

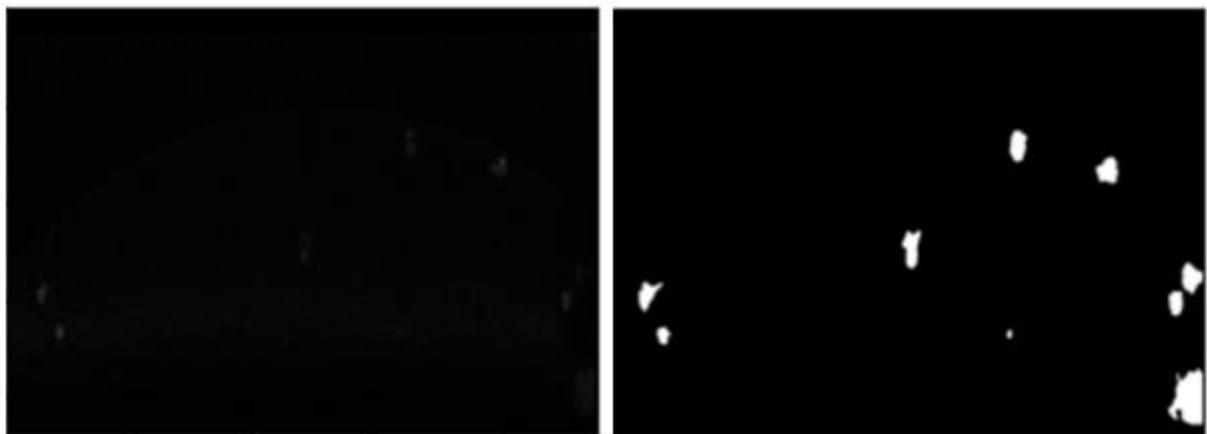
# ELABORAZIONE DI IMMAGINI MEDICHE

*Gruppo FB14: Quattrocchio S. s319792, Sera C. s315774, Truscello F. s313840*  
***Optical Coherence Tomography Challenge***

## 1. Introduzione

La OCT (Optical Coherence Tomography), è una tecnica di imaging non invasiva e non distruttiva, che permette di ottenere scansioni tomografiche di volumi di tessuti in analisi. Recenti studi hanno evidenziato le grandi potenzialità di questa tecnica nell'analisi di organoidi: colture cellulari in tre dimensioni che mimano organi o tumori in miniatura. In particolare, la OCT si è rivelata in grado di distinguere le regioni necrotiche da quelle sane degli organoids, e consente di tracciare la crescita di singoli organoidi nel tempo. Per effettuare analisi quantitative, il primo passo fondamentale è una corretta segmentazione di ogni unità presente all'interno del volume. Tuttavia, le immagini ottenute con questa tecnica sono caratterizzate da un basso rapporto segnale-rumore e dalla presenza di artefatti di riflesso.

Pertanto, l'obiettivo principale di questa challenge è lo sviluppo di un algoritmo in grado di eseguire la segmentazione automatica delle immagini 3D dei volumi sottoposti ad analisi e di fornire una stima affidabile del numero di organoidi presenti. Data l'elevata difficoltà del task, viene proposto un metodo di segmentazione basato su deep learning: la rete neurale utilizzata è una U-Net in grado di discriminare tra due classi, rispettivamente lo sfondo e l'organoide.



*Figura 1. Immagine OCT bidimensionale di organoidi tumorali e corrispondente maschera manuale a destra.*

## 2. Materiali e metodi

### 2.1. Dataset

Il dataset assegnato è composto da 12 volumi di 698 slice ciascuno e le relative maschere binarie ottenute mediante segmentazione manuale di un operatore esperto. I volumi sono forniti in formato .MAT con codifica su 16 bit, mentre le segmentazioni, in formato .PNG, sono codificate su 8 bit. Poiché i volumi hanno delle dimensioni spaziali variabili si è ritenuto opportuno ridimensionare le immagini di ogni volume alle dimensioni del volume più piccolo (800x1200).

Inoltre, si nota anche una grande variabilità dimensionale degli organoidi, di cui si propone un esempio nella figura 2.

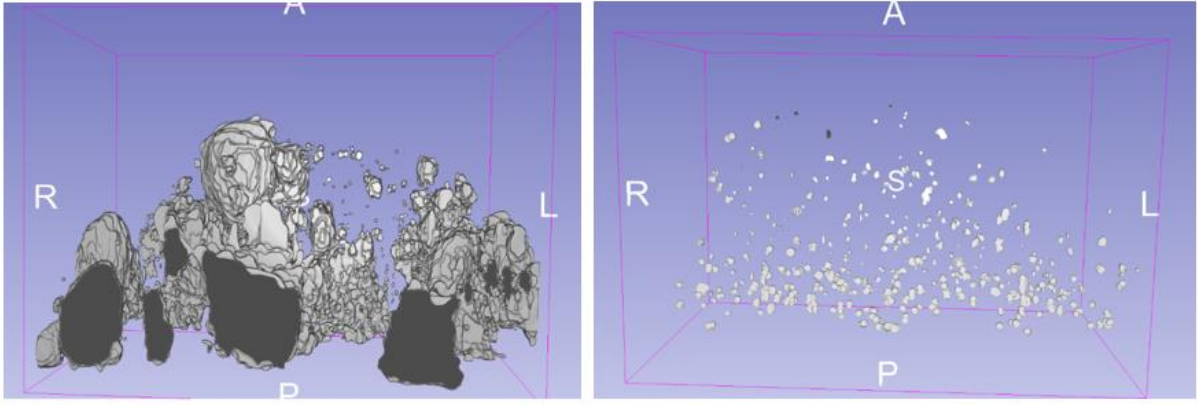


Figura 2. Ricostruzione 3D degli organoids a partire dalle maschere manuali mediante il software 3D Slicer (a sinistra il volume w6\_d13 mentre a destra il volume d1\_p1\_w2). Le immagini sono state ottenute rimuovendo le slice bianche presenti nel dataset.

