



UNIMORE

UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Facoltà di Medicina e Chirurgia

Corso di Laurea Magistrale in Medicina e Chirurgia

Dipartimento di Scienze Biomediche, Metaboliche e Neuroscienze

Struttura Complessa di Radiologia

**MACHINE LEARNING PER IL LESION MAPPING NELLE
LESIONI GLIALI DIFFUSE DELL'ADULTO: CONFRONTO
PRELIMINARE DELLA PERFORMANCE DELL'ALGORITMO
ATLAS-BASED CON I REFERTI CLINICI**

Relatore:

Chiar.ma Prof.ssa ANNARITA PECCHI

Correlatore:

Dr.ssa GIULIA SPRUGNOLI

Candidato:

MATTEO GIALDI

Anno Accademico 2025-2026

ABSTRACT	1
INTRODUZIONE	3
NEOPLASIE DEL SISTEMA NERVOSO CENTRALE	3
Epidemiologia	3
Eziologia	4
Presentazione clinica del paziente con tumore cerebrale	5
Cefalea	5
Crisi epilettiche	5
Alterazioni cognitivo-comportamentali	6
Sintomi focali	6
Sintomi generalizzati	6
Classificazione delle neoplasie del sistema nervoso centrale	7
Status mutazionale di IDH	9
Metilazione del gene MGMT	10
Caratteristiche salienti dei gliomi diffusi dell'adulto	11
Astrocitoma IDH-mutato	11
Glioblastoma IDH-wildtype	11
Oligodendroglioma IDH-mutato e 1p/19q codeleto	12
Trattamento	13
Chirurgia	13
Radioterapia	13
Chemioterapia	14
Terapia di supporto e cure palliative	14
Prognosi	14
ARTIFICIAL INTELLIGENCE (AI)	16
Intelligenza artificiale, machine learning e deep learning	16
Machine learning	16
Deep learning	16
ML vs DL: il caso della <i>black-box</i>	17
Applicazioni in radiologia	18
Segmentazione automatica delle immagini	18
Ottimizzazione delle immagini	18
Radiomica	19
Facilitazione del flusso di lavoro	19
Applicazioni AI in neuro-oncologia	20
Automated Radiology Report Generation (ARRG)	21
Aspetti etici	22
Sostituzione medica	23
Algoritmi	23
SCOPO DELLO STUDIO	25
MATERIALI E METODI	27
Disegno dello studio	27
Campione	27
Criteri di inclusione ed esclusione	27

Protocollo d'imaging	28
Selezione dei referti	28
Pipeline dell'algoritmo e dell'autogenerazione dei referti	29
Segmentazione delle aree tumorali	29
Registrazione dell'atlante anatomico sul paziente	30
Identificazione della sede di origine	31
Identificazione delle aree coinvolte	31
Valutazione delle aree eloquenti	32
Valutazione di metriche geometriche aggiuntive	33
Attribuzione del tipo di lesione	34
Fase generativa di json e testo	34
Annotazioni manuali all'algoritmo	34
Pre-elaborazione dei referti	36
Metriche di esecuzione del confronto	37
RISULTATI	39
Analisi descrittiva della popolazione	39
Analisi del contenuto dei referti	40
Output annotazioni manuali dell'algoritmo contro algoritmo <i>raw</i>	40
Output annotazioni manuali dell'algoritmo contro referti reali	42
Numero dei centri, lato e macroarea	42
Diametro e dimensioni	42
DISCUSSIONE	44
Performance nella classificazione del numero di lesioni	44
Localizzazione anatomica e lateralizzazione	45
Aree eloquenti, tumor core e edema	45
Valutazione di diametri e dimensioni	48
Implicazioni per la refertazione automatica	48
Limiti dello studio e prospettive future	50
CONCLUSIONI	51
BIBLIOGRAFIA	52

ABSTRACT

Introduzione

I gliomi diffusi dell'adulto — astrocitoma, oligodendroglioma e glioblastoma — rappresentano le neoplasie maligne primitive del SNC più frequenti, con prognosi variabile e gestione multidisciplinare complessa. La refertazione radiologica in linguaggio libero è soggetta a variabilità inter-osservatore e a mancanza di standardizzazione. L'Automated Radiology Report Generation (ARRG) mediante intelligenza artificiale potrebbe rispondere a queste criticità, garantendo output strutturati e riproducibili. In questo contesto, è stato sviluppato in collaborazione con il laboratorio AlmageLab-Zip di UNIMORE un algoritmo deterministico di *machine learning atlas-based*, alternativo al *deep learning*, che tenta di rispondere al problema della *black-box* e della carenza di dati annotati.

Materiali e metodi

Lo studio retrospettivo ha analizzato i dati di 48 pazienti con prima diagnosi di glioma diffuso dell'adulto, sottoposti a RMN preoperatoria presso la Neuroradiologia dell'AOU di Modena (settembre 2020 – giugno 2025). L'algoritmo, a partire da maschere di segmentazione tumorale ottenute mediante *deep learning* e rifinite da neuroradiologi esperti, registra l'atlante anatomico cerebrale SRI24 sullo spazio del paziente tramite ANTs, identificando sede d'origine, aree coinvolte, aree eloquenti e metriche geometriche quantitative. L'output viene tradotto in un file *json* strutturato e in un testo descrittivo. Le annotazioni manuali dell'operatore hanno consentito la correzione dell'output algoritmico, successivamente confrontato con 45 referti clinici testuali mediante metriche di *accuracy* (proporzione delle classificazioni corrette), *precision* (proporzione delle classificazioni positive effettivamente positive), *recall* (proporzione di positivi reali classificati correttamente come positivi), *F1 score* (misura sintetica che bilancia *precision* e *recall*) e *prevalence* (intesa come numero dei campi presenti in media su tutti i campi possibili).

Risultati

Nel confronto algoritmo-annotazioni manuali, l'*accuracy* nella classificazione del numero di lesioni è risultata del 70,49%, quella nella lateralità del 97,92% e quella

nell'area d'origine del 65,18% su 281 aree annotate. Per le aree eloquenti, la *precision* ha raggiunto il 91,89% con *recall* del 73,91%. Per il *tumor core* il confronto fra le aree coinvolte ha mostrato *precision* 62,75% e *recall* 84,91% (F1: 72,16%); per l'edema, *precision* 63,40% e *recall* 95,10%. Nel confronto con i referti clinici, il *matching* per macroarea ha raggiunto il 100%, per lateralità il 93% e per numero di lesioni il 76,7%. L'errore relativo sulle dimensioni è risultato del 19,9% considerando almeno due assi spaziali, salendo al 51,6% con un singolo asse.

Discussione

L'algoritmo mostra performance clinicamente rilevanti nella lateralità e nella localizzazione per macroarea, con elevata sensibilità per edema e *tumor core*. La *precision* del 91,89% per le aree eloquenti è particolarmente significativa per la pianificazione neurochirurgica, sebbene il *recall* del 73,91% evidenzia margini di miglioramento. L'errore sul diametro massimo lungo un singolo asse riflette la differenza tra la valutazione tridimensionale dell'algoritmo rispetto a quella clinica routinaria. La disomogeneità dei referti in linguaggio libero conferma la necessità di standardizzazione; l'ARRG si configura come uno dei metodi possibili.

Conclusione

I risultati preliminari supportano la validità dell'approccio deterministico *atlas-based* come strumento di supporto alla refertazione neuroradiologica strutturata, con il clinico nel ruolo irrinunciabile di supervisore e validatore.

