



# Computational Intelligence and Deep Learning

Developed by  
Arancio Febbo Salvatore



## Contents

Introduzione.....	4
Methods and experiments .....	5
CNN from scratch .....	5
Experiment 1: CNN basic.....	5
Experiment 2: CNN more layer .....	7
Pre -trained VGG16 .....	8
Experiment 1: Basic .....	8
Experiment 2: Incorporating Dropout.....	8
Experiment 3: Global average pooling .....	9
Experiment 4: last block (block5_conv3) with experiment1 .....	9
Experiment 4bis last block (block5_conv3) with experiment2 .....	9
Experiment 5: last block (block5_conv2, block5_conv3, block5_conv1) with experiment1 .....	10
Experiment 5bis: last block (block5_conv2, block5_conv3, block5_conv1) with experiment2 .....	10
Experiment 6: two block with experiment2 .....	11
Experiment 7: Eliminate "block5_conv1, block5_conv2, block5_conv3" Layers .....	11
Experiment 8: block4 .....	12
Conclusioni parziali sulla rete pre-trainata VGG16 .....	12
Esperimenti con CLAHE .....	13
Explainability .....	16
Conclusioni .....	17

## Introduzione

L'identificazione precoce dei tumori al seno è un problema di grande importanza per la salute delle donne in tutto il mondo. La tecnologia di imaging del seno, come la mammografia, è uno strumento diagnostico fondamentale per la rilevazione del tumore al seno, ma l'interpretazione delle immagini può essere difficile e soggettiva. Inoltre, la mammografia può essere costosa e non facilmente accessibile in alcune parti del mondo.

Per ovviare a questi problemi, l'analisi delle immagini tramite algoritmi di apprendimento automatico rappresenta un'opzione sempre più interessante. In particolare, la classificazione di immagini di patch mammarie tramite deep learning è diventata un'area di ricerca attiva negli ultimi anni.

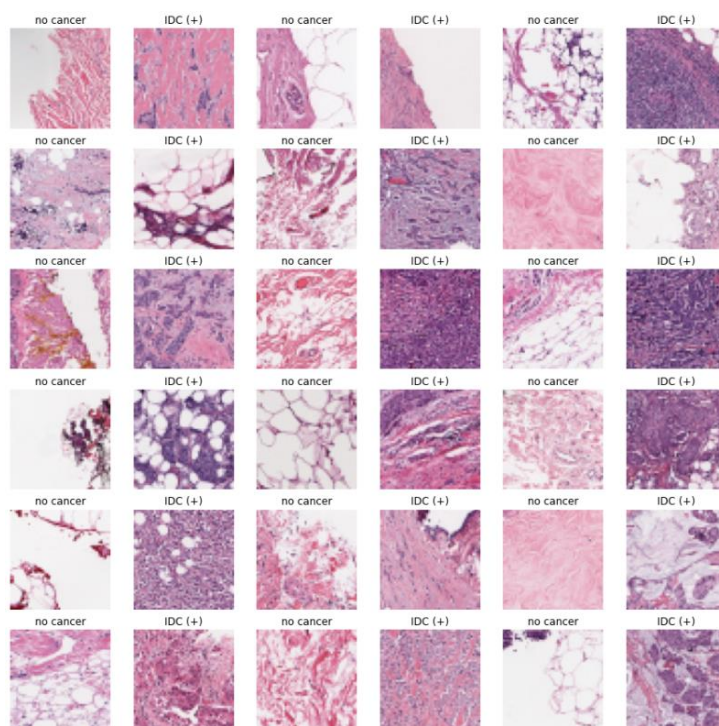
In questo progetto, ci concentriamo sulla creazione di un modello di deep learning in grado di identificare le patch mammarie contenenti tumori al seno. Il nostro dataset è composto da immagini di patch mammarie prese da diverse pazienti.

Le patch sono etichettate come 0 se non contengono un tumore e come 1 se contengono un tumore. Ogni patch è identificata dal suo nome, che contiene le coordinate per consentire la ricostruzione dell'immagine completa.

Il nostro obiettivo è quello di utilizzare il machine learning per creare un modello di classificazione delle patch mammarie in modo che possa essere utilizzato per identificare i tumori al seno in modo rapido ed efficiente. Il nostro modello sarà in grado di analizzare le patch mammarie e determinare se contengono un tumore al seno, con una precisione maggiore rispetto alla valutazione umana.

Il nostro lavoro può portare a una maggiore precisione nella diagnosi del tumore al seno, e a un maggiore accesso alle cure e alla prevenzione per le donne in tutto il mondo. Inoltre, il nostro progetto può rappresentare una base per futuri studi nel campo della diagnostica mammografica basata su intelligenza artificiale.

