del gradient vanishing, mantenendo il modello efficiente durante l'addestramento. Per l'output layer, è stata utilizzata la funzione di attivazione sigmoid, ideale per la classificazione binaria, poiché trasforma l'output in un valore compreso tra 0 e 1, interpretato come

probabilità. Dopo ogni layer convoluzionale è stata inserita la Batch Normalization, che serve a standardizzare l'output per il layer successivo, mantenendo la distribuzione dei dati consistente e migliorare l'addestramento.

Per prevenire l'overfitting e migliorare la generalizzazione del modello sono state implementate due tecniche di regolarizzazione. La prima è la tecnica del Dropout, inserita tra ogni layer ed impostata a 0.5, e consiste nel disattivare randomicamente la metà dei neuroni per ogni iterazione. La seconda è la regolarizzazione L2, impostata a 0.01, utilizzata per mantenere i pesi più piccoli e stabili, applicando una penalità proporzionale al loro quadrato.

L'ottimizzatore scelto è Adam noto per la sua efficienza nel trovare il minimo della funzione di perdita grazie alla combinazione tra velocità di convergenza e stabilità. Per migliorare la convergenza è stato impostato un learning rate iniziale di 0.01 e durante l'addestramento è stato adottato un approccio decrescente per il learning rate, riducendolo gradualmente. Come funzione di loss è stata scelta la `binary_crossentropy` in quanto è la più performante per i problemi di classificazione binaria. Per l'addestramento è stato impostato un batch size di 10 elementi e un numero di epoche pari a 100, per garantire che il modello avesse abbastanza tempo per l'addestramento. È stato impostato anche un early stop, che è una tecnica di regolarizzazione che interrompe il modello quando non apprende, in cui si monitora la funzione di perdita nel validation set e se questa non decresce per 10 epoche consecutive allora l'addestramento si interrompe, ripristinando i pesi con le migliori performance.

Oltre alle configurazioni provate in precedenza sono state fatte 4 ulteriori prove sulla configurazione che ha avuto il risultato migliore, una aggiungendo un altro layer completamente connesso sempre con 8 neuroni, e altri modificando alcuni parametri di addestramento. In particolare, è stato provato un diverso batch size ($bs=5$), un diverso learning rate iniziale ($lr=0.01$) e una combinazione di entrambi.

3.5 Metriche di valutazione

L'analisi delle performance di un modello assume un aspetto cruciale per comprenderne la qualità e l'efficacia. Per il progetto sono state adottate le seguenti tecniche:

3.5.1 Confusion Matrix

È una delle tecniche maggiormente diffuse per l'analisi delle performance di algoritmi di classificazione, in quanto permette di mostrare i risultati ottenuti in tabella semplificandone la lettura. Per le classificazioni binarie la confusion matrix è una tabella 2x2

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 3.1: Immagine di Confusion Matrix.

I valori presenti nella tabella sono:

- Veri Positivi (VP), numero di positivi correttamente classificati;
- Veri Negativi (VG), numero di negativi correttamente classificati;
- Falsi Positivi (FP), numero di negativi classificati erroneamente;
- Falsi Negativi (FN), numero di positivi classificati erroneamente.

Attraverso questi valori si possono ricavare diversi parametri che descrivono vari aspetti del modello.

L'**accuracy** è il primo valore che si va a calcolare e indica la performance generale del modello indicando la percentuale di corretti classificati rispetto all'insieme

$$acc = \frac{TP + TN}{TP + NP + TN + FN}$$

La **Specificità** è un parametro che indica la proporzione di negativi correttamente classificati

$$spec = \frac{TN}{FP + TN}$$

Dalla Specificità si può ricavare anche il suo complementare definito come **False Positive Rate (FPV)**

$$FPV = 1 - spec$$

Questo valore esprime la proporzione di negativi erroneamente classificati come positivi rispetto all'insieme.

La **Sensibilità (Recall)** è un parametro che indica la proporzione di positivi correttamente classificati

$$sens = \frac{TP}{TP + FN}$$

Negative predicted value (NPV), descrive la probabilità di un elemento di essere negativo se classificato negativo

$$NPV = \frac{TN}{TN + FN}$$

Positive predicted value (PPV), descrive la probabilità di un elemento di essere positivo se classificato positivo

$$PPV = \frac{TP}{TP + FP}$$

3.5.2 Curva ROC

La curva ROC (Receiver Operating Characteristic) è un grafico utilizzato per valutare le prestazioni di un modello di classificazione binaria. Questa curva è una rappresentazione grafica della relazione tra la sensibilità e FPR di un classificatore a diversi livelli di soglia.