INTRODUZIONE

NEOPLASIE DEL SISTEMA NERVOSO CENTRALE

Epidemiologia

Le neoplasie del sistema nervoso centrale (SNC) sono patologie relativamente poco frequenti. Tuttavia, per quanto rare, sono gravate da grande mortalità e disabilità: nell'ambito dei tumori cerebrali il classico rapporto tra malignità, dimensione e sintomatologia non è rispettato: anche una lesione di piccole dimensioni, indipendentemente dal grado di malignità, può determinare deficit rilevanti se localizzata in aree eloquenti.

Nel 2016 sono stati registrati circa 330.000 nuovi casi e 227.000 decessi attribuibili a tumori del SNC nel mondo, con un tasso di incidenza età-standardizzato di 4,63 per 100.000 persone-anno. Per quanto riguarda la distribuzione geografica, l'Asia orientale presentava l'incidenza più alta nel 2016 (circa 108.000), seguita dall'Europa occidentale e dall'Asia meridionale, mentre gli Stati con il più alto numero assoluto di casi erano Cina, Stati Uniti e India. I tassi di incidenza età-standardizzati più alti si osservano nei paesi ad alto SDI (*Socio-demographic Index*), con i valori massimi registrati negli Stati del nord-Europa (Islanda, Danimarca, Norvegia e Finlandia) e in Lussemburgo. Questa variabilità geografica può essere una verosimile conseguenza delle differenze nell'accesso alla diagnostica, a questi vanno sommati possibili fattori ambientali e genetici ancora non del tutto chiariti.¹

In Italia, i dati più solidi derivano da registri tumori regionali o provinciali: nel caso specifico dell'Emilia-Romagna, coi dati aggiornati al 2020, l'incidenza è stimata intorno 471 casi per anno, di cui 67 sono attribuibili alla sola provincia di Modena. Il tasso d'incidenza è tendenzialmente maggiore per il sesso maschile.²

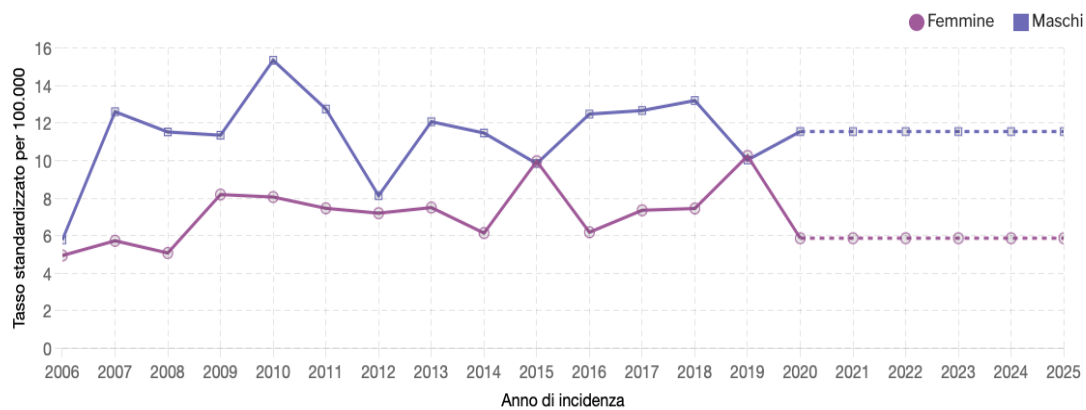


Figure 1: Tasso di incidenza standardizzato per età per anno di incidenza dei tumori cerebrali in rapporto ai due sessi nella provincia di Modena.²

Eziologia

In termini eziologici, le uniche associazioni consistenti emerse dagli studi epidemiologici sono quella positiva con le radiazioni ionizzanti e quella negativa con le condizioni della sfera atopica, quali asma, eczema e allergie alimentari. Molti altri fattori come le radiofrequenze dei telefoni cellulari, l'acido acetil-salicilico, le alterazioni ormonali, i campi magnetici a bassa frequenza, i pesticidi e le esposizioni a tossici a genesi industriale non hanno mostrato associazioni consistenti con il rischio di tumori del SNC.³ Nello specifico, da una meta-analisi è emerso che una storia di allergia si associa a un rischio di insorgenza di glioma inferiore di circa il 40%, mentre asma ed eczema atopico si associano a una riduzione del 30%.⁴

In termini genetici, da una serie di studi raccolti nella review di Ostrom et al. nel 2014, le sindromi monogeniche associate a un aumentato rischio di glioma includono la neurofibromatosi di tipo 1 e 2, la sclerosi tuberosa, la sindrome di Li-Fraumeni, la sindrome di Lynch, la FAMMM (*familial atypical multiple mole melanoma*) e la malattia di Ollier. Tuttavia, questi disordini genetici rendono conto di una piccola frazione dei casi di glioma nell'adulto. Sul versante molecolare, la stessa review mostra che tra i più importanti biomarcatori prognostici favorevoli ci sono la mutazione *gain of function* di IDH1/IDH2, la metilazione del promotore del gene MGMT, e la codelezione cromosomica 1p/19q. Studi di associazione *genome-wide* (GWAS) hanno identificato 8 polimorfismi a singolo nucleotide *germline* in 7

geni (TERT, RTEL1, EGFR, TP53, CDKN2B, PHLDB1 e CCDC26) associati a un aumentato rischio di sviluppo di specifici sottotipi di glioma.³

Presentazione clinica del paziente con tumore cerebrale

I gliomi possono manifestarsi con un ampio spettro di sintomi, determinati principalmente dalla localizzazione della lesione, dalla velocità di crescita e dalla presenza di effetto massa sul parenchima circostante. L'edema perilesionale, l'ostruzione liquorale e la presenza di emorragia intratumorale rappresentano ulteriori fattori che contribuiscono al quadro clinico d'esordio. I sintomi possono essere distinti in generalizzati, legati genericamente all'aumento della pressione intracranica, e focali, espressione del coinvolgimento di specifiche strutture del sistema nervoso centrale.⁵

Cefalea

La cefalea è tra i sintomi più comuni, sebbene raramente si presenti come unico disturbo. Al momento della diagnosi ne soffre la maggioranza dei pazienti, pur presentandosi come sintomo isolato solo in una minoranza dei casi. Il quadro cefalgico nei gliomi si discosta spesso dalla presentazione cosiddetta classica: il dolore tende ad essere intermittente piuttosto che continuo, gravativo e bilaterale, di intensità moderata, con nausea e vomito presenti solo in una minoranza dei casi. La lateralità della cefalea può tuttavia avere valore localizzatorio, poiché in una larga maggioranza dei pazienti con cefalea monolaterale essa risulta ipsilaterale alla sede del tumore.⁵

Crisi epilettiche

Le crisi epilettiche rappresentano un altro sintomo cardine della presentazione clinica dei gliomi. La loro incidenza varia tra il 30% e il 100% a seconda del tipo istologico, con i tumori a crescita lenta che risultano i più epilettogeni. Più del 50% dei pazienti con glioma sperimenta crisi ricorrenti nel corso della malattia.⁵ Un contributo importante alla comprensione del meccanismo epilettogeno viene dallo studio di Chen et al., che ha analizzato una coorte di 712 pazienti con gliomi di grado II–IV: i pazienti con mutazione di IDH1 presentavano crisi epilettiche preoperatorie nel 59–74% dei casi, rispetto al 18–34% dei pazienti con IDH1 wild-type. Il

meccanismo proposto chiama in causa il metabolita D-2-idrossiglutarato (D2HG), che presenta una marcata somiglianza strutturale con il glutammato e risulta in grado di aumentare l'attività elettrica dei neuroni corticali attraverso l'attivazione dei recettori NMDA.⁶

