SEGMENTAZIONE DI NODULI CEREBRALI

Sistemi Multimediali, A.A. 2024/2025
Villanova Simone, Alloggio Alessandra, Martiradonna Saverio
s.villanova@studenti.uniba.it, a.alloggio8@studenti.uniba.it, s.martiradonna13@studenti.uniba.it

Abstract

La segmentazione automatica dei tumori cerebrali rappresenta una sfida cruciale in ambito radiologico, con un impatto diretto sulla velocità e precisione della diagnosi. Questo lavoro presenta lo sviluppo di un sistema di segmentazione basato su **Mask R-CNN** con **backbone ResNet-50** e **Feature Pyramid Network (FPN)**, addestrato su immagini 2D del dataset *BRISC2025*, opportunamente convertite in formato *COCO* con maschere *pixel-wise* per tre classi tumorali (Glioma, Meningioma e Pituitario). Il training ha incluso una componente di *Dice Loss* personalizzata, volta a migliorare la qualità delle maschere su regioni frammentate. I risultati sul test set mostrano buone performance di segmentazione, con valori medi di **IoU** pari a **0.8189** e **Dice** coefficient di **0.8857**, associati a una **specificity** molto elevata (>**0.99**). Tuttavia, il modello evidenzia difficoltà nel rilevare morfologie tumorali particolarmente complesse, soprattutto nei gliomi, che rappresentano oltre il **75%** delle mancate rilevazioni. Questi risultati confermano l'efficacia dell'approccio **Mask R-CNN** *slice-based* come baseline, ma ne sottolineano anche i limiti intrinseci rispetto a morfologie irregolari. Lo studio apre la strada a futuri sviluppi basati su modelli volumetrici 3D e approcci multi-task per una segmentazione più accurata e robusta.

1. Introduzione

La diagnosi e la quantificazione manuale dei tumori cerebrali richiedono ore di lavoro specialistico, sono soggette a variabilità inter-operatore e comportano costi elevati oltre al rischio di omissioni critiche. Ritardi nelle valutazioni possono rallentare l'avvio delle terapie e incrementare le spese ospedaliere, mentre errori di misurazione influiscono negativamente sull'esito clinico. L'introduzione di sistemi di intelligenza artificiale per la segmentazione semantica delle lesioni non mira a sostituire il medico: si presta come supporto rapido, riproducibile e in grado di generare un secondo parere istantaneo, aumentando la produttività e riducendo significativamente tempi, costi e la probabilità di sviste umane.

1.1 Stato dell'arte

1.1.1 MUNet: fusione di SD-SSM e SD-Conv

Nel paper pubblicato su *Frontiers in Computational Neuroscience*, **MUNet** affronta la segmentazione dei tumori cerebrali integrando **U-Net** e **Mamba Network** in un'unica architettura. Il componente chiave è il blocco **SD-SSM** (*Selective-Scanning State-Space Model*), che suddivide il flusso di feature in due vie parallele:

- **Ramo locale**: utilizza il modulo **SD-Conv**, a sua volta costituito da **SCConv** (*Spatial+Channel Reconstruction Convolution*) e *depthwise separable convolutions*, per estrarre dettagli puntuali riducendo al minimo ridondanza e parametri.
- **Ramo globale**: impiega un modello di spazio di stato per catturare le dipendenze a lungo raggio e mantenere la coerenza strutturale sull'intera immagine.

Le *residual connection* e le *skip-connection* tra encoder e decoder ricompongono le due rappresentazioni, assicurando sia definizione dei bordi che contesto spaziale. Il training adotta una **loss ibrida** (**mIoU** + **Dice** + **Boundary**), bilanciata per ottimizzare sovrapposizione e nitidezza dei contorni.

Risultati:

- **BraTS2020**: **Dice Score** 0.835 (ET), 0.915 (WT), 0.823 (TC); **Hausdorff95** 2.421, 3.755, 6.437.
- **BraTS2018** e **LGG**: prestazioni analoghe e buona generalizzazione su dati indipendenti.