## 2.2. Metriche

### Dice 3D

Per la valutazione delle performance dell'algoritmo proposto è stato calcolato il Dice Similiary Coefficient (DSC) espresso come:

$$DSC(X, X_{predicted}) = \frac{2 * |X \cap X_{predicted}|}{|X| + |X_{predicted}|}$$

In cui  $X$  rappresenta l'insieme dei voxel delle maschere ground truth mentre  $X_{predicted}$  rappresenta l'insieme dei voxel delle maschere predette dall'algoritmo di segmentazione. Al fine di calcolare il DSC è stato necessario ricostruire il volume della coltura a partire dalle segmentazioni 2D, che sono state sovrapposte per formare un'unica matrice tridimensionale.

### Errore conteggio organoidi

L'accuratezza della segmentazione automatica in termini di stima del numero di organoidi è stata valutata mediante l'errore numerico nel loro conteggio.

È stato inoltre calcolato l'errore percentuale sulla stima numerica degli organoidi, per avere una maggiore rappresentatività dell'accuratezza, riferita al ground truth delle maschere manuali. La metrica è stata calcolata nel seguente modo:

$$errore\ numerico = |N_{reale} - N_{stimato}|, \quad errore\ percentuale = \frac{|N_{reale} - N_{stimato}|}{N_{reale}} * 100$$

In cui:

- $N_{reale}$  esprime il numero reale di organoidi, ottenuto a partire dalle maschere manuali fornite;
- $N_{stimato}$  indica il numero di organoidi stimato dalla rete neurale.

Il conteggio è stato effettuato sul volume utilizzando una connettività 3D ovvero, stabilendo una relazione di connessione tra i voxel nelle direzioni ortogonali e diagonali (insieme 26-connesso).

In aggiunta, in fase di allenamento è stata considerata anche la metrica dell'Intersection over Union (IoU) sulle immagini bidimensionali, per avere un ulteriore indicatore dell'avanzamento dell'apprendimento. Tuttavia, questa metrica non è stata utilizzata come criterio di confronto nella selezione delle diverse tecniche

testate. Si è preferito dare priorità alle metriche di validazione che tenessero conto della tridimensionalità dei dati; di conseguenza, i risultati numerici della IoU non sono stati riportati nelle tabelle.

Il parametro è definito come segue:

$$IoU(X,Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

In cui X e Y rappresentano l'insieme dei pixel rispettivamente della maschera manuale e della maschera generata automaticamente.

## 2.3. Rete neurale

### 2.3.1 Architettura

Il modello scelto è basato su una rete neurale convoluzionale U-Net [1], il cui design permette di combinare caratteristiche ad alto e basso livello attraverso la loro concatenazione [Fig.3a]. Questo approccio si è dimostrato particolarmente efficace nei task di segmentazione in ambito medico e, grazie alle librerie disponibili in PyTorch, risulta facilmente implementabile. Nella fase iniziale del lavoro sono state confrontate le performance di due architetture UNet differenti al fine di definire il modello migliore per lo specifico task di segmentazione.

Il primo modello di U-Net implementato è caratterizzato da 4 layer di encoder e 4 layer di decoder. Ogni layer è composto da due convoluzioni 3x3, seguite da una funzione di attivazione ReLU e da un'operazione di max pooling 2x2. I layer del decoder sono concatenati con le corrispondenti feature map dell'encoder, seguiti anch'essi da due convoluzioni 3x3 e dall'applicazione di una ReLU.

Il secondo modello è stato ottenuto apportando delle modifiche all'architettura originale, seguendo il lavoro di Zhou et al. [2]. In particolare, ad ogni layer viene concatenato un numero di feature inversamente proporzionale alla sua profondità [Fig. 3b]. L'idea di base è che i layer più superficiali presentino una maggiore distanza semantica tra le feature map dell'encoder e quelle del decoder. Per compensare questa differenza, sono stati aggiunti (i-1) layer di convoluzione ad ogni livello i-esimo, ottenuti tramite upconvolution dal livello più profondo.

Il numero di classi per la segmentazione è pari a 2, rispettivamente 0 per lo sfondo e 1 per gli organoidi.