Alterazioni cognitivo-comportamentali

Le alterazioni cognitive e comportamentali costituiscono un altro importante dominio sintomatologico. Irritabilità, cambiamenti della personalità, labilità emotiva, apatia e rallentamento psicomotorio sono osservati nel 16–34% dei pazienti e riflettono tipicamente il coinvolgimento di strutture profonde, quali le fibre talamocorticali, il corpo calloso e la formazione reticolare. Le lesioni del lobo frontale possono produrre quadri distinti a seconda della sede specifica: i tumori dorso-laterali tendono a causare apatia e abulia, mentre quelli orbito-frontali si associano a disinibizione e comportamento impulsivo.⁵

Sintomi focali

I sintomi focali variano in funzione della localizzazione del tumore. Le lesioni del lobo temporale predispongono in modo particolare alle crisi epilettiche parziali complesse e ai deficit del campo visivo per interessamento delle radiazioni ottiche, con tipici difetti dei quadranti superiori omonimi da coinvolgimento dell'ansa di Meyer. Le lesioni parietali si manifestano con disturbi sensitivi, aprassia e, nelle lesioni dell'emisfero dominante, con afasia fluente. I tumori del lobo frontale possono causare deficit motori, disturbi dell'andatura e alterazioni del linguaggio espressivo, con afasia non fluente nelle lesioni del giro frontale inferiore posteriore. I tumori della fossa posteriore e del cervelletto tendono a presentarsi con idrocefalo ostruttivo, atassia e segni da coinvolgimento dei nervi cranici, mentre le lesioni del tronco encefalico producono tipicamente quadri di paralisi dei nervi cranici.⁵

Sintomi generalizzati

I segni di ipertensione endocranica (tra cui papilledema, nausea e vomito) sebbene oggi più rari grazie alla disponibilità del *neuroimaging* precoce, rimangono espressione di una malattia in stadio avanzato o di un'ostruzione delle vie liquorali.⁵ Infine, il tromboembolismo venoso rappresenta una complicanza sistemica di

rilievo, con un rischio continuo stimato intorno all'1,5–2% per mese di sopravvivenza nei pazienti con glioblastoma, tra i più elevati in assoluto in campo oncologico.⁷

Classificazione delle neoplasie del sistema nervoso centrale

Clinicamente, le neoplasie cerebrali si dividono innanzitutto in intra-assiali ed extra-assiali. Mentre le prime originano a partire da encefalo e midollo spinale, le seconde originano da strutture extra-encefaliche come nervi cranici e meningi. A loro volta, i tumori intra-assiali possono essere suddivisi in primitivi e secondari: i primitivi sono quelli che derivano da cellule appartenenti propriamente al tessuto nervoso, i secondari sono da imputare alla metastatizzazione di tumori appartenenti ad altri organi. Epidemiologicamente, le metastasi rappresentano più della metà delle neoplasie cerebrali e la loro incidenza è in aumento; questo incremento è verosimilmente derivabile dall'allungamento progressivo della sopravvivenza dei pazienti oncologici e dall'affinamento delle pratiche diagnostiche. Le più frequenti sono quelle a partire dal tumore del polmone, della mammella e melanoma.⁸

Fino a poco tempo fa, la classificazione dei gliomi era esclusivamente di tipo morfologico. Ad oggi, si fa riferimento alla classificazione WHO del 2021 (V edizione) che consente un'integrazione sistematica tra il dato di biologia molecolare e quello istologico. I tumori sono dapprima organizzati sulla base della sede d'origine e successivamente per famiglia istologica di appartenenza: gliomi, glioneuroni e neuroni; tumori del plesso corioide; tumori embrionali; tumori pineali; tumori dei nervi cranici; meningiomi; tumori mesenchimali; tumori melanocitici; tumori ematolinfopoietici; tumori germinali; tumori della regione sellare; metastasi e sindromi genetiche. L'integrazione con dati molecolari come il sequenziamento e i profili di metilazione del DNA ha consentito la definizione di nuovi sottotipi e un affinamento della prognosi.⁹

Uno studio del *Central Brain Tumor Registry of the United States* (CBTRUS) ha mostrato come tra le neoplasie del SNC i gliomi ricoprano complessivamente il 22,2% della totalità. Tra i tumori maligni, il glioblastoma (GBM) è il tumore più frequente rappresentando da solo il 13.7% di tutti i tumori dell'SNC e il 52.2% delle neoplasie maligne. Tra i tumori benigni, i meningiomi risultano i più frequenti

rappresentando il 42.6% di tutti i tumori dell'SNC (inclusendo il meningioma maligno) e il 57.4% di tutti i tumori benigni.¹⁰

Category	Gliomas, glioneuronal tumours and neuronal tumours						Choroid plexus tumours	Embryonal tumours		Pineal tumours	Mesenchymal/Non-Meningothelial Tumours	Tumours of the sellar region
	Adult-type diffuse gliomas	Paediatric-type high-grade diffuse gliomas	Paediatric-type low-grade diffuse gliomas	Circumscribed astrocytic gliomas	Glioneuronal and neuronal tumours	Ependymal tumours		Medullo-blastoma	Other CNS Embryonal tumours		Uncertain differentiation	
Types	Astrocytoma, IDH-mutant	Diffuse midline glioma, H3 K27-altered	Diffuse astrocytoma, MYB or MYBL1-altered	High-grade astrocytoma with piloid features	Diffuse glioneuronal tumor with oligodendroglioma-like features and nuclear clusters (DGONC)	Supratentorial ependymoma, ZFTA fusion-positive	Choroid plexus papilloma	MB, WNT-activated	CNS neuroblastoma, FOXR2-activated	Desmoplastic myxoid tumor of the pineal region, SMARCB1-mutant	Intracranial mesenchymal tumor, FET-CREB fusion positive (provisional type)	Pituitary blastoma
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	Diffuse midline glioma, H3 G34-mutant	Polymorphous low-grade neuroepithelial tumour of the young (PLNTY)	Pilocytic astrocytoma	Myxoid glioneuronal tumor	Supratentorial ependymoma, YAP1 fusion-positive	Atypical choroid plexus papilloma	MB, SHH-activated, TP53-wildtype	CNS tumour with BCOR internal tandem duplication and the provisional type (CNS tumour BCOR ITD)	Pineocytoma	CIC-rearranged sarcoma	Adamantinomatous craniopharyngioma
	Glioblastoma, IDH-wildtype	Diffuse paediatric-type high-grade glioma, H3-wildtype/IDH-wildtype	Diffuse low-grade glioma, MAPK pathway-altered	Pleomorphic xanthoastrocytoma	Multinodular and vacuolating neuronal tumor (MVNT)	Posterior fossa group A (PFA) ependymoma	Choroid plexus carcinoma	MB, SHH-activated, TP53-mutant	Cribiform neuroepithelial tumour (CRINET)	Pineal parenchymal tumour of intermediate differentiation	Primary intracranial sarcoma, DICER1-mutant	Papillary craniopharyngioma
		Infant-type hemispheric glioma	Angiocentric glioma	Subependymal giant cell astrocytoma	Diffuse leptomeningeal glioneuronal tumor	Posterior fossa group B (PFB) ependymoma		MB, non-WNT/non-SHH	Atypical teratoid/rhabdoid tumour (ATRT)	Pineoblastoma	Solitary fibrous tumour	Pituitary, granular cell tumour of the sellar region, and spindle cell oncocytoma
				Chordoid glioma	Ganglioglioma	Spinal ependymoma, MYCN-amplified		MB, histologically defined	Embryonal tumour with multi-layered rosettes (ETMR)	Papillary tumor of the pineal region	Ewing sarcoma	Pituitary adenomas
				Astroblastoma, MN1-altered	Desmoplastic infantile ganglioglioma/desmoplastic infantile astrocytoma	Myxopapillary ependymoma			CNS embryonal tumor			Pituitary neuroendocrine tumour (PitNET)
				Dysembryoplastic neuroepithelial tumor		Subependymoma						
				Papillary glioneuronal tumor								
				Rosette-forming glioneuronal tumor								
				Gangliocytoma								
				Dysplastic cerebellar gangliocytoma								
				Central neurocytoma								
				Extraventricular neurocytoma								
				Cerebellar liponeurocytoma								

Note: for Mesenchymal/Non-meningothelial tumours, only the group of "Uncertain differentiation" is shown.