Link al paper: https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2025.1513059/full

1.1.2 Unified HT-CNNs Architecture: ensemble ibrido e transfer learning

Unified HT-CNNs lavora su volumi MRI 3D trattati *slice-by-slice* per coniugare il contesto spaziale locale e la consistenza volumetrica globale. In fase di pre-allenamento il modello apprende rappresentazioni di base sul dataset **BraTS**, poi affina i suoi pesi tramite i **High-Throughput Module**, una serie di strati dedicati al *transfer learning* che si adattano rapidamente a domini eterogenei come gliomi adulti e tumori pediatrici. Ogni slice viene processata da blocchi **Transformer gerarchici** che applicano attenzione assiale lungo i tre assi spaziali, integrati da convoluzioni tradizionali per catturare pattern locali ad alta risoluzione. Le maschere elaborate dai singoli *esperti* confluiscono nel metodo di fusione **STAPLE**, che combina le predizioni in una segmentazione finale più robusta e meno sensibile al rumore.

Risultati:

- **Tumori pediatrici**: incremento del **Dice** da 0,4097 a 0,6248 e riduzione dell'**Hausdorff95** da 163,37 a 37,46
- **Dataset Sub-Saharan Africa**: **Dice** da 0,7832 a 0,8647 e **Hausdorff95** da 18,38 a 10,98

Link al paper: https://www.researchgate.net/publication/387026244_Unified_HT-CNNs_Architecture_Transfer_Learning_for_Segmenting_Diverse_Brain_Tumors_in_MRI_from_Gliomas_to_Pediatric_Tumors

1.1.3 XAI-MRI: ensemble dual-modality con explainability

XAI-MRI scompone ciascun volume 3D in *slice* 2D e addestra quattro **U-Net** indipendenti su *T1*, *T2*, *T1ce* e *FLAIR*, valutando poi le performance di ogni modalità. Le due sequenze migliori vengono unite in un *ensemble* dotato di decoder con *up-sampling* e *skip-connection* per integrare informazioni di contrasto e dettaglio anatomico. Per introdurre trasparenza, si sfrutta **Grad-CAM**: le mappe di attivazione mostrano esattamente le regioni che guidano la segmentazione, facilitando il confronto e la correzione da parte del radiologo. Questo approccio combina la ricchezza multimodale con l'interattività, offrendo risultati di alta affidabilità.

Risultati:

- **BraTS2020**: **Dice Coefficient** 0,9773 e **Mean IoU** 0,6008

Link al paper: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1525240/full

1.1.4 Menzione onorevole: MedGemma, modello multimodale per la classificazione (prospettive di segmentazione)

MedGEMMA v3 è un *Large Vision Model* medicale basato su un encoder Transformer visivo con architettura simile ai **Vision Encoder** di tipo **ViT**, combinato con un modulo di *grounding multimodale* per l'allineamento immagine-testo. Il modello è pre-addestrato su milioni di immagini medicali eterogenee (*RM*, *TAC*, *radiografie*, *istologia*) e progettato per task di **classificazione**, **retrieval** e **grounding**, con capacità *zero-shot* e *few-shot*. **Non esegue segmentazione semantica**, ma l'architettura *encoder-centric* e la struttura del *grounding* rendono il modello **predisposto** per un'estensione verso la segmentazione, tramite l'integrazione di un decoder spaziale, analogamente ai paradigmi **Segment Anything**.
È stato incluso per la sua rilevanza come backbone generalista multimodale nel contesto medicale.

Risultati (classificazione):

- **AUC** di classificazione superiore a 0,95 su test interni

Link al modello: https://deepmind.google/models/gemma/medgemma/

2. Materiali e Metodi

2.1 Dataset:

Poiché il dataset inizialmente fornito era adatto alla classificazione ma non alla segmentazione, abbiamo scelto di utilizzare un **dataset** alternativo: **BRISC2025** (https://www.kaggle.com/datasets/briscdataset/brisc2025). Questo dataset fornisce immagini cerebrali accompagnate da maschere di segmentazione *pixel-wise*, ma non segue il formato *COCO* richiesto da **Detectron2**.

Per questo motivo, abbiamo implementato uno script di conversione personalizzato che trasforma ogni immagine e maschera nel formato *COCO* standard per la segmentazione, utilizzando la **codifica RLE** (*Run-Length Encoding*).
Durante la conversione, sono state definite tre categorie tumorali: Glioma, Meningioma e Pituitario.
Il **dataset finale** è strutturato come segue:

- **Training set**
  - **Glioma**: 1147 immagini
  - **Meningioma**: 1329 immagini
  - **Pituitario**: 1457 immagini
- **Test set**
  - **Glioma**: 254 immagini
  - **Meningioma**: 306 immagini
  - **Pituitario**: 300 immagini

Questa struttura bilanciata ha consentito un training efficace e una valutazione accurata su ciascuna classe tumorale.