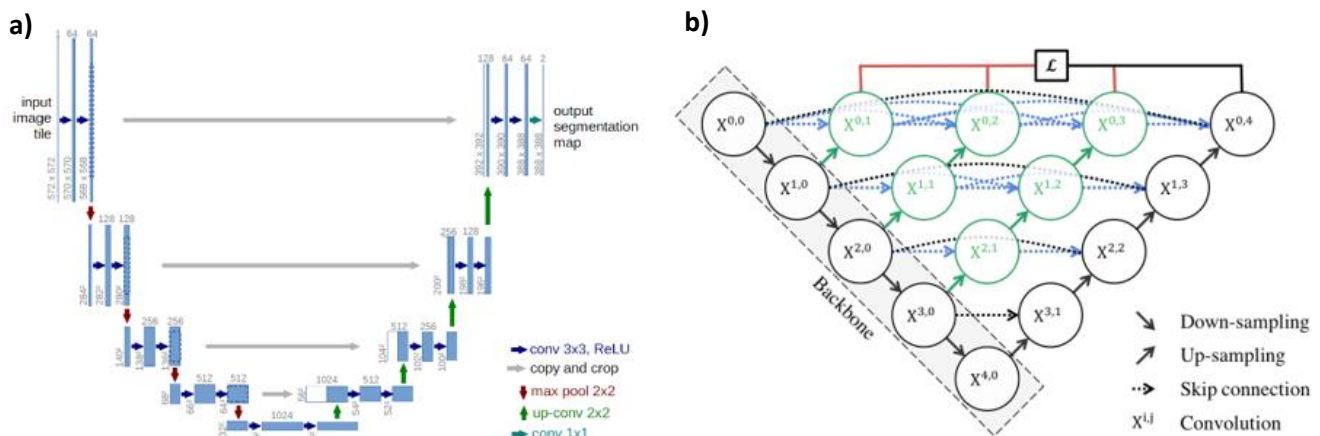


Figura 3. Architetture delle reti neurali utilizzate: a) UNet e b) UNet++.

### 2.3.2 Iperparametri dell'allenamento

Gli allenamenti sono stati realizzati con learning rate pari a 0.0005 e batch size di 4. È stato necessario raggiungere un compromesso tra ridimensionamento delle immagini e batch size in modo tale da mantenere delle dimensioni confrontabili con quelle originali e massimizzare il numero di immagini per batch nei limiti della RAM disponibile aumentando la velocità di convergenza dell'allenamento. È stato quindi necessario eseguire un ridimensionamento da 800x1200 pixel a 512x512 pixel

Il numero massimo di epoche è stato impostato a 20, con un criterio di “early stopping” grazie al quale l'allenamento può terminare precocemente nel caso in cui per 5 epoche consecutive non si verifichi un miglioramento delle performance sul validation set.

Inoltre, sono state valutate le performance della rete utilizzando due differenti loss function che sono comunemente utilizzate in task di classificazione binaria: dice loss e binary cross entropy loss (BCE) [3].

$$L_{Dice}(X, X_{predicted}) = 1 - \frac{2 * X \cap X_{predicted}}{X \cup X_{predicted} + \epsilon}$$

$$L_{BCE}(X, X_{predicted}) = -(X * \log(X_{predicted}) + (1 - X) * \log(1 - X_{predicted}))$$

Come ottimizzatore è stato scelto Adam fornito dalle librerie di Python.

### 2.4. Data Augmentation

È stato poi eseguito data augmentation per aumentare la variabilità dei dati in ingresso alla rete e quindi, migliorarne la capacità di generalizzazione. A tal fine, vengono generate delle nuove immagini applicando una o più trasformazioni alle immagini originali e alle maschere corrispondenti.

Per garantire la coerenza spaziale dei dati sintetici è importante considerare le caratteristiche dello specifico dataset: nel nostro caso, sono state ad esempio evitate trasformazioni come la rotazione random di 90°, in quanto avrebbe creato delle immagini con parte degli organoidi nella porzione superiore, dove invece sono presenti i riflessi, compromettendo la capacità della rete di discernere quest'ultimi dagli organoidi. Inoltre, ogni tipo di rotazione è stata evitata, per non pregiudicare la coerenza e la continuità spaziale degli stessi nel successivo impilamento necessario per il calcolo delle metriche tridimensionali in fase di allenamento.

È stata usata la libreria albumentations [4], con le seguenti trasformazioni: Resize (512x512), aggiunta di luminosità random (p=0.2), random gamma correction (p=0.5).

### 2.5. Preparazione Dataloader

Dal momento che le metriche di valutazione sono calcolate sui voxel, i 12 volumi sono stati suddivisi nei set di training, validation e test mantenendo tutte le slice appartenenti ad uno stesso volume all'interno dello stesso set. In particolare, 8 volumi sono stati utilizzati per la costruzione del training set, mentre i restanti 4 sono stati equamente suddivisi tra il validation ed il test set.

I dati sono stati suddivisi in modo che ciascun set fosse rappresentativo della popolazione totale da cui sono stati estratti: le immagini tomografiche delle colture sono state acquisite in giorni diversi di osservazione, di conseguenza, gli organoidi contenuti hanno delle dimensioni variabili [Fig. 2]. Ogni set è provvisto sia di volumi con organoidi mediamente più piccoli sia di scansioni tomografiche in cui le dimensioni degli organoidi sono maggiori.

Per il set di train sono stati scelti i volumi: d1\_p1\_w2, w2\_d7, w2\_d9, w2\_d11, w2\_d13, d3\_p1\_w2, w6\_d7 e w6\_d9. Mentre per il validation sono stati selezionati i volumi w6\_d5 e w6\_d11, infine il test set è composto dai volumi w2\_d5 e w6\_d13.

In fase di allenamento, poi, per garantire una maggiore capacità di generalizzazione della rete neurale, le slice dei volumi sono state fornite in modo random (shuffle=True). Successivamente, attraverso il nome dei file contenuti nel batch corrente, è stata eseguita la ricostruzione del volume per il calcolo del dice tridimensionale.

Nella sezione seguente vengono illustrate le prove effettuate per definire le caratteristiche dell'architettura, loss function, pre-processing e post processing, mentre i risultati sono descritti nel paragrafo 3.

## 2.6. Selezione dell'architettura e della loss function

Inizialmente, per determinare la migliore combinazione tra le due architetture di rete neurale proposte al paragrafo 2.3.1, e le due funzioni di perdita citate [Par. 2.3.2], si è deciso di testare le capacità di predizione sul dataset non processato, svolgendo quattro allenamenti con le caratteristiche mostrate in tabella 1. La combinazione con le performance migliori sui set di allenamento e validazione è stata utilizzata successivamente per valutare gli effetti delle tecniche di pre-processing e post-processing sulla segmentazione.