Figure 2: classificazione WHO 2021 dei tumori cerebrali.¹¹

Il *grading* dei gliomi segue la suddetta classificazione e divide le neoplasie da un grado 1 a un grado 4 in numeri arabi. All'aumentare del grado, l'aggressività clinico-biologica di queste neoplasie aumenta: infatti, il grado riflette la biologia naturale del tumore soprattutto in relazione a potenziale di crescita e capacità di recidiva, integrando l'aspetto isto-morfologico e il dato molecolare.⁹ Questo ha consentito di ridurre l'eterogeneità diagnostica, migliorare la capacità di prevedere prognosi e risposta ai trattamenti e di ottenere popolazioni più omogenee ai trial clinici.¹²

Operativamente, nella classificazione WHO i gliomi diffusi vengono suddivisi in quattro gruppi principali: *adult-type diffuse gliomas*, *pediatric-type diffuse low-grade gliomas*, *pediatric-type diffuse high-grade gliomas* e *circumscribed astrocytic gliomas*. Nel contesto di questo lavoro, la trattazione verterà sui gliomi diffusi dell'adulto, composti da: astrocitoma IDH-mutato, oligodendroglioma IDH-mutato 1p/19q codeleto e glioblastoma IDH-wildtype.¹³ Questi sono tumori altamente

infiltranti, responsabili della maggior parte delle neoplasie maligne del SNC in età adulta e con prognosi molto variabile a seconda del sottotipo e del grado WHO.

Status mutazionale di IDH

L'innovazione recente più significativa per i gliomi è rappresentata dalla mutazione del gene IDH (isocitrato deidrogenasi) e dalla co-delezione dei cromosomi 1p19q: la presenza contemporanea di questi fattori identifica le forme oligodendrogliali; al contrario, la presenza della mutazione di IDH in assenza della co-delezione identifica le forme astrocitarie. Grazie a questo dato molecolare è stato possibile discernere tra astrocitoma, IDH-mutato, CNS WHO grado 4 – la cui origine deriva dalla trasformazione di un glioma a basso grado – e glioblastoma IDH-*wildtype* – che è insorto de novo. Complessivamente, le forme con IDH-mutato hanno un *outcome* più favorevole rispetto a quelle con IDH-*wildtype*.⁸

Uno studio retrospettivo su 66 pazienti, di cui 55 con glioblastoma IDH-*wildtype* e 11 con astrocitoma IDH-mutato CNS WHO grado 4, ha valutato per ciascuna neoplasia l'*absolute tumor blood flow* e il *relative tumor flow within the enhancing portion* mediante tecniche di *arterial spin labeling* (ASL) e ha estratto la *maximum necrosis area*, la percentuale di *cross-sectional necrosis area inside the enhancing lesions*, il *minimum e mean ADC* da sequenze post-contrasto T1w e da DWI. Si è notato che l'*absolute tumor blood flow*, il *relative tumor flow inside the enhancing lesions*, la *maximum necrosis area* e la *percentuale di cross-sectional necrosis area inside the enhancing lesions* sono significativamente più alti nei pazienti IDH-*wildtype* rispetto a quelli che mostrano la mutazione di IDH, a conferma della maggiore aggressività in assenza di mutazione.¹⁴

Tra le metodiche non invasive è possibile inserire anche la *magnetic resonance spectroscopy* (MRS) che si è mostrata estremamente sensibile e specifica nella caratterizzazione dello status mutazionale di IDH attraverso la misurazione di un suo oncometabolita: il 2-idrossiglutarato.¹⁵ Questo mostra come l'imaging RM/MRS può stimare o predire lo stato IDH, fermo restando che la diagnosi molecolare resta di pertinenza istopatologica.

Metilazione del gene MGMT

Un ruolo importante è legato anche alla metilazione del gene MGMT: è stato visto da uno studio che i pazienti con glioblastoma associato a metilazione del promotore di MGMT mostrano una migliore risposta alla temozolomide rispetto a coloro che non la presentano.¹⁶

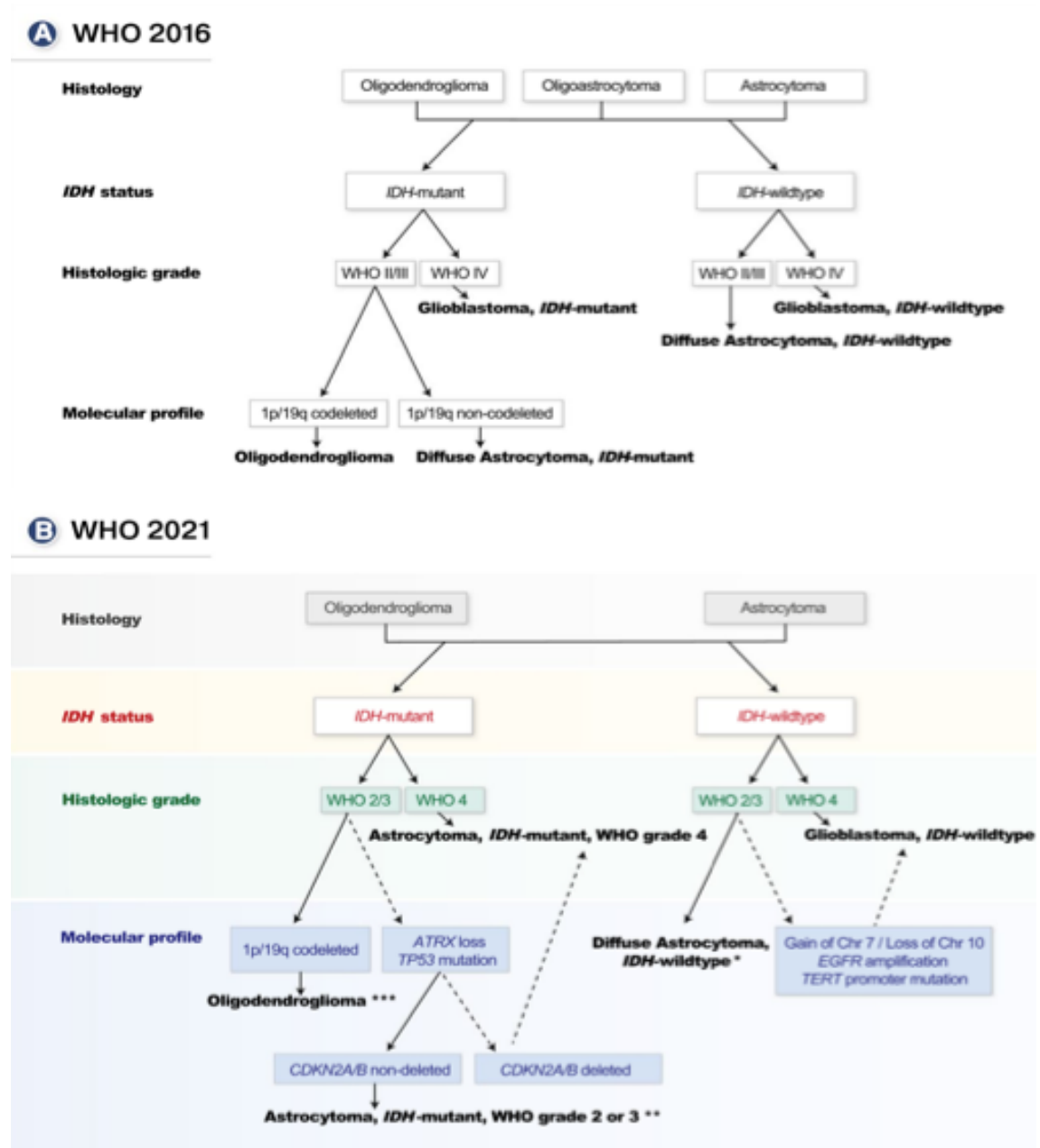


Figure 3: classifiche WHO 2016 e 2021 a confronto sulla base di IDH status, grado istologico e profilo molecolare.¹²