A seguire saranno presentate delle patch esplicative in cui le colonne dispari rappresentano le parti del seno senza tumore e quelle pari le parti con tumore.



*Figura 1: patches with cancer and without cancer*

## Methods and experiments

Il capitolo "Methods and experiments" di questo report descrive la metodologia utilizzata per valutare l'efficacia di modelli di Deep Learning nella classificazione di immagini. In particolare, verranno presentati i risultati di due differenti approcci: uno basato sull'addestramento da zero di una rete neurale convoluzionale, e l'altro basato sull'utilizzo di una rete pre-addestrata VGG16, nota per le sue prestazioni in campo di classificazione di immagini.

Per entrambi gli approcci, sono stati utilizzati gli stessi dataset di immagini, e sono state condotte diverse fasi di addestramento e testing. Nel primo caso, la rete è stata addestrata partendo da zero, utilizzando un'architettura customizzata. Nel secondo caso, invece, la rete VGG16 è stata utilizzata in due modalità: come estrattore di features, con i layer convoluzionali già addestrati, oppure come modello completo, con l'intera architettura pre-addestrata.

Verranno presentati i risultati delle sperimentazioni, con una valutazione delle prestazioni dei modelli in termini di accuratezza nella classificazione delle immagini, e un'analisi comparativa tra i due approcci.

Questo capitolo fornisce una panoramica completa dei metodi e degli esperimenti condotti, con l'obiettivo di fornire una solida base per la valutazione delle prestazioni dei modelli di Deep Learning utilizzati nella classificazione di immagini.

### CNN from scratch

#### Experiment 1: CNN basic

L'esperimento 1 sembra essere un buon inizio per il riconoscimento dei tumori al seno tramite deep learning. Il modello è basato su una rete neurale convoluzionale (CNN) con una struttura piuttosto semplice, composta da un layer di convoluzione con 64 filtri e una dimensione di kernel di 3x3, seguito da un layer di max pooling con una dimensione di pool di 2x2.

Il modello viene poi regolarizzato tramite un layer di dropout, che riduce il rischio di overfitting. Infine, ci sono due layer densi, uno con 64 neuroni e uno di output con 2 neuroni e una funzione di attivazione softmax.

	precision	recall	f1-score
0	0.9041	0.8634	0.8833
1	0.6906	0.7689	0.7277
accuracy	0.8366		

In termini di prestazioni, il modello raggiunge un'accuratezza complessiva del 83,66%, con una precisione del 69,06% per la classe 1 e del 90,41% per la classe 0. Il recall, che indica la capacità del modello di identificare correttamente i campioni di una classe, è del 76,89% per la classe 1 e dell'86,34% per la classe 0.

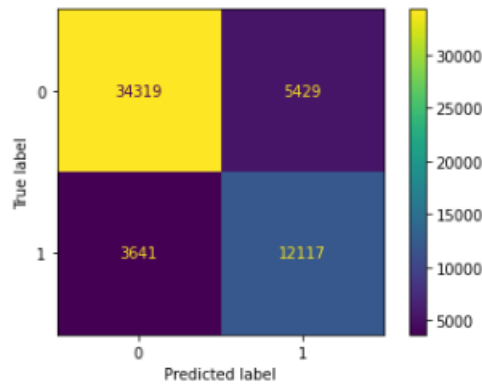


Figura 2: Confusion matrix

La matrice di confusione indica che il modello ha fatto un buon lavoro nell'identificare correttamente i campioni della classe 0, ma ha avuto qualche difficoltà nell'identificare correttamente i campioni della classe 1.

In generale, questo esperimento rappresenta un'ottima base per l'elaborazione di ulteriori modelli di deep learning per il riconoscimento dei tumori al seno, con un buon equilibrio tra semplicità e prestazioni. Tuttavia, ci sono ancora alcune aree in cui il modello può essere migliorato, in particolare nella capacità di identificare correttamente i campioni della classe 1.