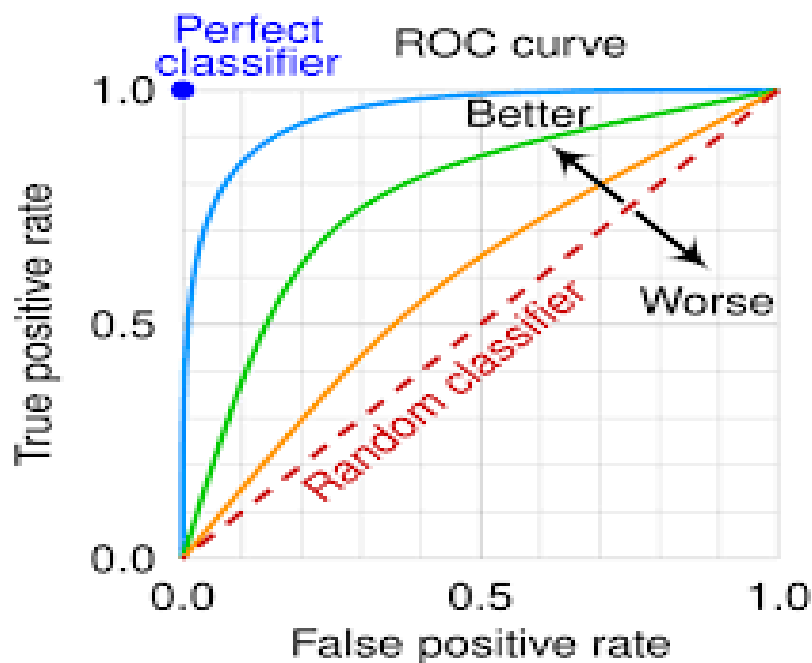


Figura 3.2: Rappresentazione di una Curva Roc

Da questa curva si ricava un altro valore chiamato **AUC (Area Under the Curve)**. Esso determina l'area sotto la curva ROC. In un modello ideale che non fa nessun errore il valore dell'AUC è pari a 1. Per valori compresi tra 1 e 0.5 il classificatore è in grado di discriminare le due classi, più ci si avvicina all'1 e migliore sarà la performance del modello, viceversa se il valore si avvicina a 0.5.

Un valore $AUC = 0.5$ indica un classificatore casuale, suggerendo che non è in grado di distinguere le classi. Un $AUC < 0.5$ rappresenta un modello con correlazione inversa tra predizione e valore reale.

Risultati

Le configurazioni utilizzate presentano quasi tutte le stesse performance, con un'accuratezza che oscilla tra il 55% e il 64%. Le uniche due configurazioni che presentano un'accuratezza migliore sono le configurazioni 5 e 14.

configurazioni CNN delle prove 5 e 14

	Prova 5	Prova 14
Input layer		
Conv2D	8	32
Conv2D	8	32
Conv2D	/	32
Conv2D	/	32
Conv2D	/	32
Fully Connected	8	32
Fully Connected	2	2

A parità di accuratezza per decidere quale modello fosse il migliore è stata considerata la sensibilità della rete, in quanto si è preferito un modello che “sbagliando” individuasse una lesione non grave come clinicamente significativa, piuttosto che il contrario. Tra le varie configurazioni provate la migliore quindi in termini di accuratezza e sensibilità è stata la n. 5, la quale ha raggiunto un'accuratezza del 68,18% e una sensibilità del 81,82%, inoltre, presenta un valore AUC di 0,68.

Queste sono le confusion matrix rispettivamente del training set e del validation set, del fold 1, dove per classe 0 si intende lesione con GG=2 e per classe 1 GG>2.

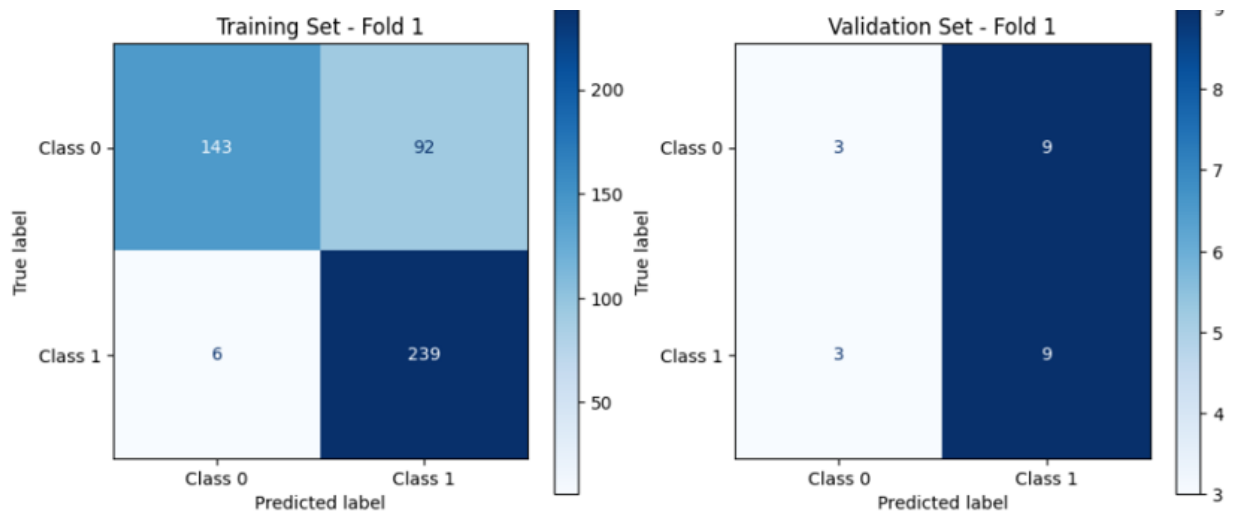


Figura 4.1: Confusion Matrix del Training set (a sinistra) e del Validation set (a destra), per la configurazione n. 5

Da queste CM si possono ricavare i valori esplicitati in precedenza quali:

Tabella 1. Valori Confusion matrix del training set

ACC	SPEC	SENS	PPV	NPV
79,58%	60,85%	97,55%	72,21%	95,97%

Tabella 2. Valori Confusion matrix del validation set

ACC	SPEC	SENS	PPV	NPV
50,00%	25,00%	75,00%	50,00%	50,00%

Mentre la Confusion matrix del test set è

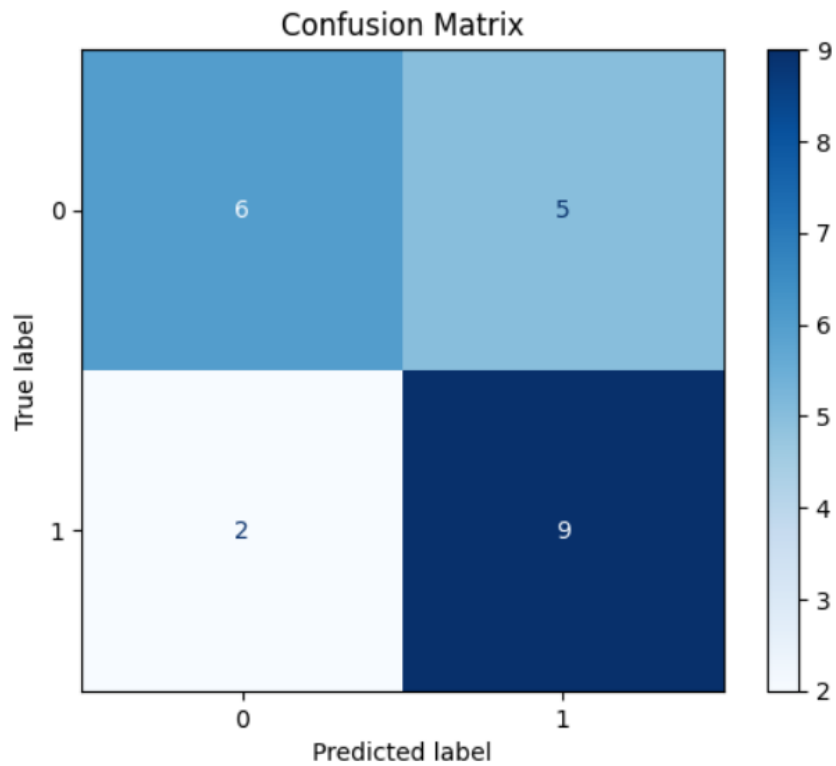


Figura 4.2: Figura 5.1: Confusion Matrix del Test set, per la configurazione n. 5

Tabella 3. Valori Confusion matrix del test set

ACC	SPEC	SENS	PPV	NPV
68,18%	54,55%	81,82%	64,29%	75,00%

Oltre alle confusion matrix e ai vari valori ricavabili è stata altresì costruita la curva ROC del test set

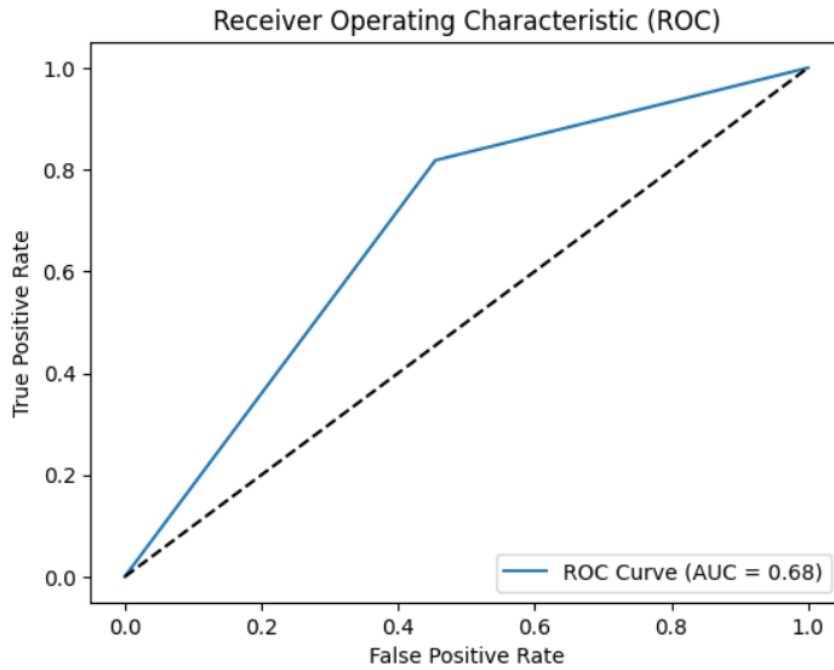


Figura 4.3: Curva ROC della configurazione n. 5

Modificando i parametri di questa rete come batch size e learning rate la performance del modello è sempre stata più bassa rispetto a quella della rete originale, l'unica con prestazioni simili è quella che presenta un ulteriore layer completamente connesso.