I risultati delle quattro prove sono stati confrontati mediante il calcolo, in fase di allenamento della rete, del coefficiente di Dice volumetrico per ciascun volume del training e validation set e infine mediando i valori ottenuti. A ciascun allenamento sono stati associati due valori medi di Dice 3D, uno per il training set ed uno per il validation set, con la relativa deviazione standard.

		Architettura rete	
		Unet Base	Unet ++
Loss function	Dice loss	Training 1	Training 2
	BCE loss	Training 3	Training 4

Tabella 1. Combinazioni di allenamenti per definire l'architettura e la funzione di perdita con le performance migliori

## 2.7. Pre-processing

Data la significativa presenza di artefatti da riflesso e di rumore delle immagini OCT, sono state esplorate diverse strategie per migliorare la qualità dei dati in ingresso alla rete neurale. Le immagini sono state processate mantenendo la loro dinamica di 16 bit, e successivamente sono state salvate su 8 bit. Di seguito, vengono approfondite le tecniche implementate.

- **Processing 1: Denoising (filtro mediano+ filtro gaussiano)**

Dato che le immagini acquisite con la tecnica OCT sono caratterizzate dalla presenza di rumore di tipo “speckle” si è scelto di applicare un filtro gaussiano ed un filtro mediano in cascata. Quest’ultimo, infatti, risulta particolarmente efficace nel ridurre le variazioni rapide e isolate di intensità che caratterizzano il rumore “speckle”. Il primo filtraggio è stato realizzato con un kernel di dimensione 3 ed un sigma pari a 1, mentre il secondo con un kernel quadrato di dimensione 5x5.

- **Processing 2: Denoising ed equalizzazione adattiva (CLAHE)**

Al fine di evidenziare le strutture di interesse si è pensato di applicare un’equalizzazione adattiva, in cui l’aumento di contrasto viene eseguito sulla base dell’intensità dei pixel limitrofi, e non sull’immagine globale in quanto l’equalizzazione uniforme ha mostrato di amplificare eccessivamente il rumore.

- **Processing 3: Denoising, equalizzazione e sharpening con gradienti obliqui**

In aggiunta alle strategie di pre-processing in precedenza descritte, si è cercato un metodo per evidenziare i contorni delle strutture di interesse, senza accentuare ulteriormente i riflessi. Dal momento che si è notata la tendenza di questi ultimi a presentarsi come linee orizzontali o verticali, si è optato per il calcolo di un gradiente che non li enfaticasse. Pertanto, è stato calcolato il gradiente obliquo, di inclinazione pari a  $45^\circ$  e  $-45^\circ$ , e successivamente è stato eseguito lo sharpening sulla base di questi ultimi.

- **Processing 4: Denoising con filtraggio a media non locale ed equalizzazione CLAHE.**

L'ultima strategia di pre-processing implementata ha previsto la ricerca di un metodo maggiormente incisivo nell'eliminazione del rumore che caratterizza le immagini dell'optical coherence tomography; pertanto, è stato implementato un filtraggio a media non locale (funzione fornita dalla libreria di skimage: "non local means", patch size=9, patch distance=5, h=0.1), seguito da un'equalizzazione adattiva [5].

La figura 4 mostra un confronto delle quattro tecniche di pre-processing applicate, a partire dalla stessa immagine originale.

Infine, le immagini sono state convertite in uint8, e fornite alla rete in formato .png, come le relative maschere.

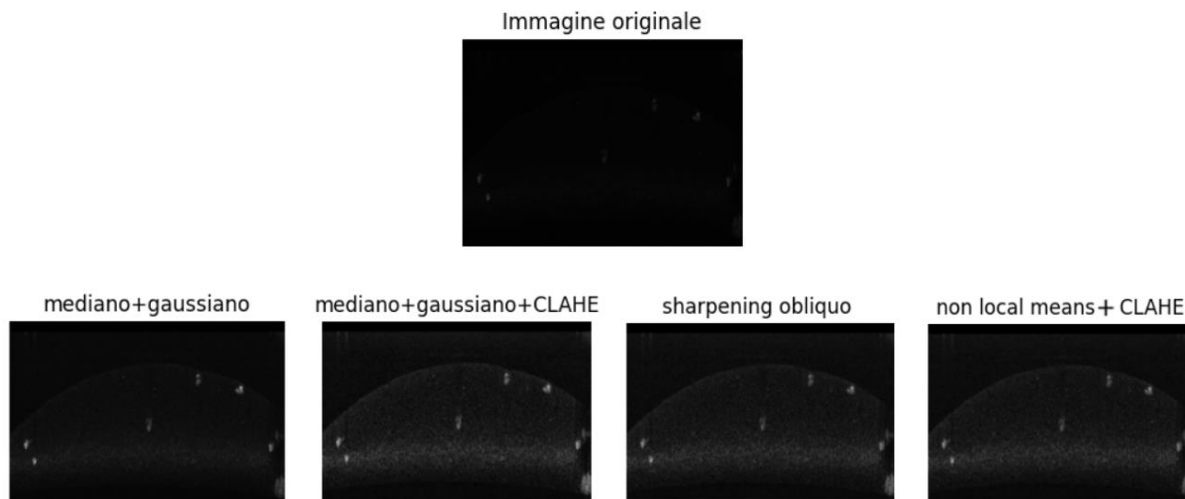


Figura 4. Rappresentazione degli effetti del processing introdotto sul volume w2\_d9