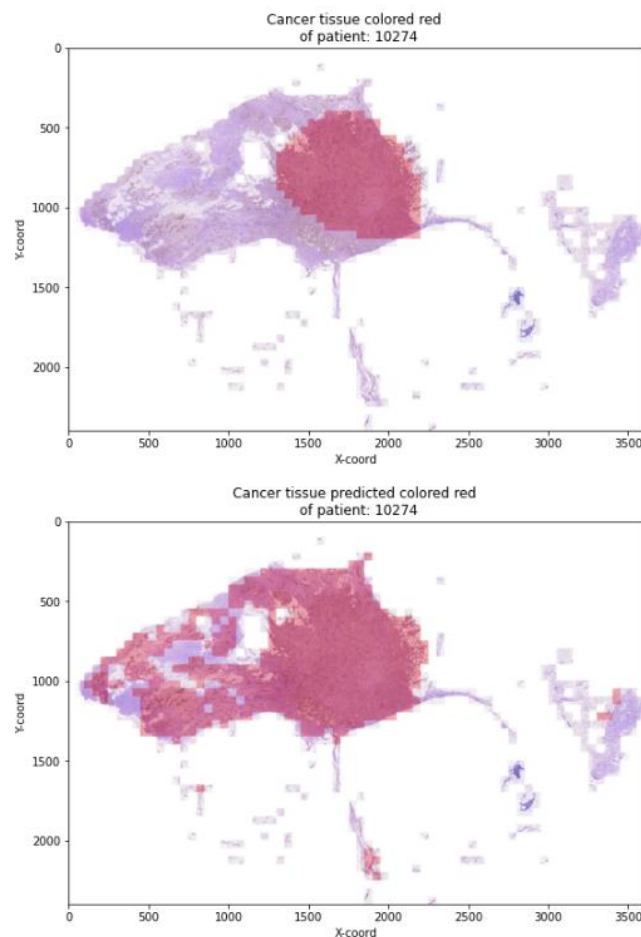


Figura 3: Esperimento1, differenza tra immagine reale e predetta

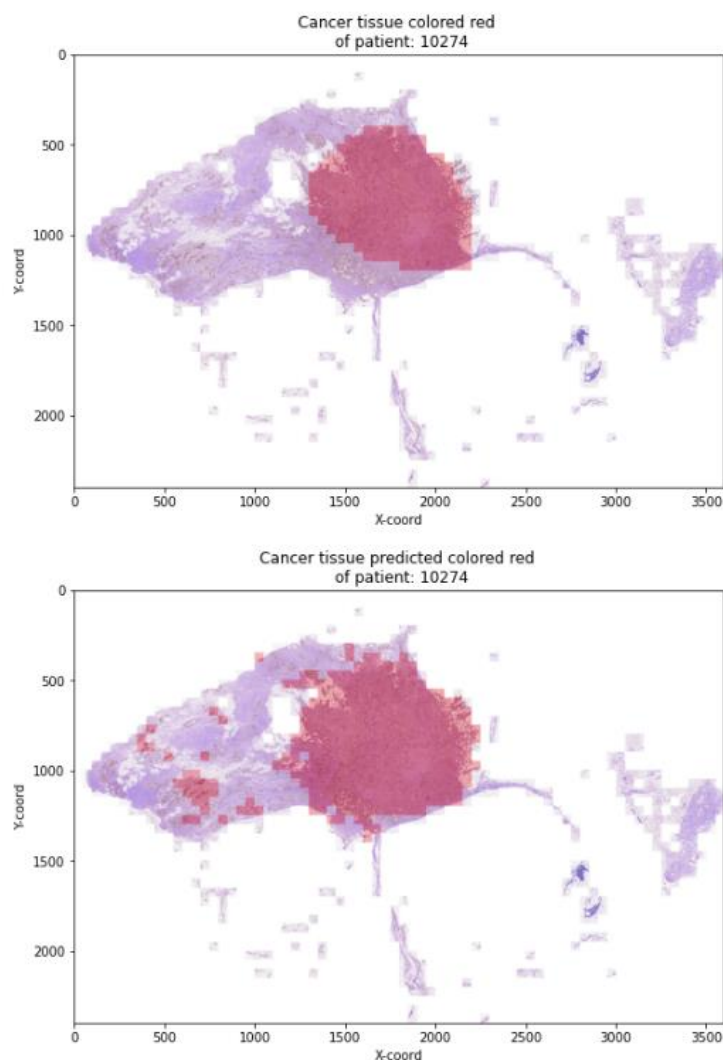
### Experiment 2: CNN more layer

L'Experiment 2 utilizza una CNN con più strati rispetto all'esperimento 1. In particolare, questo modello ha una serie di strati Conv2D e MaxPooling2D, seguiti da alcuni strati completamente connessi. Il modello utilizza anche tecniche di regolarizzazione, come il dropout e la normalizzazione del batch.

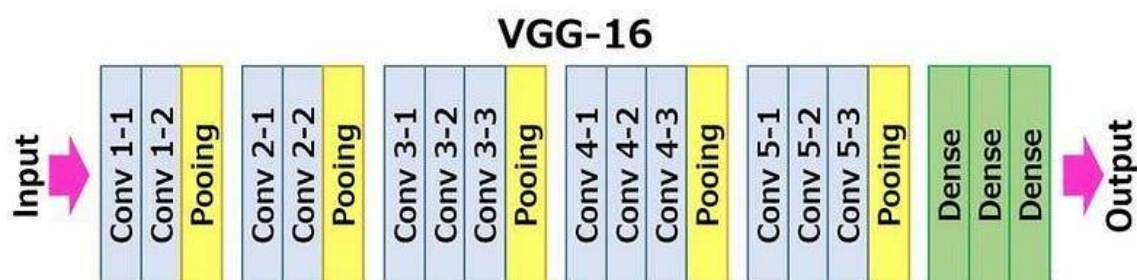
	precision	recall	f1-score
<b>0</b>	0.9466	0.844	0.8924
<b>1</b>	0.691	0.8799	0.7741
<b>accuracy</b>	0.8542		