La confusion matrix della rete modificata è

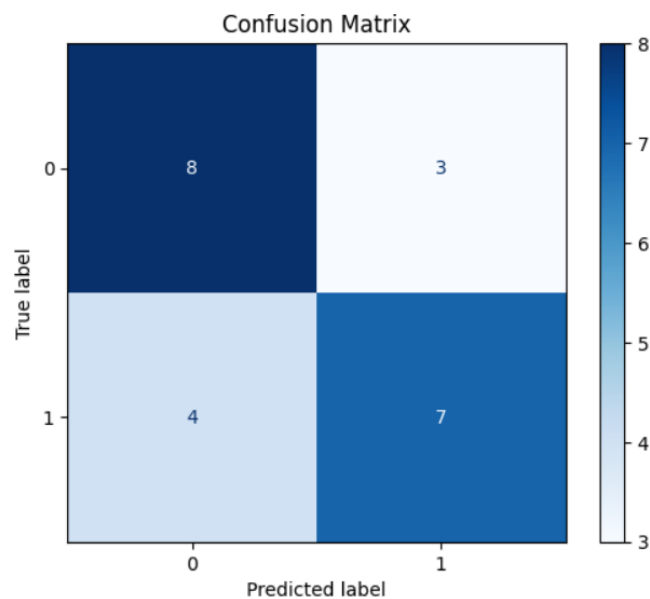


Figura 4.4: Confusion Matrix del Test set, per la configurazione n. 5 modificata

Tabella 4. Valori Confusion matrix del test set per la configurazione n. 5 modificata.

ACC	SPEC	SENS	PPV	NPV
68,18%	72,73%	63,64%	66,67%	70,00%

Con la seguente curva ROC

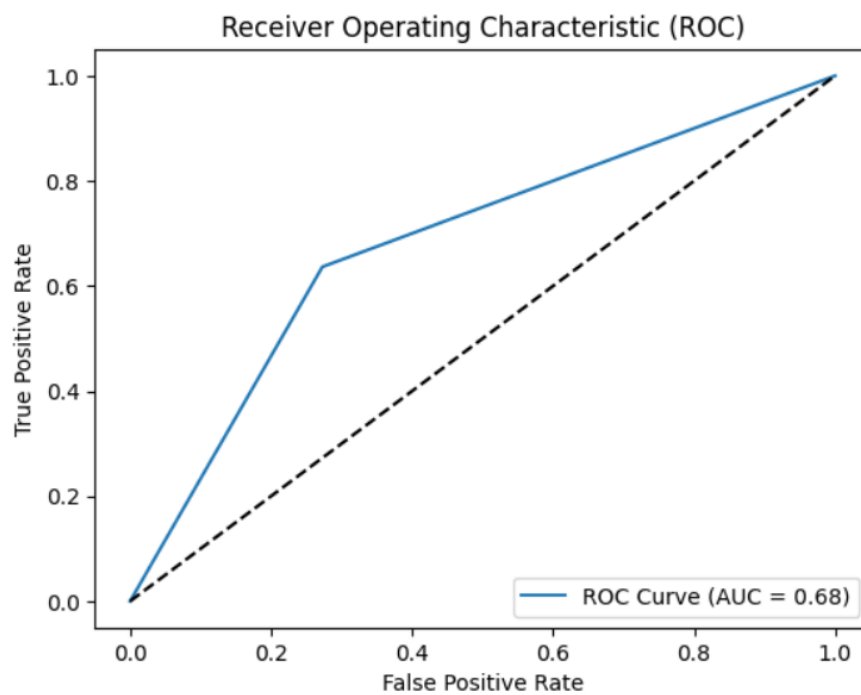


Figura 4.5: Curva ROC della configurazione n. 5 modificata.

I risultati migliori sono stati ottenuti però combinando le due reti più performanti, questo perché combinare più reti o modelli in un ensemble porta ad una percentuale di accuratezza più elevata perché riduce la varianza e il rumore, migliorando la capacità di generalizzazione. Ogni modello può avere imperfezioni e punti di forza unici; quindi, la loro combinazione permette di compensare errori individuali e catturare diversi pattern nei dati. Questa diversità rende l'ensemble più robusto e accurato rispetto a un singolo modello. Per combinare l'output delle due reti sono state utilizzate due tecniche, la media e la valutazione maggioritaria. Con la prima tecnica è stata calcolata la media con i valori di output delle due reti per ogni singola predizione. Sulla base del valore da essa ricavato è stata impostata una soglia di cut off pari a 0.3, per cui se la media delle predizioni è inferiore alla soglia la classe predetta è 0, viceversa la classe predetta è 1.