## 2.8. Post-processing

Definite l'architettura di rete e la tecnica di pre-processing migliori, un post-processing opportuno è in grado di migliorare le performance di predizione ottenute. Le tecniche di post-processing sono state applicate alle maschere binarie, ricavate imponendo una soglia sulle softmax pari a 0.5. Si è deciso di valutare l'effetto dei seguenti operatori morfologici, con l'obiettivo di rimuovere piccole imprecisioni della segmentazione:

- Erosione, disco 5x5;
- Apertura, disco 5x5;
- "remove\_small\_objects"

La complessità intrinseca del task di segmentazione, dovuta alla variabilità morfologica e dimensionale degli organoidi, non ha permesso di implementare tecniche maggiormente complesse, come ad esempio un criterio di analisi dell'eccentricità o dei bordi.

## 3. Risultati

Di seguito viene riportato un flowchart riepilogativo delle prove eseguite per definire la pipeline di segmentazione [Fig.5]. I risultati ottenuti e le conseguenti scelte sono descritti nei paragrafi successivi.

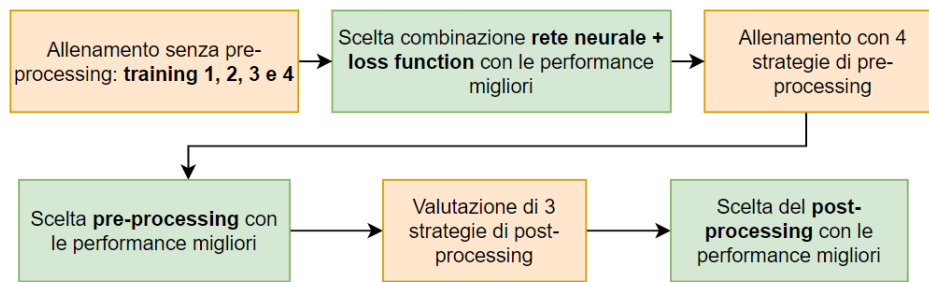


Figura 5. Flowchart riepilogativo

### 3.1. Selezione dell'architettura e della loss function

In questa fase preliminare si è scelto di utilizzare come unico parametro di validazione il Dice 3D. La rete ha mostrato, in tutti gli allenamenti, di avere delle performance peggiori nella segmentazione del volume “d1\_p1\_w2” a causa del suo basso rapporto segnale rumore [Fig.6]. Infatti, come si può osservare nella tabella 2, la deviazione standard dei valori di Dice 3D dei volumi del training set risulta significativamente più alta includendo il volume d1\_p1\_w2.

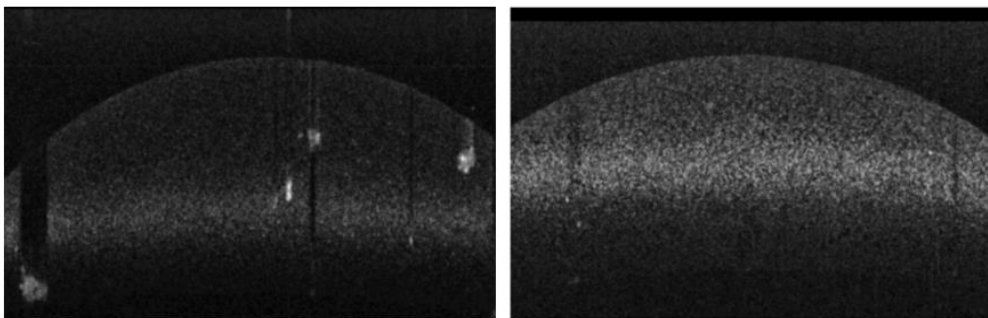


Figura 6. Immagini di volumi differenti a confronto. A destra una slice del volume d1\_p1\_w2 risulta molto più rumorosa dell'immagine a sinistra che fa parte del volume w2\_d7. Le immagini sono state elaborate mediante filtraggio ed equalizzazione adattiva a partire da quelle non processate.

Per questo motivo, lo stesso volume è stato considerato come un outlier e, di conseguenza, è stato escluso dal training set nei successivi allenamenti. Inoltre, le valutazioni in merito alla combinazione migliore di funzione di perdita e architettura della rete sono state effettuate sulle performance ottenute escludendo il volume citato.

Il training 4 (BCE loss e U-Net++) è risultato il migliore in termini di valore medio e deviazione standard del coefficiente di Dice 3D dei volumi del training set [Tab.2]. Per quanto concerne i risultati sul validation set, il training 2 ha mostrato le prestazioni migliori. Poiché tutti i modelli di rete che sono stati testati, dimostrano di avere performance sul validation set confrontabili tra loro si è deciso di scegliere la combinazione con i risultati migliori sul training set. Le prove successive sono state, quindi, effettuate con la U-Net++ e BCE loss function.

	TRAINING		VALIDATION
	DICE [MEAN ± STD]	DICE senza volume [MEAN ± STD]	DICE [MEAN ± STD]
Training 1	0.7347±0.2754	0.8274±0.0906	0.7825±0.1464
Training 2	0.7446± 0.2773	0.8381±0.0887	0.7957±0.1410
Training 3	0.7282± 0.2815	0.8237±0.0850	0.7534±0.1684
Training 4	0.7576±0.2956	0.8589±0.0778	0.7735±0.1550

Tabella 2. Risultati in termini di dice volumetrico [media± deviazione standard] degli allenamenti descritti nel paragrafo 2.4 e nella tabella 1.