Il modello ha mostrato un miglioramento dell'accuratezza rispetto all'esperimento 1, con un'accuratezza complessiva del 85,42%. La precisione e il recall per le immagini di classe 0 (senza tumore) sono aumentati rispetto all'Experiment 1, mentre per le immagini di classe 1 (con tumore) la precisione è diminuita ma il recall è aumentato, indicando una maggiore capacità del modello di individuare i casi di tumore.

In generale, questo modello sembra essere una buona scelta per il riconoscimento delle immagini di tumore al seno, grazie alla sua capacità di rilevare con precisione sia i casi positivi che quelli negativi.



## Pre-trained VGG16



Nella rete preaddestrata VGG16 sono stati eseguiti più esperimenti per valutare l'effetto di diversi livelli di complessità del modello.

### Experiment 1: Basic

Nel **primo esperimento**, si è utilizzata una configurazione di base del modello, con un layer di flatten e due layer fully connected. Il modello è stato addestrato per 20 epoche e si è ottenuta una precisione del 81,54%. La matrice di confusione mostra che il modello ha classificato correttamente la maggior parte delle immagini senza cancro al seno, ma ha ottenuto una precisione più bassa nella classificazione delle immagini contenenti un tumore al seno.

	precision	recall	f1-score
<b>0</b>	0.8773	0.863	0.8701
<b>1</b>	0.668	0.6955	0.6815
<b>accuracy</b>	0.8154		

### Experiment 2: Incorporating Dropout

Nel **secondo esperimento**, si è aggiunto un layer di dropout per ridurre l'overfitting del modello. Il modello addestrato ha ottenuto una precisione del 82%, leggermente superiore rispetto al modello di base. La matrice di confusione mostra che il modello ha classificato correttamente la maggior parte delle immagini senza cancro al seno, ma ha ottenuto una precisione leggermente più alta nella classificazione delle immagini contenenti un tumore al seno.

	precision	recall	f1-score
<b>0</b>	0.8881	0.8566	0.8721
<b>1</b>	0.668	0.7278	0.6966
<b>accuracy</b>	0.82		



### Experiment 3: Global average pooling

Nel **terzo esperimento**, si è utilizzato un layer di Global Average Pooling in sostituzione del layer di Flatten per ridurre il numero di parametri del modello. Il modello addestrato ha ottenuto una precisione del 82,11%. La matrice di confusione mostra che il modello ha classificato correttamente la maggior parte delle immagini senza cancro al seno, ma ha ottenuto una precisione leggermente più bassa nella classificazione delle immagini contenenti un tumore al seno rispetto al secondo esperimento, abbastanza trascurabile.

	precision	recall	f1-score
<b>0</b>	0.8869	0.8598	0.8731
<b>1</b>	0.6717	0.7234	0.6966
<b>accuracy</b>	0.8211		

### Experiment 4: last block (block5\_conv3) with experiment1

Nell'**esperimento 4**, è stato utilizzato il modello allenato nell'esperimento 1, con l'aggiunta di una modifica al livello del blocco 5 "block5\_conv3". In particolare, il livello è stato reso addestrabile impostando il parametro "trainable" su True. Tutti gli altri strati del modello sono stati lasciati come non addestrabili. Successivamente, il modello è stato compilato con una learning rate di  $1e-5$ .

	precision	recall	f1-score
<b>0</b>	0.9047	0.8625	0.8831
<b>1</b>	0.6897	0.7707	0.728
<b>accuracy</b>	0.8365		

Il report di classificazione mostra che il modello ha raggiunto un'accuratezza del 83,65%, con una precisione del 90,47% per la classe 0 (niente tumore) e del 68,97% per la classe 1 (con tumore). La recall è del 86,25% per la classe 0 e del 77,07% per la classe 1, mentre l'f1-score è rispettivamente del 88,31% e del 72,80%. Il valore di macro avg è del 80,55%, mentre il weighted avg è del 83,91%.

### Experiment 4bis last block (block5\_conv3) with experiment2

L'**esperimento 4bis** è simile all'esperimento 4, ma utilizza il modello allenato nell'esperimento 2.

	precision	recall	f1-score
<b>0</b>	0.9066	0.828	0.8655
<b>1</b>	0.6439	0.7847	0.7074
<b>accuracy</b>	0.8157		

Analizzando l'esperimento 4bis, notiamo che ha ottenuto un punteggio di precisione relativamente alto per la classe 0, indicando che è in grado di classificare correttamente la mancanza di tumore al seno nella maggior parte dei casi. Tuttavia, il punteggio di precisione per la classe 1 è piuttosto basso, indicando che il modello ha difficoltà a distinguere correttamente i frammenti di immagini di seni contenenti tumori.