La confusion matrix del test set

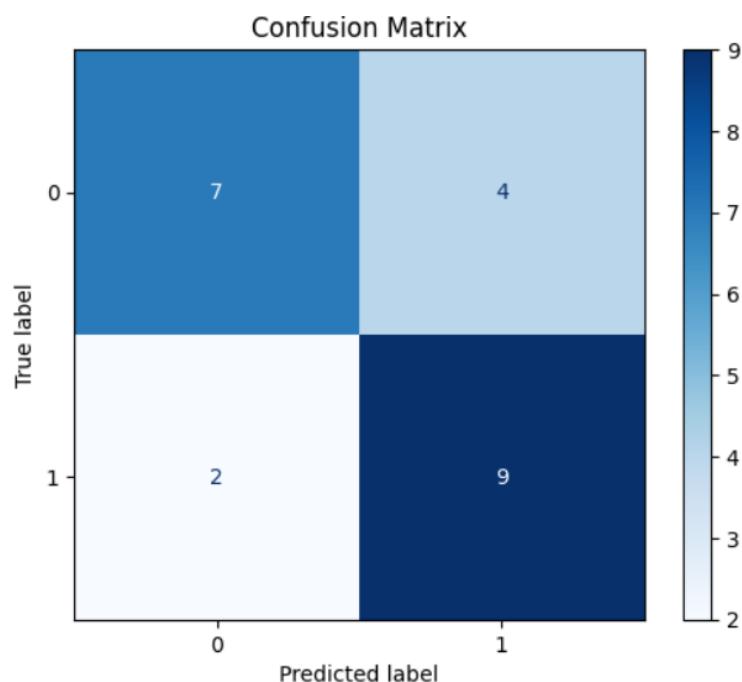


Figura 4.6: Confusion Matrix del test set delle reti combinate con la tecnica della media.

Tabella 5. Valori Confusion matrix del test set delle reti combinate con la tecnica della media.

ACC	SPEC	SENS	PPV	NPV
72,73%	81,82%	63,64%	69,23%	77,78%

Curva ROC del test set

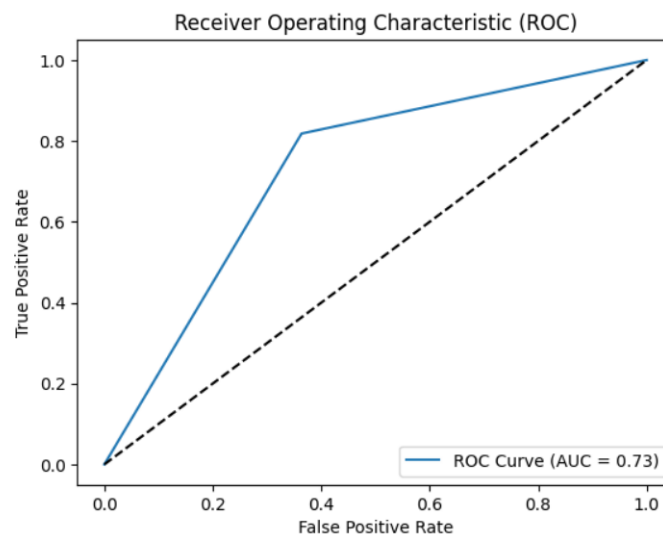


Figura 4.7: Curva ROC delle reti combinate con la tecnica della media.

La seconda tecnica utilizzata consiste nel prendere come output finale la classe predetta dalla maggioranza dei modelli. Se i due modelli classificano diversamente si può impostare la soglia di classificazione finale o al 50%, in questo caso basta solo

una rete che ha come output 1 per avere lo stesso output finale. Se invece viene impostata la soglia al 100%, per avere come output 1 entrambe le reti devono aver classificato 1. All'interno del progetto, per privilegiare un margine di errore in eccesso, si è scelto di impostare la soglia al 50%.

I risultati sono

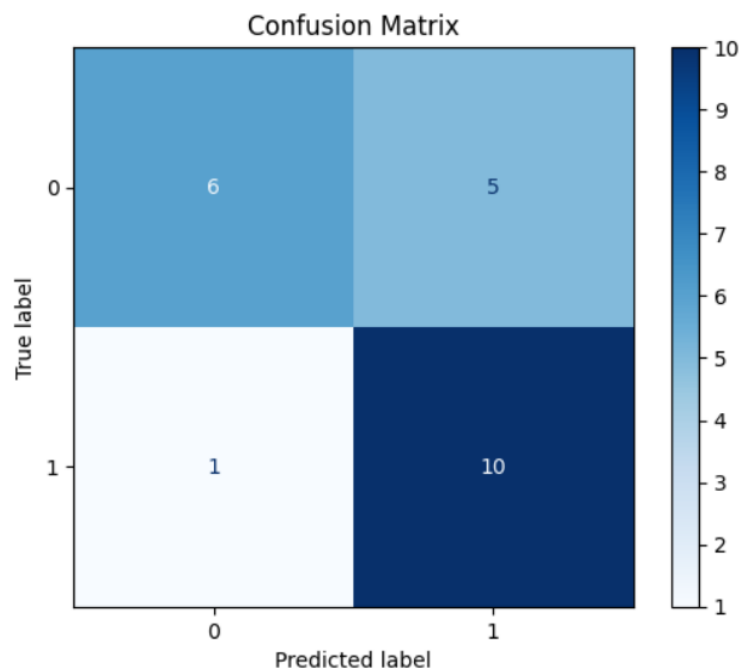


Figura 4.8: Confusion Matrix del test set delle reti combinate con la tecnica della valutazione maggioritaria.

Tabella 6. Valori Confusion matrix del test set delle reti combinate con la tecnica della valutazione maggioritaria.