Caratteristiche salienti dei gliomi diffusi dell'adulto

Astrocitoma IDH-mutato

Per quanto riguarda l'astrocitoma IDH-mutato di grado 2, le caratteristiche imaging tipiche includono solitamente una massa sopratentoriale omogenea circoscritta, iperintensa in T2, a margini relativamente definiti, localizzata tendenzialmente nei lobi frontale o temporale e solitamente sprovvista di calcificazioni ed *enhancement* post-contrasto. Un reperto particolarmente utile è il cosiddetto *T2-FLAIR mismatch sign*, caratterizzato da iperintensità del segnale in T2 con segnale relativamente ipointenso nella maggior parte della lesione in FLAIR, fatta eccezione per un alone periferico iperintenso.¹³ Tra i gliomi a basso grado, il *T2-FLAIR mismatch sign* si è rivelato estremamente specifico per il sottotipo molecolare IDH-mutante non codeleto.¹⁷

L'astrocitoma IDH-mutato di grado 3 può essere radiologicamente indistinguibile dal grado 2, sebbene possa presentare eterogeneità del segnale in T2, *enhancement* e valori elevati di volume ematico cerebrale relativo massimo (rCBV).¹³

Glioblastoma IDH-wildtype

Un glioma diffuso astrocitario IDH-*wildtype* dopo un'analisi istologica e molecolare che escluda la mutazione di IDH può essere classificato come glioblastoma CNS WHO grado 4 in presenza di anche solo uno dei seguenti criteri: presenza di proliferazione microvascolare, necrosi, mutazione del promotore di TERT, amplificazione del gene EGFR o alterazioni del numero di copie cromosomiche +7/-10.¹³ Si può notare come anche la sola presenza di una delle alterazioni molecolari è sufficiente per poter definire un glioma diffuso come glioblastoma.

All'imaging tende ad avere un *enhancement* irregolare "ad anello" (*ring-enhancing*) con la presenza di necrosi centrale ipointensa *non enhancing*. Queste lesioni sono normalmente circondate da una porzione peritumorale infiltrativa iperintensa in T2-FLAIR. Alla perfusione si nota solitamente ipervascolarizzazione, mentre la DWI/ADC mostra ipercellularità.¹⁸

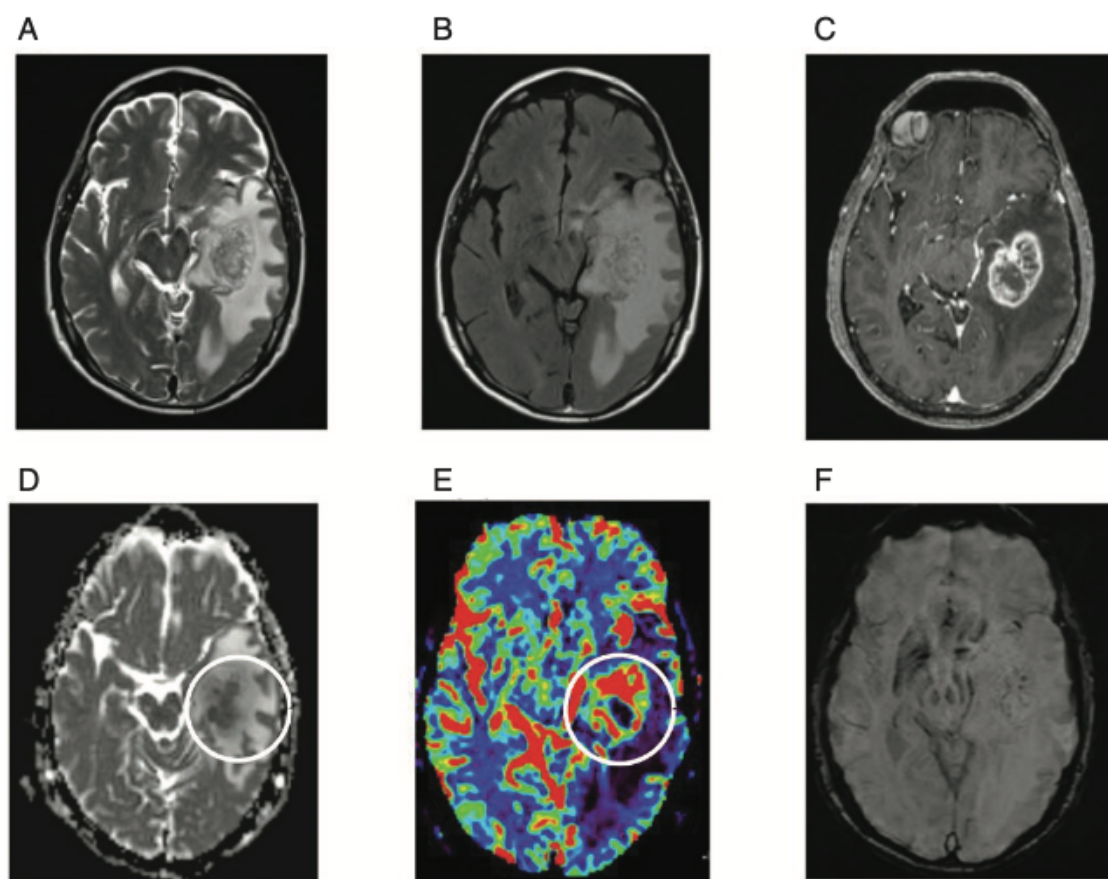


Figure 4:immagini che mostrano il glioblastoma IDH-wildtype secondo varie sequenze: (A), T2-FLAIR (B), T1w con contrasto (C), ipercellularità in mappa ADC (D), ipervascolarizzazione all'imaging di perfusione (E) e una microemorragia pesata all'imaging di suscettività magnetica (F).¹⁸

Oligodendroglioma IDH-mutato e 1p/19q codeleto

L'oligodendroglioma IDH-mutato e 1p/19q codeleto si localizza più spesso nel lobo frontale, mostra un aspetto morfologicamente eterogeneo con calcificazioni ed *enhancement* variabile.¹³ Da uno studio effettuato su 148 pazienti è emerso che questi oligodendrogliomi geneticamente definiti sono più spesso scarsamente circoscritti e mostrano praticamente sempre un'eterogeneità di segnale in T1 e T2. Oltretutto, nei pazienti con ADC disponibile, un valore inferiore di ADC risultava predittivo della codelezione 1p/19q.¹⁹