### Experiment 5: last block (block5\_conv2, block5\_conv3, block5\_conv1) with experiment1

Nell'**esperimento 5**, l'intero blocco 5 è stato attivato e ricompilato il modello creato nell'esperimento 1, con un tasso di apprendimento di  $1e-5$ .

	precision	recall	f1-score
0	0.9126	0.869	0.8903
1	0.7051	0.7901	0.7452
accuracy	0.8466		

La matrice di confusione del modello addestrato ha mostrato 34540 veri negativi, 5208 falsi positivi, 3307 falsi negativi e 12451 veri positivi. Inoltre, il rapporto di classificazione ha mostrato un'accuratezza del 84,66%, con una precisione del 70,51% per la classe 1 e del 91,26% per la classe 0. Il richiamo è stato del 79,01% per la classe 1 e del 86,90% per la classe 0, con un punteggio F1 del 74,52% per la classe 1 e del 89,03% per la classe 0. Questi risultati mostrano che l'attivazione del blocco 5 e l'utilizzo del modello allenato nell'esperimento 1 ha portato a un miglioramento delle prestazioni del modello.

Dal report di classificazione possiamo notare che il modello ha un'accuratezza del 84.4%, il che significa che è in grado di classificare correttamente l'84.4% delle patch. La precisione per la classe 0 è alta (91.75%), il che significa che il modello è in grado di identificare correttamente la maggior parte delle patch senza tumore. La recall per la classe 1 è abbastanza elevata (80.52%), il che significa che il modello è in grado di individuare correttamente la maggior parte delle patch con tumore al seno. Tuttavia, la precisione per la classe 1 è abbastanza bassa (69.42%), il che significa che il modello commette molti falsi positivi, ovvero classifica erroneamente alcune patch senza tumore come se avessero un tumore. Questo potrebbe portare a interventi invasivi inutili e potenzialmente pericolosi per i pazienti.

### Experiment 5bis: last block (block5\_conv2, block5\_conv3, block5\_conv1) with experiment2

Confrontando i risultati dell'**esperimento 5bis** con quelli dell'esperimento 5, si nota che l'accuracy è leggermente superiore nell'esperimento 5bis (0.8440 rispetto a 0.8466), ma la differenza non è significativa. Tuttavia, l'esperimento 5bis sembra avere una maggiore precisione e recall per la classe 0 (patches senza tumore al seno), mentre l'esperimento 5 ha una maggiore precisione e recall per la classe 1 (patches con tumore al seno). In entrambi i casi, l'F1-score è piuttosto basso per la classe 1, suggerendo che il modello fatica a rilevare correttamente le patches con tumore al seno.

	precision	recall	f1-score
0	0.9175	0.8594	0.8875
1	0.6942	0.8052	0.7456
accuracy	0.844		

### Experiment 6: two block with experiment2

Il modello nell'**esperimento 6** viene addestrato sui blocchi 4 e 5 e ricompilato il modello creato nell'esperimento 2. In particolare, le convoluzioni 1, 2 e 3 di questi blocchi vengono addestrate insieme ai livelli completamente connessi del modello. Gli altri livelli della rete VGG16 vengono mantenuti non addestrabili.

	precision	recall	f1-score
<b>0</b>	0.9218	0.8842	0.9026
<b>1</b>	0.7352	0.8107	0.7711
<b>accuracy</b>	0.8633		

Il rapporto di classificazione riporta una precisione del 92.18% per la categoria 0 e del 73.52% per la categoria 1. La recall è del 88.42% per la categoria 0 e del 81.07% per la categoria 1. L'f1-score è del 90.26% per la categoria 0 e del 77.11% per la categoria 1. L'accuratezza del modello è del 86.33%.

In generale, l'esperimento 6 sembra avere prestazioni simili o leggermente migliori rispetto agli altri esperimenti già visti che utilizzano il modello VGG16 con alcuni dei suoi blocchi addestrati.