ACC	SPEC	SENS	PPV	NPV
72,73%	91,92%	54,55%	85,71%	66,67%

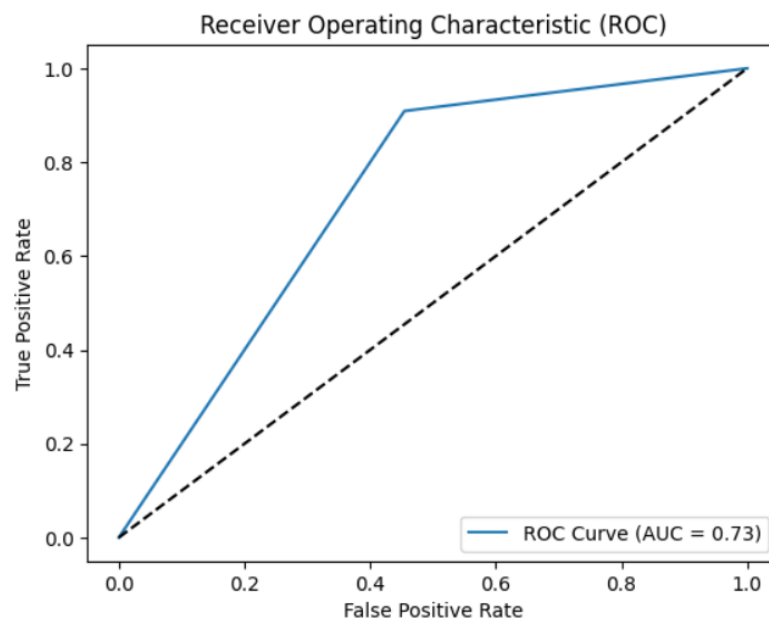


Figura 4.9: Curva ROC delle reti combinate con la tecnica della valutazione maggioritaria.

Grazie all'utilizzo di queste tecniche è stato possibile migliorare la performance delle singole architetture, portando l'accuratezza a valori di 72,73% e una specificità del 81,82% con la prima tecnica e arrivando a 91,82% con la seconda.

Conclusioni

Le conclusioni di questa tesi evidenziano il potenziale significativo delle reti neurali convoluzionali applicate alla diagnosi del tumore alla prostata. Attraverso la combinazione di tecniche di pre-processing avanzate, come la registrazione non rigida e la normalizzazione delle immagini MRI, è stato possibile addestrare modelli di deep learning capaci di distinguere con successo tra tumori clinicamente significativi e meno pericolosi. La migliore configurazione del modello ha raggiunto un'accuratezza del 68,18% e una sensibilità dell'81,82%, mostrando un buon bilanciamento tra precisione e identificazione dei casi critici. Con la combinazione delle due migliori reti si è giunti ad un valore di accuratezza più elevato, pari a 71,72% e una sensibilità del 91,92%. I risultati così ottenuti confermano l'efficacia del deep learning nel supportare i processi diagnostici, riducendo il margine d'errore e offrendo un valido aiuto ai radiologi. Questo studio contribuisce all'avanzamento delle applicazioni dell'intelligenza artificiale in medicina, aprendo la strada a sistemi diagnostici sempre più robusti e integrati nella pratica clinica. L'adozione di tecniche di ensemble, come la combinazione di reti multiple, ha dimostrato di migliorare ulteriormente le performance, sottolineando l'importanza della diversificazione dei modelli per ottenere una maggiore affidabilità. Il lavoro suggerisce che, se integrati con set di dati più ampi e con ulteriori ottimizzazioni, questi algoritmi potrebbero diventare strumenti standard nella classificazione dell'aggressività tumorale. Ulteriori sviluppi potrebbero includere l'implementazione di tecniche di deep learning più avanzate, come i modelli transformer, e l'analisi di nuove modalità di imaging per migliorare ulteriormente la precisione diagnostica. In sintesi, questa tesi sottolinea il ruolo cruciale dell'apprendimento profondo nella diagnosi medica e getta le basi per future applicazioni volte a migliorare l'efficienza e la qualità delle cure oncologiche.

Bibliografia

1. Singh O, Bolla SR. "Anatomy, Abdomen and Pelvis, Prostate". In: Stat Pearls [Internet]. Jan. 2020. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK540987/>;
2. Ittmann M. "Anatomy and histology of the human and murine prostate". Cold Spring Harbor Perspectives in Medicine [Internet]. 2018. Available from: <https://doi.org/10.1101/cshperspect.a030346>;
3. Aaron, LaTayia; Franco, Omar; Hayward, Simon W. "Review of Prostate Anatomy and Embryology and the Etiology of BPH." Urologic Clinics of North America, 2016. DOI: 10.1016/j.ucl.2016.04.012;
4. Toivanen, Roxanne; Shen, Michael M. "Prostate Organogenesis: Tissue Induction, Hormonal Regulation and Cell Type Specification." Development, 2017. DOI: 10.1242/dev.148270;
5. Ferlay, Jacques; Ervik, Mathilde; Lam, Florence; et al. "GLOBOCAN 2022: Global Cancer Observatory." International Agency for Research on Cancer (IARC). Available at: <https://gco.iarc.who.int>;
6. Rawla, Prashanth. "Epidemiology of Prostate Cancer." World Journal of Oncology, 2019;10(2):63-89. DOI: [10.14740/wjon1191](https://doi.org/10.14740/wjon1191);
7. <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-della-prostata>
8. Jones D, Friend C, Dreher A, Allgar V, Macleod U. "The diagnostic test accuracy of rectal examination for prostate cancer diagnosis in symptomatic patients: a systematic review". BMC Fam Pract. 2018; 19:79. DOI:10.1186/s12875-018-0765-y;
9. Soares SCM, Cancela MC, Migowski A, Souza DLB. "Digital rectal examination and its associated factors in the early detection of prostate cancer: a cross-sectional population-based study." BMC Public Health. 2019; 19:1573. DOI:10.1186/s12889-019-7946-z;
10. Sekhoacha M, Riet K, Motloug P, Gumenku L, Adegoke A, Mashele S. "Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches." Molecules. 2022;27(5730):1-33. DOI:10.3390/molecules27175730;