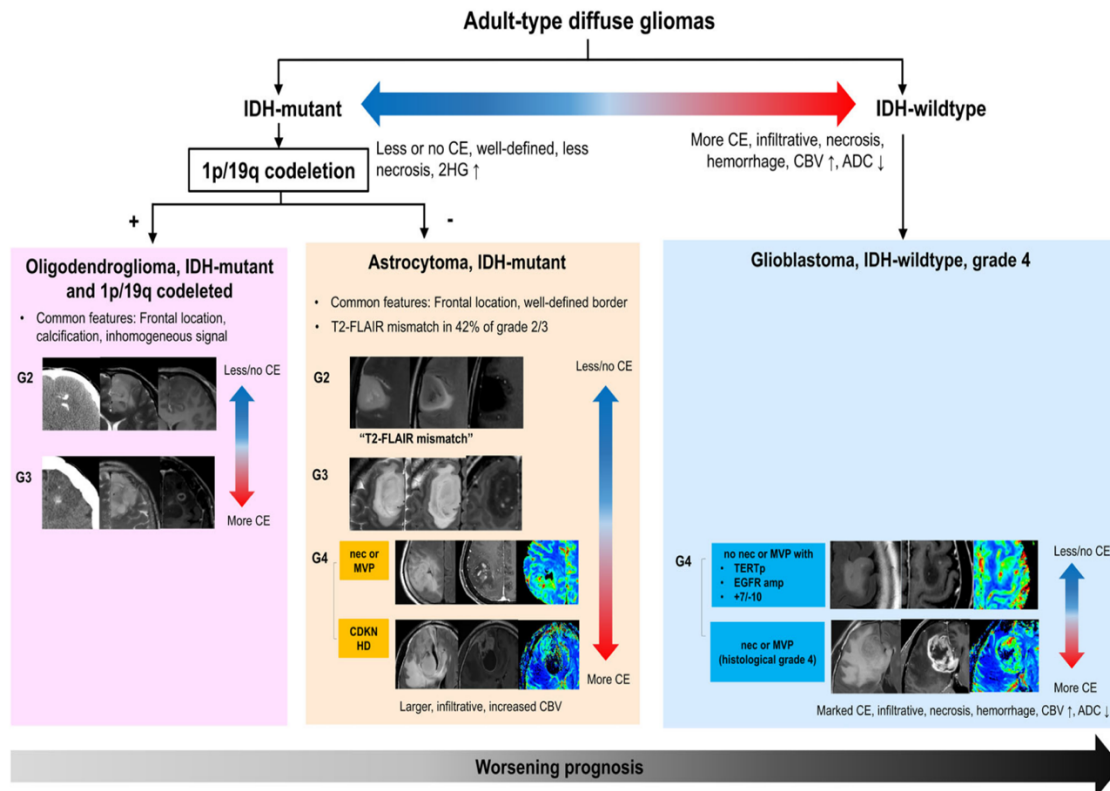


Figure 5: correlazioni morfologiche, molecolari e d'imaging nei gliomi diffusi dell'adulto.²⁰

Trattamento

Chirurgia

La chirurgia, soprattutto quella microchirurgica, costituisce il primo approccio terapeutico per la maggioranza dei gliomi. L'obiettivo è quello di arrivare a rimuovere il massimo volume tumorale possibile, che prende il nome di "massima resezione sicura": all'aumentare del volume tumorale residuo, infatti, la prognosi peggiora in tutte le categorie diagnostiche. Tuttavia, la priorità assoluta rimane quella di preservare la funzione neurologica che all'aumentare dell'estensione della resezione e a seconda della zona potrebbe risultare irrimediabilmente inficiata. Grazie a tecniche come la neuronavigazione e il monitoraggio intra-operatorio si è riusciti a ridurre il rischio di deficit post-operatori permanenti.^{21,22}

Radioterapia

Per quanto riguarda la radioterapia, le linee guida congiunte ASCO-SNO prevedono prevalentemente due tipi di utilizzi principali: il trattamento standard adiuvante in combinazione con temozolomide (TMZ) nel glioblastoma e l'utilizzo nelle

combinazioni curative nei gliomi IDH-mutati (in regimi che prevedono RT+PCV o RT+TMZ a seconda del sottotipo e della clinica).²³

Secondo le linee guida ASCO-SNO-ASTRO, nel caso specifico delle metastasi cerebrali, la RT è indicata in due principali regimi: la radiocirurgia stereotassica (SRT) e la radioterapia panencefalica (WBRT). L'SRT viene utilizzata per la malattia limitata (da 1 a 4 metastasi), mentre la WBRT viene utilizzata nel caso della malattia non trattabile con SRT o chirurgicamente. Inoltre, la RT non è indicata nel caso di malattia metastatica asintomatica con *Karnofski Performance Status* (KPS) ≤ 50 o <70 senza opzioni terapeutiche sistemiche praticabili.²⁴

Chemioterapia

Sempre le linee guida ASCO-SNO definiscono che nei tumori cerebrali primitivi dell'adulto la chemioterapia si può effettuare con due principali formulazioni farmaceutiche: la temozolomide (TMZ) e la polichemioterapia con procarbazine, lomustina e vincristina (PCV). La TMZ rappresenta lo standard in combinazione con la RT nel glioblastoma e la PCV in associazione a RT è fondamentale negli oligodendrogliomi 1p/19q codeleti. Inoltre, vi è una forte raccomandazione all'utilizzo adiuvante della chemioterapia in molti astrocitomi IDH-mutati.²³

Terapia di supporto e cure palliative

L'utilizzo dei corticosteroidi (in particolare il desametasone) con finalità antiedemigena, consente di ottenere sollievo dai sintomi legati all'edema vasogenico. Dal versante delle crisi epilettiche, l'utilizzo della terapia è raccomandato solo in prevenzione secondaria e viene condotta tendenzialmente con levitiracetam e lamotrigina.²⁵

Nell'ottimizzazione del processo di cura, le linee guida EANO raccomandano una rapida integrazione con le cure palliative precoci.²⁶

Prognosi

Uno studio di Ostrom et al. (2023), condotto su 8.651 pazienti diagnosticati nel 2018 attraverso il *National Cancer Database* statunitense, ha fornito le prime stime di *overall survival* (OS) a livello nazionale per i gliomi diffusi classificati secondo i criteri

molecolari integrati. I risultati mostrano differenze prognostiche marcate in base al tipo molecolare e al grado WHO. Il glioblastoma IDH-*wildtype* di grado 4 si conferma la forma più aggressiva e prognosticamente più infausta, con una OS a un anno del 53,7% e una OS mediana di 17,1 mesi nei tumori con metilazione del promotore di MGMT rispetto ai 12,4 mesi nei tumori non metilati; nei pazienti trattati con chemioradioterapia combinata, la OS mediana raggiungeva 21,8 mesi nei tumori metilati contro 14,1 mesi nei non metilati. Gli astrocitomi IDH-mutati hanno un profilo prognostico nettamente più favorevole e proporzionale al grado: la OS a un anno era del 98,0% per il grado 2, del 92,4% per il grado 3 e del 76,3% per il grado 4, con un'associazione significativa tra chemio-radioterapia e migliore sopravvivenza nei gradi 3 e 4. Gli oligodendrogliomi IDH-mutati e 1p/19q codeleti presentano la prognosi migliore in assoluto, con una OS a un anno del 97,9% per il grado 2 e del 94,4% per il grado 3. Trasversalmente a tutti i sottotipi, l'età inferiore a 50 anni, una maggiore estensione della resezione chirurgica e il trattamento con chemio-radioterapia si associano a una migliore sopravvivenza, mentre le dimensioni tumorali più elevate correlano a una prognosi peggiore.²⁷ Ne consegue che in termini di *disability adjusted life years* (DALYs), i gliomi siano prevalentemente legati alla componente di anni di vita persi (YLL), più che a quella di anni vissuti con disabilità (YLD).¹