2.2 Preprocessing e Augmentation

Il dataset è organizzato in formato compatibile con **Detectron2**, con annotazioni in formato **bitmask**. Ogni immagine viene preprocessata tramite un **custom mapper**, che applica una pipeline di **augmentation 2D** su immagini *slice* (non su volumi 3D completi). Le trasformazioni applicate includono:

- **Resize** fisso a 512×512 pixel
- **Random flip** orizzontale (probabilità 0.5)
- **Random rotation** entro un intervallo di ±15°
- **Random crop** relativo all'80% della dimensione originale

Queste trasformazioni vengono propagate anche alle maschere mediante le trasformazioni geometriche corrispondenti, preservando l'allineamento immagine-maschera. Tutte le immagini vengono convertite in tensori **float32** nel formato (C, H, W).

2.3 Architettura del modello

Il modello base è una **Mask R-CNN** con **backbone ResNet-50 + FPN** (**Feature Pyramid Network**), configurato per segmentazione semantica su 3 classi tumorali. L'architettura **GeneralizedRCNN** è mantenuta come struttura principale.

La head di mask utilizza **ROI Align** per estrarre regioni candidate e applica una **convoluzione deconvolutiva** per generare una maschera binaria per ciascun oggetto.

Un **hook personalizzato** estende il modello standard introducendo una componente di **Dice Loss** durante l'ottimizzazione: viene aggiunta al termine standard **loss_mask** di **Detectron2**, migliorando la qualità della segmentazione su aree tumorali poco bilanciate o frammentate. La **perdita combinata** usata durante il training è quindi:

```
loss_totale = loss_mask_RCNN + dice_loss
```

La **Dice Loss** è calcolata in modo differenziabile su ogni maschera predetta e *ground truth*, usando una versione smussata per evitare divisioni instabili.

2.4 Training: Iperparametri e Modalità

Il training è effettuato su GPU con *mixed precision* (AMP abilitato), **batch size** effettivo di 14 immagini per iterazione, per un massimo di **8000 iterazioni**.
Gli **iperparametri** chiave sono:

- **Learning rate iniziale**: 0.004, con **scheduler multistep**
- **Warmup**: 500 iterazioni con metodo **lineare**
- **Momentum**: 0.9
- **Weight decay**: 0.0001
- **Scheduler**: step a 3000 e 6000 iterazioni, con **gamma** 0.6

Le immagini di input e output sono tutte mantenute a dimensione fissa di 512x512. L'intero addestramento è gestito con **DefaultTrainer** esteso, integrando **mapper** e **hook personalizzati**.

3. Risultati

Per valutare le prestazioni del nostro modello di segmentazione **Mask R-CNN**, abbiamo considerato un approccio in cui, per ciascuna immagine del set di test, viene selezionata una singola predizione: quella con la confidenza più alta, purché superiore a una **soglia del 70%**. Se la classe associata alla predizione coincide con quella *ground truth* dell'immagine, vengono calcolate una serie di metriche di qualità della segmentazione. In caso contrario (classe errata o assenza di predizioni sopra soglia), l'immagine è comunque conteggiata nei totali ma esclusa dal calcolo delle metriche.

Le metriche così ottenute sono poi aggregate per classe di *ground truth*, per valutare il comportamento del modello su ciascun tipo di tumore (Glioma, Meningioma, Pituitario). Questo metodo consente di isolare la qualità della segmentazione dai problemi di classificazione, focalizzandosi solo sulle predizioni corrette.

3.1 Metriche utilizzate

Le seguenti metriche sono state calcolate a livello pixel tra la maschera predetta e quella ground truth, dove: **TP** - *True Positive*, **FP** - *False Positive*, **TN** - *True Negative*, **FN** - *False Negative*, **d(a,b)** - distanza euclidea tra i pixel a e b.