11. Humphrey PA. "Histopathology of Prostate Cancer." *Cold Spring Harb Perspect Med.* 2017;7: a030411. DOI:10.1101/cshperspect.a030411;
12. Humphrey PA. "Gleason grading and prognostic factors in carcinoma of the prostate." *Mod Pathol.* 2004; 17:292–306. DOI:10.1038/modpathol.3800054
13. <https://www.ncbi.nlm.nih.gov/books/NBK553178/>
14. Brawley S, Mohan R, Nein CD. "Localized Prostate Cancer: Treatment Options." *Am Fam Physician.* 2018;97(12):798-805;
15. Evans AJ. "Treatment effects in prostate cancer." *Mod Pathol.* 2018; 31: S110–S121. DOI:10.1038/modpathol.2017.158;
16. Teo MY, Rathkopf DE, Kantoff P. "Treatment of Advanced Prostate Cancer." *Annu Rev Med.* 2019; 70:479–499. doi:10.1146/annurev-med-051517-011947;
17. Wasim S, Lee SY, Kim J. "Complexities of Prostate Cancer." *Int J Mol Sci.* 2022;23(14257). DOI:10.3390/ijms232214257;
18. Tempany CMC, Carroll PR, Leapman MS, et al. "The role of magnetic resonance imaging in prostate cancer." *UpToDate.* 2023. Available from: <https://www.wolterskluwer.com/en/solutions/uptodate>;
19. Scialpi M, Martorana E, Scialpi P, et al. "MRI apparent diffusion coefficient (ADC): A biomarker for prostate cancer after radiation therapy." *Turk J Urol.* 2021;47(6):448-451. DOI: 10.5152/tud.2021.21274;
20. Würnschimmel C, Chandrasekar T, Hahn L, et al. "MRI as a screening tool for prostate cancer: current evidence and future challenges." *World J Urol.* 2023; 41:921–928. DOI: 10.1007/s00345-022-03947-y;
21. Fusco R, Sansone M, Granata V, et al. "A systematic review on multiparametric MR imaging in prostate cancer detection." *Infect Agents Cancer.* 2017; 12:57. DOI:10.1186/s13027-017-0168-z;
22. Barentsz JO, Weinreb JC, Verma S, et al. "Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use." *Eur Urol.* 2016;69(1):41–49. DOI:10.1016/j.eururo.2015.08.038;
23. Xu L, Zhang G, Shi B, et al. "Comparison of biparametric and multiparametric MRI in the diagnosis of prostate cancer." *Cancer Imaging.* 2019; 19:90. doi:10.1186/s40644-019-0274-9;
24. LeCun Y, Bengio Y, Hinton G. "Deep learning." *Nature.* 2015; 521(7553): 436-444. DOI:10.1038/nature14539;
25. <https://www.ibm.com/it-it/topics/supervised-learning>

26. Machine learning and deep learning Christian Janiesch¹ & Patrick Zschech² & Kai Heinrich³ Janiesch C, Zschech P, Heinrich K. "Machine learning and deep learning: Fundamentals." *Electron Markets*. 2021; 31: 685-695.
DOI:10.1007/s12525-021-00475-2;
27. Jiang Y, Yang M, Wang S, Li X, Sun Y. "Emerging role of deep learning-based artificial intelligence in tumor pathology." *Cancer Commun (Lond)*. 2020;40(3):154-166. doi:10.1002/cac2.12012;
28. Jha GK. "Artificial neural networks." *Indian Agricultural Research Institute*. 2021:1-13
29. Anderson D, McNeill G. "Artificial neural networks technology." *Rome Laboratory*. 1992:1-71
30. Yang GR, Wang XJ. "Artificial neural networks for neuroscientists: A primer." *Neuron*. 2020;107(6):1048-1071. doi:10.1016/j.neuron.2020.09.005;
31. Nwankpa CE, Ijomah W, Gachagan A, Marshall S. "Activation functions: Comparison of trends in practice and research for deep learning." *arXiv preprint*. 2018. arXiv:1811.03378;
32. Dubey SR, Singh SK, Chaudhuri BB. "Activation functions in deep learning: A comprehensive survey and benchmark." *Neurocomputing*. 2022; 470:1-24. doi:10.1016/j.neucom.2021.08.059;
33. Kılıçarslan S, Adem K, Çelik M. "An overview of the activation functions used in deep learning algorithms." *J New Results Sci*. 2021;10(3):75-88. doi:10.54187/jnrs.1011739;
34. *Journal of New Results in Science* <https://dergipark.org.tr/en/pub/jnrs>
35. <https://deepai.org/machine-learning-glossary-and-terms/feed-forward-neural-network>
36. <https://builtin.com/data-science/feedforward-neural-network-intro>
37. https://en.wikipedia.org/wiki/Feedforward_neural_network#cite_note-10
38. Svozil D, Kvasnička V, Pospíchal J. "Introduction to multi-layer feed-forward neural networks." *Chemom Intell Lab Syst*. 1997;39:43-62. doi:10.1016/S0169-7439(97)00061-0;
39. O'Shea K, Nash R. "An introduction to convolutional neural networks." *ArXiv Preprint*. 2015. arXiv:1511.08458