### Experiment 7: Eliminate "block5\_conv1, block5\_conv2, block5\_conv3" Layers

Nell'**esperimento 7**, è stato utilizzato un modello sequenziale con base convoluzionale VGG16 pre-addestrata. Sono stati congelati tutti i layer della rete, ad eccezione dei layer "block5\_conv1", "block5\_conv2" e "block5\_conv3" i quali sono stati direttamente rimossi e successivamente aggiunto uno strato di GlobalAveragePooling2D, che riduce la dimensione dell'output e lo rende utilizzabile come input per uno strato fully-connected. Successivamente, è stato addestrato un classificatore su top della rete utilizzando il set di dati di addestramento.

	precision	recall	f1-score
<b>0</b>	0.8983	0.8869	0.8925
<b>1</b>	0.7235	0.7466	0.7349
<b>accuracy</b>	0.847		

L'addestramento del classificatore ha prodotto un'accuratezza del 84,7%. La classe 0 ha una precisione del 89,83%, una recall del 88,69% e un f1-score del 89,25%, mentre la classe 1 ha una precisione del 72,35%, una recall del 74,66% e un f1-score del 73,49%. La matrice di confusione mostra che il classificatore ha predetto correttamente 35251 campioni appartenenti alla classe 0 e 11765 campioni appartenenti alla classe 1. Tuttavia, ha classificato erroneamente 4497 campioni della classe 0 come appartenenti alla classe 1 e 3993 campioni della classe 1 come appartenenti alla classe 0.

La recall della classe 0 è buona, mentre la recall della classe 1 è relativamente bassa rispetto ad alcuni degli altri esperimenti. La matrice di confusione mostra un certo grado di confusione tra le due classi, con un alto numero di campioni classificati erroneamente. Complessivamente, l'esperimento 7 sembra produrre risultati mediamente buoni.

#### Experiment 8: block4

Nell'esperimento 8, è stato preso il modello dell'esperimento 7 e resto addestrabile il blocco 4

	precision	recall	f1-score
<b>0</b>	0.9235	0.8799	0.9012
<b>1</b>	0.7293	0.8161	0.7703
<b>accuracy</b>	0.8618		

Siamo riusciti ad eguagliare le prestazioni del modello ottenuto nell'esperimento 6 senza il blocco 5.

#### Conclusioni parziali sulla rete pre-trainata VGG16

Nei nostri esperimenti mirati al riconoscimento di tessuti con cancro al seno come appartenenti alla classe 0 (senza cancro) o alla classe 1 (con cancro), ci siamo concentrati sull'utilizzo degli ultimi due blocchi della nostra rete di deep learning, ovvero il blocco 5 e il blocco 4. Abbiamo osservato che addestrando solo questi due ultimi blocchi, possiamo raggiungere ottime prestazioni. Inoltre, rimuovendo il blocco 5, possiamo comunque ottenere prestazioni confrontabili, sebbene leggermente inferiori. Questo approccio ci consente di risparmiare risorse computazionali e temporali, che hanno un costo.

Sulla base di questi risultati, stiamo ora cercando di migliorare ulteriormente il nostro pipeline di elaborazione delle immagini. In particolare, miriamo a migliorare le caratteristiche delle immagini in ingresso utilizzando l'Equalizzazione dell'Istogramma Limitato del Contrasto Adattivo (CLAHE), ed estendere la nostra analisi a tutti gli altri blocchi della rete, escludendo comunque il blocco 5. Ci aspettiamo che questo approccio ci consentirà di ottenere prestazioni ancora migliori su questo complesso compito.

## Esperimenti con CLAHE

Nei seguenti esperimenti è stato utilizzato CLAHE sulla rete neurale pre-addestrata VGG16 per migliorare le prestazioni della rete nel rilevare tumori in immagini mediche. Sono stati implementati due approcci.

Nel primo approccio, sono state create reti che utilizzano modello VGG16, con l'allenamento di un blocco alla volta, partendo dal blocco 4 e scendendo fino al blocco 1. Vi sono state in totale 6 fasi di addestramento.

Nella prima fase, il quinto blocco della rete VGG16 è stato rimosso per i motivi sopra definiti, mentre tutti gli altri blocchi sono stati "frizzati", cioè non sono stati aggiornati durante il training con l'aggiunta di un livello GlobalAveragePooling2D.

Nelle fasi successive, i blocchi sono stati "sbloccati" uno alla volta, cioè sono stati resi aggiornabili durante il training, in modo da permettere alla rete di apprendere nuove features più specifiche per il nuovo task.

Nella fase 6, tutti i blocchi sono stati completamente "unfrizzati", cioè sono stati resi completamente aggiornabili durante il training, in modo da permettere alla rete di apprendere anche features molto specifiche e complesse.

Di seguito verranno mostrate le fasi con i corrispettivi report.