## 3.2. Pre-processing

La selezione della tecnica di pre-processing migliore è stata effettuata valutando i valori medi del coefficiente di Dice 3D e l'errore nel conteggio degli organoids, sia in termini di valore numerico assoluto, sia come percentuale rispetto al numero reale di organoids delle maschere manuali, come spiegato al paragrafo 2.2.

Inoltre, è stato eseguito un confronto rispetto alle performance del training set non processato, come si riporta nella tabella 3. Le quattro tecniche di pre-processing testate sono state descritte al paragrafo 2.7. Si precisa che nella seguente tabella è stato inserito l'errore di conteggio in termini di valor medio e deviazione standard, per rendere di più immediata comprensione i risultati intermedi finora ottenuti. Una maggiore attenzione ai singoli errori di conteggio di ciascun volume processato sarà prestata in fase di post-processing [Par. 3.3].

UNET ++, BCE loss	TRAINING			VALIDATION		
	DICE [MEAN $\pm$ STD]	ERR # (MEAN $\pm$ STD)	ERRORE % (MEAN $\pm$ STD)	DICE [MEAN $\pm$ STD]	ERR # (MEAN $\pm$ STD)	ERRORE % (MEAN $\pm$ STD)
SENZA PRE-PROCESSING	0.8226 $\pm$ 0.1083	257 $\pm$ 100	108.5 $\pm$ 63.8	0.7815 $\pm$ 0.1355	214 $\pm$ 162	136.4 $\pm$ 159.5
PROCESSING 1	0.8154 $\pm$ 0.0887	76 $\pm$ 80	26.2 $\pm$ 22.4	0.7127 $\pm$ 0.1535	68 $\pm$ 28	28.5 $\pm$ 22.2
PROCESSING 2	0.8666 $\pm$ 0.0871	187 $\pm$ 60	76.2 $\pm$ 44.8	0.7950 $\pm$ 0.0820	206 $\pm$ 109	131.2 $\pm$ 108.2
PROCESSING 3	0.8240 $\pm$ 0.1093	571 $\pm$ 184	228.9 $\pm$ 113.8	0.7689 $\pm$ 0.1519	163 $\pm$ 259	84.6 $\pm$ 162.6
PROCESSING 4	0.7395 $\pm$ 0.0875	104 $\pm$ 115	34.6 $\pm$ 40.0	0.6251 $\pm$ 0.1713	37 $\pm$ 4	17.5 $\pm$ 11.7

Tabella 3. Risultati in termini di dice volumetrico ed errore di conteggio (assoluto e percentuale) delle quattro strategie di pre-processing testate su training set e validation set [Media  $\pm$  deviazione standard].

Si osserva che, in termini di Dice 3D sul training set, le strategie di pre-processing 2 e 3 risultano migliorative rispetto al caso senza pre-processing; per quanto riguarda l'errore numerico del conteggio degli organoids, invece, sono le strategie di pre-processing 1,3 e 4 a mostrare un progresso. Analizzando il validation set si mostra vincente il caso 2 in termini di Dice 3D, ed i casi 1 e 4 per l'errore numerico di conteggio.

Data la necessità di effettuare una scelta della strategia migliore, si è deciso di favorire il parametro del Dice 3D, a discapito dell'errore numerico, con l'obiettivo di ottenere un suo miglioramento con un post-processing adatto al contesto che si è mostrato al Paragrafo 3.3. Infatti, attraverso una visualizzazione 3D di alcuni dei volumi a disposizione, si è notato che l'errore numerico relativamente elevato nel caso del processing 2 è determinato dalla presenza di un buon numero di falsi positivi di piccole dimensioni, che può essere ottimizzato con l'utilizzo di operatori morfologici. Al contrario, il minor errore numerico presente ad esempio in seguito al pre-processing 1 risulta in parte da attribuirsi ad una minore capacità della rete di riconoscere organoids anche di dimensioni mediamente grandi che si riflette in un minore valore medio di Dice 3D [Fig.7]. Pertanto, in conclusione, è stato scelto il pre-processing 2, basato su un doppio filtraggio, mediano e gaussiano, e da un'equalizzazione adattiva (CLAHE), come evidenziato nella tabella 2.