40. Yamashita R, Nishio M, Do RK, Togashi K. "Convolutional neural networks: An overview and application in radiology." *Insights Imaging*. 2018;9:611-629. doi:10.1007/s13244-018-0639-9;
41. <https://www.ibm.com/topics/convolutional-neural-networks>
42. Li Z, Yang W, Peng S, Liu F. "A survey of convolutional neural networks: Analysis, applications, and prospects." *ArXiv Preprint*. 2018. arXiv:1801.01078;
43. Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S. "Recent advances in recurrent neural networks." *ArXiv Preprint*. 2018. arXiv:1801.01078;
44. <https://www.ibm.com/topics/recurrent-neural-networks>
45. <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>
46. Creswell A, White T, Dumoulin V, et al. "Generative adversarial networks: An overview." *IEEE Signal Process Mag*. 2017;35(1):53-65. doi:10.1109/MSP.2017.2765202;
47. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. "Generative adversarial nets." *Adv Neural Inf Process Syst*. 2014;27:2672-2680;
48. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. "Generative adversarial networks." *Commun ACM*. 2020;63(11):139-144. doi:10.1145/3422622;
49. Wang KF, Gou C, Duan YJ, et al. "Generative adversarial networks: Introduction and outlook." *IEEE/CAA J Autom Sinica*. 2017;4(4):588-598. doi:10.1109/JAS.2017.7510583;
50. Yeasmin MN, Al Amin M, Joti TJ, Aung Z, Azim MA. "Advances of AI in image-based computer-aided diagnosis: A review." *Array*. 2024;23:100357. doi:10.1016/j.array.2024.100357;
51. Doi K. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential." *Comput Med Imaging Graph*. 2007;31(4-5):198-211. doi:10.1016/j.compmedimag.2007.01.002;
52. Petrick N, Sahiner B, Armato SG III, et al. "Evaluation of computer-aided detection and diagnosis systems." *Med Phys*. 2013;40(8):087001. doi:10.1118/1.4816310;
53. Gasmi A. "Deep learning and health informatics for smart monitoring and diagnosis." *Unpublished manuscript*. 2020;

54. Ronneberger O, Fischer P, Brox T. "U-Net: Convolutional networks for biomedical image segmentation." LNCS. 2015;9351:234-241. doi:10.1007/978-3-319-24574-4_28;
55. Wu C, Montagne S, Hamzaoui D, et al. "Automatic segmentation of prostate zonal anatomy on MRI: A systematic review of the literature." *Insights Imaging*. 2022;13:202. doi:10.1186/s13244-022-01340-2;
56. Badiezadeh A, Malekmohammadi A, Mirhassani SM, et al. "Segmentation strategies in deep learning for prostate cancer diagnosis: A comparative study of Mamba, SAM, and YOLO." Unpublished manuscript. 2024;
57. Talaat FM, El-Sappagh S, Alnowaiser K, Hassan E. "Improved prostate cancer diagnosis using a modified ResNet50-based deep learning architecture." *BMC Med Inform Decis Mak*. 2024;24:23. doi:10.1186/s12911-024-02419-0;
58. Campanella G, Hanna MG, Geneslaw L, et al. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images." *Nat Med*. 2019;25(8):1301-1309. doi:10.1038/s41591-019-0508-1;
59. Müller D, Meyer P, Rentschler L, et al. "DeepGleason: A system for automated Gleason grading of prostate cancer using deep neural networks." Unpublished manuscript. 2024;
60. Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE. "Deep learning-based Gleason grading of prostate cancer from histopathology images—Role of multiscale decision aggregation and data augmentation." *IEEE J Biomed Health Inform*. 2020;24(5):1413–1425. doi:10.1109/JBHI.2019.2944643.

Ringraziamenti

In primo luogo, un sentito ringraziamento va al mio relatore, Valentina Giannini, per la sua guida, disponibilità e per l'attenzione che ha dedicato al mio lavoro. La sua competenza e il suo incoraggiamento sono stati fondamentali nel superare le difficoltà di questo percorso.

Un ringraziamento speciale ad Anna, che con il suo amore, la sua pazienza e la sua costante vicinanza è stata la mia forza nei momenti più difficili. Grazie per aver creduto in me, per i tuoi incoraggiamenti e per esserti presa cura di me anche quando ero troppo immerso nel lavoro per farlo da solo. Questo traguardo è anche il frutto del tuo amore e del tuo sostegno.

Un grazie speciale alla mia famiglia, soprattutto a mia mamma, che mi ha sempre sostenuto con pazienza e affetto, credendo in me anche nei momenti più impegnativi.

Grazie ai miei amici di sempre, Mario e Xhulio, per avermi sostenuto con costanza e per essere sempre al mio fianco.

Infine, ringrazio tutte le persone che, direttamente o indirettamente, hanno contribuito alla realizzazione di questa tesi. Il vostro aiuto e supporto sono stati inestimabili.

Con affetto e gratitudine,

Francesco Spinola