Metrica	Descrizione sintetica	Formula
IoU	Intersezione su unione (Jaccard Index)	`IoU = TP / (TP + FP + FN)`
Dice	Coefficiente di similarità Dice	`Dice = 2*TP / (2*TP + FP + FN)`
Precision	Accuratezza delle predizioni positive	`Precision = TP / (TP + FP)`
Recall	Capacità di identificare i pixel positivi	`Recall = TP / (TP + FN)`
Specificity	Capacità di escludere i pixel negativi	`Specificity = TN / (TN + FP)`
Balanced Accuracy	Media tra Recall e Specificity	`BalancedAcc = (Recall + Specificity) / 2`
Hausdorff (px)	Massima distanza tra i bordi delle maschere	`H(A, B) = max(sup_a∈A inf_b∈B d(a,b), sup_b∈B inf_a∈A d(b,a))`

3.2 Risultati quantitativi

Per ogni immagine, è stata considerata una sola maschera predetta (la più confidente, se presente).
Le metriche sono state calcolate solo per le predizioni con classe corretta, e successivamente aggregate per classe di *ground truth*. Questo approccio garantisce una valutazione accurata della qualità della segmentazione in assenza di errori di classificazione.

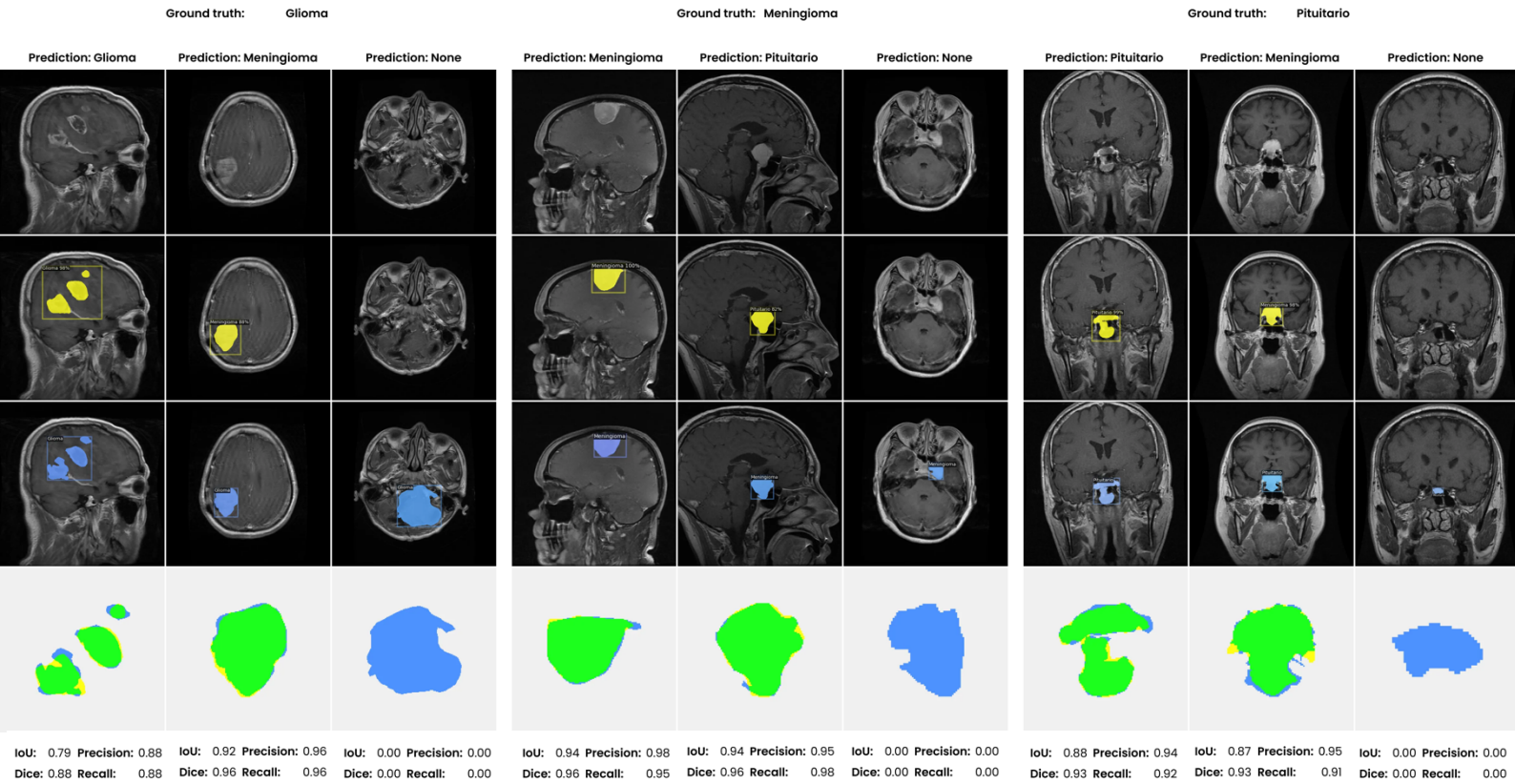
Ground truth	Predetto	Errato	Non predetto	IoU	Dice	Precision	Recall	Specificity	Balanced Acc	Hausdorff (px)
Glioma	235	3	16	0.7200	0.8035	0.8230	0.8211	0.9968	0.9089	29.93
Meningioma	303	2	1	0.8990	0.9442	0.9571	0.9374	0.9988	0.9681	9.07
Pituitario	293	4	3	0.8145	0.8905	0.8703	0.9344	0.9988	0.9666	10.96