- Fase 1: Blocco4 trainable = false, Blocco3 = false, Blocco2 = false, Blocco1 = false

Fase 1			
	precision	recall	f1-score
<b>0</b>	0.8882	0.909	0.8985
<b>1</b>	0.756	0.7113	0.733
<b>accuracy</b>	0.8529		

- Fase 2: Blocco4 trainable = true, Blocco3 = true, Blocco2 = false, Blocco1 = false

Fase 2			
	precision	recall	f1-score
<b>0</b>	0.9295	0.8729	0.9003
<b>1</b>	0.7221	0.8331	0.7736
<b>accuracy</b>	0.8616		

- Fase 3: Blocco4 trainable = false, Blocco3 = true, Blocco2 = false, Blocco1 = false

Fase 3			
	precision	recall	f1-score
<b>0</b>	0.9308	0.8778	0.9035
<b>1</b>	0.7304	0.8353	0.7793
<b>accuracy</b>	0.8657		

- Fase 4: Blocco4 trainable = false, Blocco3 = false, Blocco2 = true, Blocco1 = false

Fase 4			
	precision	recall	f1-score
<b>0</b>	0.926	0.8861	0.9056
<b>1</b>	0.7408	0.8214	0.779
<b>accuracy</b>	0.8677		

- Fase 5: Blocco4 trainable = false, Blocco3 = false, Blocco2 = false, Blocco1 = true

Fase 5			
	precision	recall	f1-score
<b>0</b>	0.9269	0.8892	0.9077
<b>1</b>	0.7466	0.8232	0.783
<b>accuracy</b>	0.8705		

- Fase 6: Blocco4 trainable = true, Blocco3 = true, Blocco2 = true, Blocco1 = true

Fase 6			
	precision	recall	f1-score
<b>0</b>	0.9224	0.8987	0.9104
<b>1</b>	0.7601	0.8093	0.7839
<b>accuracy</b>	0.8733		

Il report evidenzia un progressivo miglioramento delle prestazioni del modello al variare delle fasi di allenamento. In particolare, l'aggiunta di uno o più blocchi di rete trainabili ha portato ad un miglioramento significativo delle prestazioni rispetto alla fase precedente. Inoltre, si può notare come la migliore performance è stata raggiunta nella fase 6, in cui tutti i blocchi della rete erano trainabili.

Nel secondo approccio, la rete è stata allenata “sbloccando” tutti i blocchi senza però allenare blocco dopo blocco, ottenendo risultati simili ma con una migliore recall della classe 1. in pratica come nel primo approccio ma effettuando addestramento implementando solo la fase 6.

	precision	recall	f1-score
<b>0</b>	0.9401	0.8689	0.9031
<b>1</b>	0.7225	0.8605	0.7854
<b>accuracy</b>	0.8665		

Per la classe 0, la precisione è leggermente aumentata al 94,01%, il che significa che il modello è più preciso nel predire che una patch non contiene un tumore al seno. Tuttavia, il richiamo è leggermente diminuito al 86,89%, il che significa che il modello ha una minore capacità di individuare le patch senza tumore al seno quando queste sono presenti.

Per la classe 1, la precisione è leggermente diminuita al 72,25%, il che significa che il modello è meno preciso nel predire che una patch contiene un tumore al seno. Tuttavia, il richiamo è aumentato al

86,05%, il che significa che il modello è in grado di individuare correttamente una maggiore percentuale di patch con tumore al seno.

In un contesto medico, come quello relativo alla diagnosi del tumore al seno, il riconoscimento della presenza di un tumore è di fondamentale importanza, in quanto l

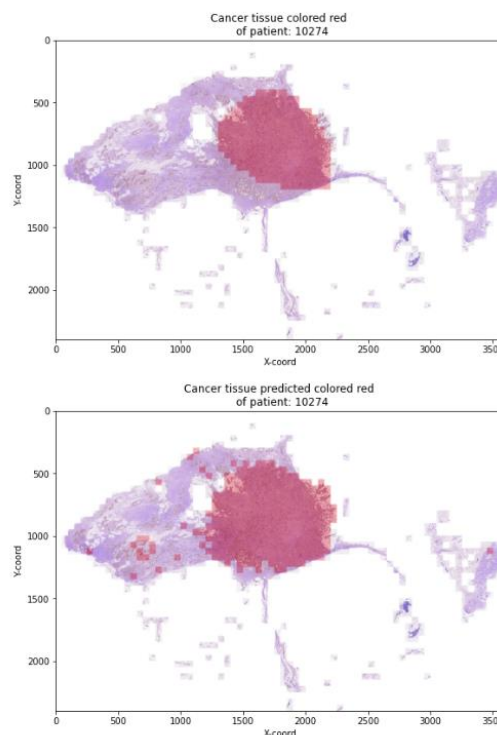
a sua mancata individuazione potrebbe avere conseguenze gravi per la salute della paziente.

Pertanto, il modello con una recall più elevata nella classe 1 (86,05% nel secondo esempio) rispetto al primo esempio (80,93%), anche a discapito di una precisione leggermente più bassa (72,25% nel secondo esempio rispetto al 76,01% nel primo), potrebbe essere preferito.