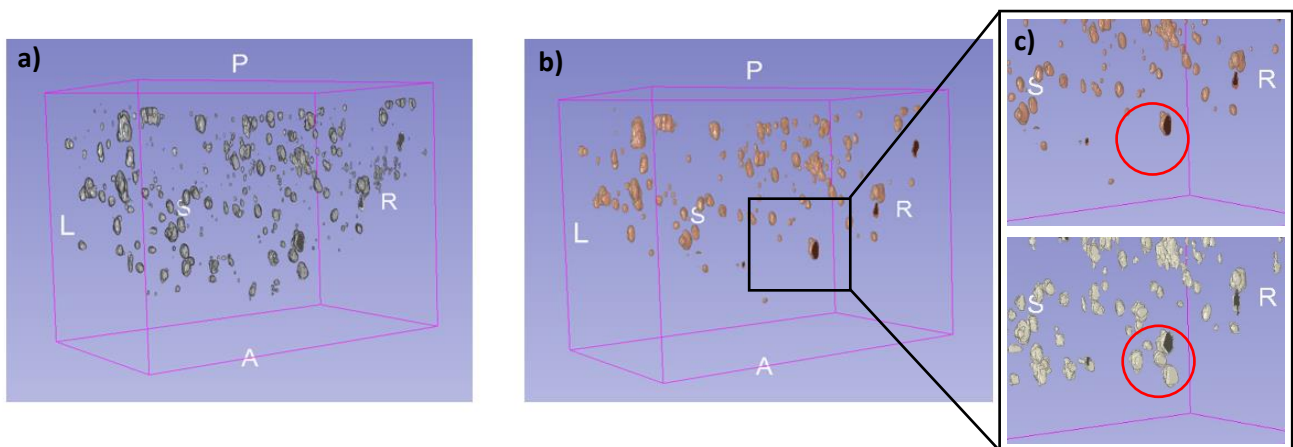


Figura 7. a) Rappresentazione 3D delle segmentazioni del volume w6\_d5 ottenute applicando alle immagini a) il processing 2, b) il processing 1. Dal volume nella figura b) è stata estratta una regione di interesse c): in alto si trova la ROI del volume segmentato con la rete mentre in basso la ROI dello stesso volume segmentato manualmente. Nel cerchio rosso sono evidenziati gli organoids che non sono stati riconosciuti come tali dalla rete.



### 3.3. Post-processing

L'applicazione delle tecniche di post-processing che sono state descritte nel paragrafo 2.8 hanno determinato un miglioramento delle capacità di predizione della rete che si possono apprezzare nella figura 8. In particolare, risulta che l'apertura e la funzione "remove small objects" non incidono significativamente sul valore medio di Dice volumetrico calcolato sul training set, che rimane molto simile a quello ottenuto senza alcuna tecnica di post-processing. Si nota, inoltre, che l'erosione causa un peggioramento del Dice 3D sui volumi del validation set, il quale scende dal 79,5 % al 59,6% [Tab.4].

Per quanto concerne, invece, l'errore nel conteggio degli organoidi, le tecniche di erosione e apertura permettono un notevole decremento di questo coefficiente in entrambi i set di dati.

La scelta della strategia di post-processing è, quindi, ricaduta sull'apertura 5x5 in quanto presenta le performance maggiori sul training set, mentre sul validation set è la tecnica che offre il migliore compromesso tra Dice 3D ed errore nel conteggio delle cellule: essa permette di ridurre il numero di falsi positivi, migliorando la specificità dell'algoritmo.

	TRAINING			VALIDATION		
	DICE (MEAN±STD)	ERR # (MEAN±STD)	ERRORE % (MEAN±STD)	DICE (MEAN±STD)	ERR # (MEAN±STD)	ERRORE % (MEAN±STD)
SENZA POST-PROCESSING	0.8666±0.0871	187±60	76.2±44.7	0.7950±0.0820	206±109	131.2±108.2
EROSIONE	0.7163±0.1573	86±78	30.2±24.6	0.5961±0.1891	73±49	49.1±43.4
APERTURA 5X5	0.8671±0.0877	71±79	23.7±20.2	0.7864±0.1252	90.5±27.6	50.1±47.1
REMOVE SMALL OBJECTS	0.8664±0.0875	137±59	53.4±31.5	0.7954±0.0815	138±98	94.1±84.7

Tabella 4. Risultati in termini di dice volumetrico ed errore di conteggio (assoluto e percentuale) delle quattro strategie di post-processing testate su training set e validation set. I risultati sono riportati in termini di media ± deviazione standard.

Al fine di ottenere una maggiore comprensione circa l'errore numerico di conteggio degli organoidi viene di seguito riportato l'errore su ciascun volume appartenente al training set ed al validation set [Tab.5].

	ERRORE NUMERO ORGANOIDI								
	TRAINING							VALIDATION	
VOLUMI	w2_d7	w2_d9	w2_d11	w2_d13	d3_p1_w2	w6_d7	w6_d9	w6_d5	w6_d11
SENZA POST-PROCESSING	245	110	197	298	153	172	136	316	97
EROSIONE	65	5	70	181	6	51	224	122	24
APERTURA 5X5	52	14	16	130	11	33	239	110	71
REMOVE SMALL OBJECTS	163	54	120	227	82	115	200	236	40

Tabella 5. Risultati in termini di errore di conteggio assoluto delle quattro strategie di post-processing testate su training set e validation set. I risultati sono riportati per ciascun volume.

Si può osservare come l'errore numerico di conteggio sia molto dipendente dal volume in analisi, come si può anche notare dai valori di deviazione standard [Tab.4]. In particolare, data la forte presenza di falsi positivi, si può sostenere che l'algoritmo causi una sovrastima del numero di organoidi presenti. I volumi che presentano un maggiore errore sono w2\_d7 e w2\_d13 sul training set, mentre sul validation set si osserva un errore più grande nel caso del volume w6\_d5 rispetto al secondo campione (w6\_d11). In generale, le tecniche di post-