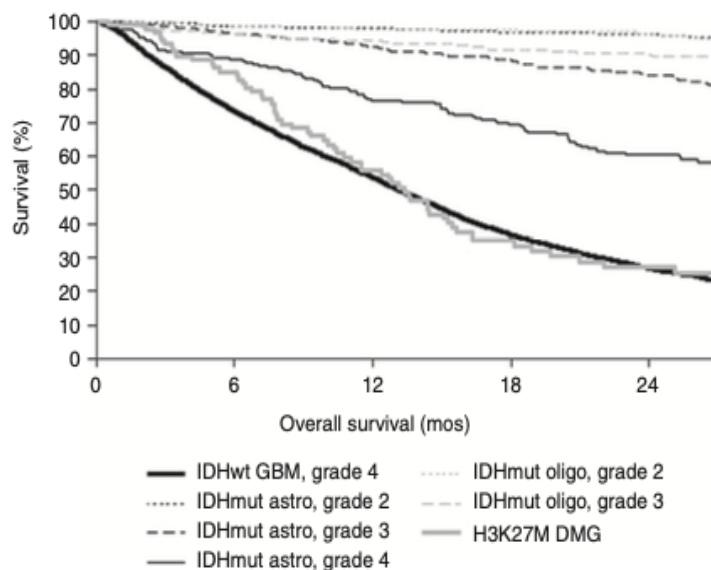


Figure 6: curva di Kaplan-Meier dell'overall survival per sottotipo tumorale.²⁷

ARTIFICIAL INTELLIGENCE (AI)

Intelligenza artificiale, machine learning e deep learning

L'intelligenza artificiale è una delle attuali aree di sviluppo più rapido nell'ambito informatico e gli applicativi in radiologia sono di estremo rilievo. Per AI si intende la capacità delle macchine di simulare funzioni cognitive umane come possono essere l'apprendimento, il *problem solving*, la percezione e il *decision making*.

Machine learning

Uno dei sottoinsiemi dell'AI è il *machine learning* (ML), termine da far coincidere col concetto di "apprendimento automatico". Infatti, il ML consente ai sistemi di migliorare automaticamente le proprie prestazioni attraverso l'esposizione progressiva ai dati, senza essere esplicitamente programmati per ogni singolo compito. Sulla base di quanto viene osservato e raccolto, gli algoritmi di ML altro non sono che un modello statistico che consente di fare predizioni sulla realtà. In quanto modello statistico, il ML è direttamente collegato al concetto di probabilità, in particolare al concetto di "probabilità d'errore": un apprendimento è infatti tale perché da un'azione che non è andata come previsto si può imparare qualcosa di nuovo. Seguendo questa logica è chiaro che maggiore è la quantità di dati disponibili, maggiore sarà la capacità dell'algoritmo di generalizzare. Le modalità di apprendimento sono tre: supervisionato (in genere utilizzato per compiti di classificazione), non supervisionato (utilizzato per generare clusterizzazioni di dati) e con rinforzo (nel quale è prevista l'esplorazione dell'ambiente di lavoro a fronte di una ricompensa quando l'azione svolta porta a un miglioramento).²⁸

Deep learning

Un ulteriore sottoinsieme del *machine learning* è il *deep learning*. Quest'ultimo costituisce la base della maggior parte degli strumenti di AI finalizzati all'interpretazione delle immagini. Dal punto di vista matematico, si fonda su reti neurali artificiali composte da molteplici strati (*layers*) di neuroni artificiali interconnessi fra loro. Durante la fase di addestramento, questi strati elaborano gli *input* ed estraggono progressivamente caratteristiche rilevanti dei dati (*features*),

generando output sempre più accurati.²⁹ Per *features* si intendono caratteristiche o proprietà rilevanti estratte dai dati che consentono all'algoritmo di riconoscere *pattern* e formulare previsioni: nella radiomica in particolare si tratta di variabili numeriche estratte da regioni o volumi di interesse (detti ROI o VOI) segmentate che si configurano su più strati di complessità partendo da quelli più semplici inerenti ad esempio la forma (come volume, superficie, sfericità, compattezza e irregolarità dei margini).³⁰

Tra le architetture di *deep learning* più rilevanti in ambito biomedico spiccano U-Net e V-Net, progettate specificamente per la segmentazione di immagini mediche tridimensionali. Esse sono caratterizzate da percorsi di contrazione ed espansione simmetrici che consentono una localizzazione precisa delle strutture anatomiche.³¹

ML vs DL: il caso della *black-box*

All'atto pratico, la differenza fondamentale è che mentre il ML richiede un intervento umano per identificare e selezionare le caratteristiche rilevanti a partire dai dati di partenza, il DL elabora le *features* in maniera autonoma, senza che vi sia necessità di selezione manuale delle variabili. Proprio per questo motivo, il ML è capace di funzionare anche su dataset ridotti a differenza del DL che richiede un addestramento fine e una grande quantità di dati non strutturati.³²

Il problema maggiore per un clinico che utilizza *tools* AI rimane la cosiddetta *black-box*, quella condizione per cui un *output* prodotto dalla macchina è frutto di un procedimento non interpretabile per l'uomo. Questo aspetto è problematico soprattutto nelle discipline ad elevata responsabilità decisionale (tra cui sanità e giustizia) ed è pervasivo quando si parla di modelli di ML molto avanzati o di DL. Tuttavia, la *black-box* risulta minimizzata o anche annullata (*white-box*) nel contesto del cosiddetto *shallow* ML: il machine learning più semplice e "rudimentale". Proprio per questo motivo alcuni autori sostengono la necessità di riprendere a generare modelli interpretabili piuttosto che progredire nell'alimentare modelli provvisti di *black-box*.^{32,33}

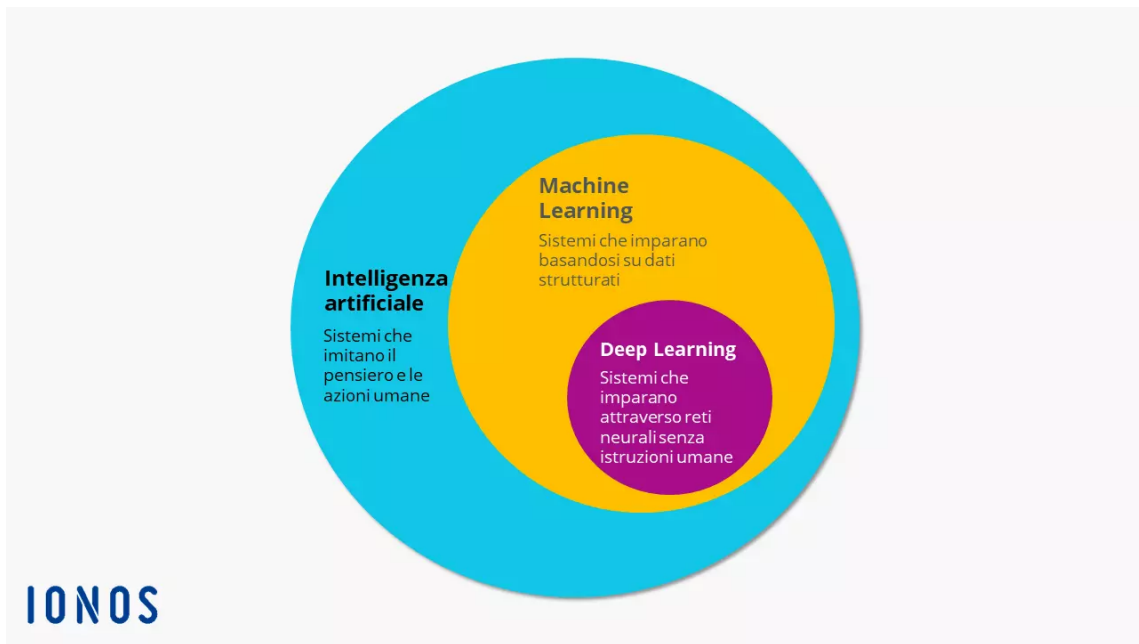


Figure 7: principali sottoinsiemi dell'AI.³⁴

Applicazioni in radiologia

Le applicazioni cliniche dell'AI in radiologia sono innumerevoli e sono destinate a crescere con l'avanzare delle possibilità tecnologiche.