Questa scelta dipende dalla priorità che si attribuisce alla minimizzazione di falsi negativi rispetto alla minimizzazione di falsi positivi. Nel contesto della diagnosi del tumore al seno, è meglio evitare di trascurare la presenza di un tumore anche a costo di effettuare ulteriori verifiche su pazienti che, in realtà, non hanno un tumore, piuttosto che comunicare ai pazienti che il loro esame è negativo, trascurando invece la presenza del tumore.

Tuttavia, va tenuto presente che ogni contesto medico ha le proprie specificità e priorità, quindi la scelta finale dipende dalle esigenze specifiche del contesto e delle persone coinvolte.

In generale, gli esperimenti appaiono interessanti e possono fornire informazioni utili alla comunità scientifica riguardo l'uso di CLAHE sulla rete neurale pre-addestrata VGG16 per il rilevamento di tumori in immagini mediche. Tuttavia, sarebbe importante fornire maggiori dettagli sulle tecniche e sulle metriche di valutazione utilizzate per ottenere un quadro più completo e preciso dei risultati ottenuti.



*Figura 4: Differences between real cancer and predicted cancer*

## Explainability

L'impiego di tecniche di deep learning per la classificazione delle patch del cancro al seno sta diventando sempre più popolare in ambito medico, grazie alla loro capacità di analizzare grandi quantità di dati e individuare eventuali anomalie. Tuttavia, uno dei problemi principali delle reti neurali profonde è la loro complessità e la loro scarsa interpretabilità, il che le rende difficili da spiegare agli esperti medici. Questo è particolarmente critico nel contesto medico, dove la trasparenza e la giustificazione delle decisioni prese dalla rete sono fondamentali per ottenere la fiducia degli esperti.

In questo studio, abbiamo analizzato la spiegabilità delle reti neurali profonde addestrate da zero e pre-addestrate su dataset di immagini di mammografie per la classificazione delle patch del cancro al seno. In particolare, abbiamo indagato la capacità di queste reti di rilevare le caratteristiche delle immagini che portano alla classificazione corretta, con l'obiettivo di rendere la classificazione più interpretabile ed esplicabile.

Siamo stati in grado di identificare specifici strati delle reti neurali profonde che sembrano non fornire informazioni rilevanti per la classificazione delle patch del cancro al seno. In particolare, abbiamo notato che gli ultimi livelli della rete VGG16, ovvero "block4\_conv3" e "block4\_pool", sembrano produrre immagini completamente nere e probabilmente inutili per la classificazione.

Tuttavia, è importante sottolineare che la rimozione di questi strati non garantisce necessariamente una maggiore spiegabilità e interpretabilità delle reti neurali. Infatti, i modelli di deep learning sono spesso imprevedibili e ciò che può sembrare inutile ai fini della classificazione potrebbe comunque essere importante per le prestazioni computazionali della rete. Quindi, la rimozione di questi strati dovrebbe essere valutata con cautela e potrebbe richiedere ulteriori studi.



## Conclusioni

In questo studio è stata valutata l'efficacia di diversi modelli di deep learning per la classificazione delle patch del cancro al seno. È stato confrontato i risultati di due reti neurali profonde addestrate da zero e pre-addestrate su dataset di immagini di mammografie, utilizzando diverse tecniche di elaborazione delle immagini.

I risultati mostrano che il modello from-scratch dell'esperimento 2 ha raggiunto il miglior risultato per quanto riguarda la classificazione delle patch del cancro al seno, mentre per quanto riguarda il modello pre-addestrato VGG16, il secondo esperimento con la tecnica di elaborazione delle immagini CLAHE ha prodotto il miglior report. In particolare, la rete from-scratch ha ottenuto una recall della classe 1 addirittura superiore rispetto alla VGG16, con un valore di 0.8799 rispetto a 0.8605.

Questi risultati sono estremamente incoraggianti per il futuro sviluppo di modelli di deep learning per la classificazione del cancro al seno, poiché dimostrano che l'uso di tecniche di elaborazione delle immagini può migliorare significativamente le prestazioni dei modelli.

Inoltre, questo studio ha analizzato anche la spiegabilità delle reti neurali profonde, rivelando la presenza di strati che sembrano non fornire informazioni rilevanti per la classificazione delle patch del cancro al seno. Questi risultati potrebbero aprire la strada a future ricerche volte a ridurre la complessità delle reti neurali e migliorare la loro spiegabilità e interpretabilità.

In conclusione, il nostro studio ha dimostrato che i modelli di deep learning possono rappresentare un'importante risorsa nella lotta contro il cancro al seno. Tuttavia, ci sono ancora molte sfide da affrontare e molti aspetti da esplorare per migliorare ulteriormente le prestazioni dei modelli e renderli più interpretabili ed esplicabili.