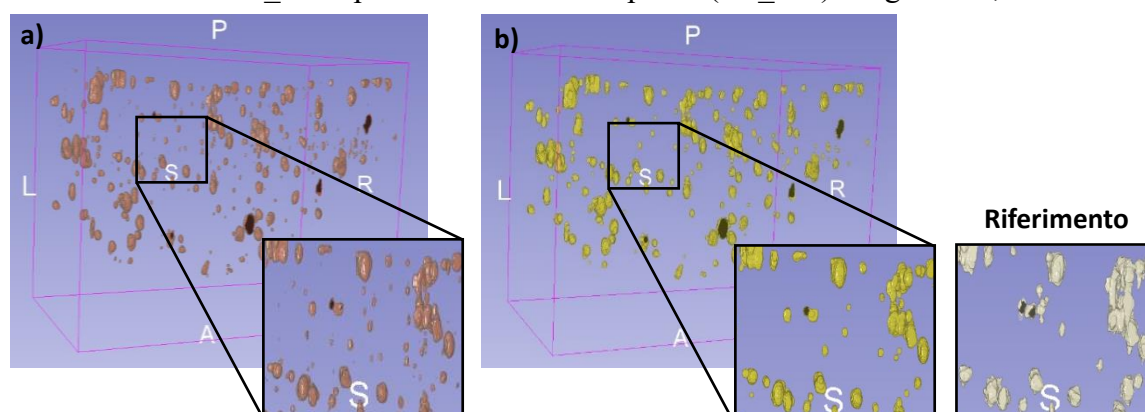


Figura 8. a) Rappresentazione 3D del volume w6\_d5 segmentato in modo automatico senza applicazione del post-processing. b) Rappresentazione 3D del volume w6\_d5 segmentato in modo automatico applicando il post-processing alle predizioni della rete. In ciascuna immagine è stata evidenziata una ROI per valutare l'effetto del post processing nella riduzione dei falsi positivi.

processing testate si mostrano efficienti, con qualche eccezione, come nel caso del volume w6\_d9. La scelta della chiusura offre un sostanziale miglioramento dell'errore, che si traduce in una riduzione del numero di falsi positivi [Fig.8b].

## 4. Discussione

L'obiettivo della challenge consisteva nello sviluppo di un algoritmo automatico in grado di segmentare cancer organoids presenti all'interno di volumi acquisiti con l'Optical Coherence Tomography (OCT). La maschera 3D binaria è ricostruita a partire dalle singole segmentazioni 2D, di conseguenza la rete ha una visione di insieme limitata nella classificazione degli organoidi. Mediante l'analisi dei risultati e delle performance ottenute si può notare che la scelta del pre e del post-processing contribuiscono al miglioramento delle predizioni ottenute dalla rete neurale; mentre la scelta della rete neurale e della loss non incide in modo significativo sulle prestazioni, come argomentato nel paragrafo 3.1.

In generale, il coefficiente di dice risultante può ritenersi soddisfacente, mentre si osserva una maggiore difficoltà nell'ottenere un errore numerico di conteggio affidabile. Come mostrato, tale problema è da attribuirsi al riconoscimento di piccoli organoidi che non risultano presenti nelle maschere manuali (e che si classificano, quindi, come falsi positivi). Sebbene il post processing scelto offra un miglioramento del parametro, senza influire negativamente sul dice, si nota ancora la presenza di un errore numerico non trascurabile [Fig.8b]. Possibili sviluppi futuri risiedono nell'esplorazione di tecniche in grado di migliorare ulteriormente l'accuratezza della predizione.

Sarebbe opportuno indagare tecniche diverse di denoising: le immagini ottenute mediante la tomografia a coerenza ottica risultano rumorose e con una forte presenza di riflessi.

Una soluzione potrebbe essere l'isolamento della regione emisferica in cui sono contenuti gli organoidi, in modo da eliminare gli artefatti da riflesso.

A tal scopo, si è cercato inizialmente di eseguire questa segmentazione calcolando i gradienti delle immagini per evidenziare la discontinuità tra lo sfondo e la regione emisferica contenente gli organoidi. Successivamente è stata identificata, per ogni colonna, la riga corrispondente al primo picco del gradiente a partire dall'alto. In questo modo sono state ricostruite le coordinate bidimensionali del bordo della ROI; tuttavia, non è stato possibile implementare questa tecnica a causa dell'elevata rumorosità (speckle) e della presenza di riflessi sul bordo della stessa area che non consentono di garantire una corrispondenza univoca tra "primo picco" del gradiente e profilo del gel.

Risulta necessario individuare una tecnica più robusta della precedente, come ad esempio una rete neurale per classificazione non supervisionata.

## 5. Conclusioni

In conclusione, l'algoritmo sviluppato ha dimostrato di essere efficace nella segmentazione dei cancer organoids in immagini OCT, ottenendo un buon compromesso tra coefficiente di Dice ed errore di conteggio grazie all'uso di tecniche di pre e post-processing specifiche.

Questi risultati possono avere un impatto significativo nella ricerca sul cancro, fornendo uno strumento automatizzato per la segmentazione accurata degli organoidi, facilitando così studi più rapidi ed efficienti.

Futuri lavori potrebbero esplorare ulteriormente tecniche avanzate di image processing e l'integrazione di reti neurali più sofisticate per migliorare ulteriormente la precisione della segmentazione degli organoidi. La continua evoluzione delle tecniche di imaging e di analisi dei dati offrirà nuove opportunità per migliorare la qualità e l'efficacia degli strumenti di segmentazione automatica, contribuendo significativamente ai progressi nella ricerca biomedica.

## Bibliografia

- [1] Ronneberger, O., Fischer, P., & Brox, T. **U-Net: Convolutional Networks for Biomedical Image Segmentation**. Retrieved from: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [2] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., & Liang, J. **UNet++: A Nested U-Net Architecture for Medical Image Segmentation**. Arizona State University. Retrieved from: [arXiv:1911.07067v1](https://arxiv.org/abs/1911.07067v1)
- [3] **Deep semantic segmentation of natural and medical images: a review**. ResearchGate. Retrieved from: [10.1007/s10462-020-09854-1](https://doi.org/10.1007/s10462-020-09854-1)
- [4] **Albumentations Documentation**. Retrieved from: <https://albumentations.ai/docs/>
- [5] Yu, H., Gao, J., & Li, A. **Probability-based non-local means filter for speckle noise suppression in optical coherence tomography images**. Optical Letters, 41(5), 994. Retrieved from: [http://dx.doi.org/10.1364/OL.41.000994](https://doi.org/10.1364/OL.41.000994)