Di seguito sono sintetizzate le performance complessive del modello in termini di rilevamento e segmentazione. Le metriche riportate descrivono la capacità del sistema di identificare tumori e di segmentarli accuratamente, indipendentemente dalla corretta classificazione della loro categoria.

Metriche globali		Metriche aggregate per tumori rilevati						
Tumori rilevati	**Tumori non rilevati**	**IoU**	**Dice**	**Precision**	**Recall**	**Specificity**	**Balanced Acc**	**Hausdorff (px)**
840	20	0,8189	0,8857	0,8888	0,9036	0,9982	0,9509	15,5481

3.3 Risultati qualitativi

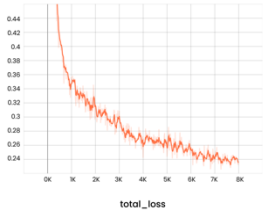
Per una valutazione qualitativa del modello, sono stati selezionati alcuni esempi rappresentativi suddivisi per classe (Glioma, Meningioma, Pituitario) e per tipologia di risultato: Segmentazione corretta (*True Positive*), Classificazione errata (*False Positive*), Tumore non riconosciuto (*False Negative*).
Ogni colonna rappresenta uno di questi tre scenari per una specifica classe tumorale.
Ogni riga mostra, nel seguente ordine: l'immagine originale, la segmentazione predetta dal modello, la maschera reale (*ground-truth*) e la sovrapposizione delle due maschere.
Al di sotto di ciascuna colonna sono riportati quattro indicatori quantitativi (**IoU**, **Dice**, **Precision**, **Recall**) calcolati sulla maschera segmentata, indipendentemente dalla correttezza della classificazione.

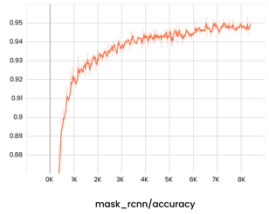


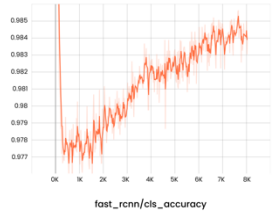
4. Discussione

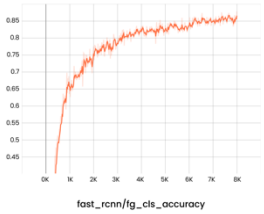
L'analisi dei risultati sul dataset di test mostra che il modello è particolarmente efficace nel mantenere un'elevata **specificity**, frequentemente superiore a **0.99**. Questo riflette una tendenza a evitare falsi positivi e a non segmentare erroneamente tessuti sani, favorito sia dalla *pipeline* di *augmentation* conservativa sia dall'introduzione della componente di **Dice Loss**, che migliora la qualità della segmentazione su regioni sbilanciate. Tuttavia, questa

caratteristica si accompagna a una **recall** non sempre soddisfacente, con **20** casi di completa assenza di predizione. È particolarmente significativo osservare che ben **16** di questi **20** casi riguardano tumori della classe Glioma, rappresentando oltre il **75%** delle mancate rilevazioni. Questo dato suggerisce che il principale fattore limitante non è tanto la dimensione o il contrasto delle lesioni, quanto piuttosto la loro elevata eterogeneità morfologica, tipica dei gliomi, che presentano contorni meno definiti e pattern meno regolari rispetto a meningiomi e pituitari.
L'andamento del training mostra una progressiva riduzione della *loss* complessiva, con valori di `mask_rcnn/accuracy` stabilmente superiori a **0.95** e un `fast_rcnn/cls_accuracy` vicino a **0.98**, indicando una buona capacità del modello nell'apprendere sia le caratteristiche semantiche che quelle spaziali delle lesioni.
L'andamento della `total_loss` e in particolare della `loss_cls`, ancora non pienamente convergenti, suggeriscono che il modello non ha esaurito la propria capacità di apprendimento. Questo giustifica l'opportunità di prolungare il training oltre le **8000** iterazioni, accompagnato da un ulteriore abbassamento del *learning rate*.
Quando il modello riesce a rilevare la presenza del tumore, indipendentemente dall'accuratezza nella classificazione della classe, la qualità della segmentazione risulta molto buona, con valori medi di **IoU** superiori a **0.81**, **Dice** intorno a **0.89**, e una **Hausdorff distance** contenuta, che conferma la coerenza spaziale delle maschere rispetto al *ground truth*. Questo significa che il modello possiede una buona capacità di apprendere la morfologia generale delle masse tumorali, a condizione che riesca effettivamente a identificarne l'esistenza nell'immagine.
D'altra parte, l'elevato numero di falsi negativi nella classe Glioma evidenzia un limite strutturale del modello nel generalizzare su morfologie tumorali particolarmente variabili. Il fenomeno è confermato anche da un *False Negative Rate* sensibilmente più elevato rispetto alle altre classi, a sottolineare come il problema sia sistemico e non legato a specifici casi isolati. Le metriche raccolte durante il training mostrano inoltre un certo scollamento tra l'accuratezza globale e quella relativa agli oggetti tumorali (`fast_rcnn/fg_cls_accuracy`), indicando che anche la componente di classificazione risente di questa complessità. Questo fenomeno potrebbe essere almeno in parte mitigato adottando una **backbone** più espressiva, come ***ResNet-101***, ***ResNeXt-101*** o architetture più recenti come ***ConvNeXt*** o ***Swin Transformer***, in grado di catturare pattern morfologici più articolati e relazioni spaziali non locali.









5. Conclusioni

Il modello **Mask R-CNN** addestrato su *slice* **2D** ha dimostrato buone capacità di segmentazione, con metriche di **IoU** e **Dice** solide (**0.81** e **0.89**). Tuttavia, la difficoltà nel rilevare morfologie complesse, in particolare nei gliomi, evidenzia i limiti di un approccio **2D** *slice-based*. Il lavoro svolto rappresenta un valido punto di partenza per sviluppi futuri in ambito volumetrico **3D** e modelli più avanzati.

5.1 Proposte future

Un'evoluzione naturale di questo lavoro consiste nell'adozione di modelli volumetrici **3D**, capaci di elaborare simultaneamente l'intero *stack* di **155** *slice* per paziente, acquisite in quattro diverse sequenze di risonanza magnetica (***T1***, ***T1-contrast***, ***T2***, ***FLAIR***). Sebbene ciascuna slice rappresenti un'immagine **2D**, l'insieme delle sequenze fornisce un contesto tridimensionale estremamente ricco, contenente informazioni spaziali, morfologiche e funzionali. I modelli **3D** sono in grado di sfruttare la continuità lungo l'asse *interslice* per migliorare la coerenza spaziale delle predizioni, ridurre segmentazioni incoerenti e catturare dettagli anatomici distribuiti su più piani.
Uno dei principali vantaggi di questo approccio è la maggiore sensibilità nel rilevare tumori di piccole dimensioni, lesioni frammentarie o aree pretumorali, che potrebbero non essere visibili ad occhio nudo o risultare impercettibili su una singola *slice* **2D**. Il contesto volumetrico, infatti, consente al modello di comprendere le strutture nel loro insieme, migliorando la capacità di identificare pattern deboli o sfumati ma clinicamente rilevanti.
In tale scenario, è particolarmente utile l'impiego di funzioni di perdita avanzate, come quella adottata da **MUNet**, che integra **IoU Loss**, **Dice Loss** e una **Boundary Loss** orientata alla precisione dei contorni. L'obiettivo non è solo aumentare la sovrapposizione tra maschere predette e *ground truth*, ma anche garantire un'aderenza più accurata ai margini reali delle lesioni, aspetto cruciale in ambito radioterapico e chirurgico.
Ulteriori sviluppi includono l'utilizzo di modelli pre-addestrati su dataset ampi come **BRaTS**, che oltre alle immagini forniscono informazioni cliniche supplementari (es. trattamenti ricevuti, aspettativa di sopravvivenza, stato molecolare), abilitando una possibile estensione del modello verso approcci *multi-task*. Questi modelli potrebbero prevedere simultaneamente la segmentazione, la classificazione del tipo tumorale, e persino stadi prognostici, con impatti significativi in termini di personalizzazione terapeutica.