Segmentazione automatica delle immagini

Uno degli applicativi più conosciuti è la segmentazione automatica delle immagini. Con segmentazione si intende un processo che consiste nell'identificare e raggruppare regioni di un'immagine che condividono caratteristiche simili (questo può essere fatto ad esempio sulla base della *texture* o del colore) assegnando a ciascuna zona un'etichetta di classe. Grazie ad algoritmi di *machine learning* e *deep learning*, è stato possibile migliorare la precisione e la velocità delle diagnosi. Le tecniche di segmentazione si suddividono in due grandi gruppi: supervisionate, che richiedono dati annotati, e non supervisionate, che invece non richiedono dati annotati. Alcuni metodi supervisionati – tra cui quelli *atlas based*, *support vector machine* (SVM) e *Random Forest* – sono stati molto studiati negli ultimi anni.³⁵

Ottimizzazione delle immagini

Altre possibilità riguardano l'ottimizzazione delle immagini: nel caso della TC, l'AI consentirebbe un miglioramento del posizionamento del paziente, l'evitamento

dell'*overscanning*, una riduzione del mezzo di contrasto e favorirebbe la ricostruzione delle immagini riducendo rumore e artefatti. Anche per la RM la riduzione degli artefatti gioca un ruolo chiave, in particolare la possibilità di ridurre quelli legati al movimento non intenzionale del paziente, così come nella possibilità di utilizzare meno gadolinio.³⁶

Radiomica

La radiomica rappresenta un'area emergente nel campo dell'analisi quantitativa consentita dall'AI. Il suo obiettivo è quello di relazionare *features* estratte dalle varie metodiche di *imaging* per ottenere informazioni significative in campo clinico o biologico. Questo consentirebbe un miglioramento per la diagnosi, la selezione dei trattamenti e la valutazione della risposta soprattutto in ambito oncologico ma non solo, nell'ottica di una medicina sempre più "*tailored*" sul singolo individuo.³⁷

Ad esempio, studi condotti in UNIMORE mostrano anche le potenzialità nello stimare lo stato mutazionale di BRCA1 nelle pazienti con carcinoma della mammella triplo negativi mediante DCE-MRI e radiomica.^{38,39}

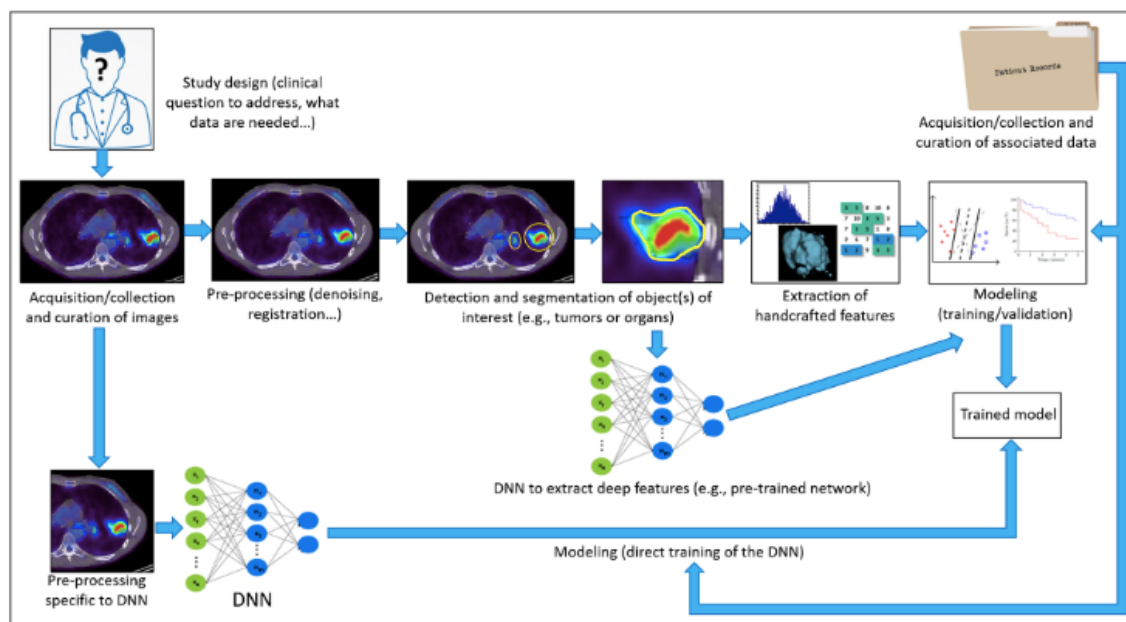


Figure 8: Tipico workflow di radiomica.³⁰

Facilitazione del flusso di lavoro

L'AI consente di migliorare i flussi di lavoro e l'efficienza dei lavori non-interpretativi connessi alla professione radiologica, ad esempio, automatizzando la

precompilazione di un referto. In questo ambito uno degli aspetti principali è la prioritizzazione degli esami basata sul grado di urgenza, che consentirebbe di riconoscere prima e quindi dare priorità a patologie tempo-dipendenti (come ictus ed emorragie). Questo consentirebbe anche di velocizzare gli aspetti più ripetitivi e tediosi della professione.⁴⁰

D'altro canto, il risparmio di tempo consentirebbe di concentrarsi sulle reali priorità del momento (come può essere la comunicazione al paziente di una diagnosi o la partecipazione a una valutazione multidisciplinare per la pianificazione di trattamento di un paziente complesso).

Applicazioni AI in neuro-oncologia

In campo neuro-oncologico, le decisioni terapeutiche richiedono estesa multidisciplinarietà. Professionisti come neuro e radio-oncologi, neuroradiologi, neurochirurghi e patologi collaborano quotidianamente proprio in virtù dell'estrema complessità gestionale di queste neoplasie. Le principali sfide riguardano la diagnosi integrata, la valutazione dell'estensione di malattia, il monitoraggio post-trattamento, il *grading* e la caratterizzazione molecolare. L'AI è in grado di supportare questi professionisti a vari livelli: può assistere il chirurgo nella valutazione dei margini di resezione e nella diagnosi intraoperatoria in tempo reale, può assistere il neuropatologo e il patologo molecolare nell'effettuare analisi integrate isto-molecolari sul campione, analizzare dati da sequenziamento genico e identificare biomarcatori per la predizione della risposta al trattamento e l'identificazione di varianti genetiche rilevanti.⁴¹

In ambito neuroradiologico, l'AI ha portato innovazioni nell'identificazione automatica dei margini tumorali, nella segmentazione automatica delle lesioni, nell'identificazione di sottotipi molecolari e nella distinzione tra tessuto tumorale e zone necrotiche.⁴¹

Nel merito della segmentazione automatica delle lesioni, si porta come esempio un modello che ha combinato due reti neurali distinte (una U-Net e una 3D-CNN) e che ha portato a ottenere risultati migliori rispetto agli approcci antecedenti su tre aree tumorali che normalmente vengono utilizzate nella valutazione clinica dei gliomi:

whole tumor, tumor core ed enhanced tumor. Questo studio conferma come la combinazione di reti neurali differenti, pratica ormai consolidata e facilmente implementabile, possa ottenere risultati diagnostici di elevatissima accuratezza nella segmentazione automatica in RM.⁴²

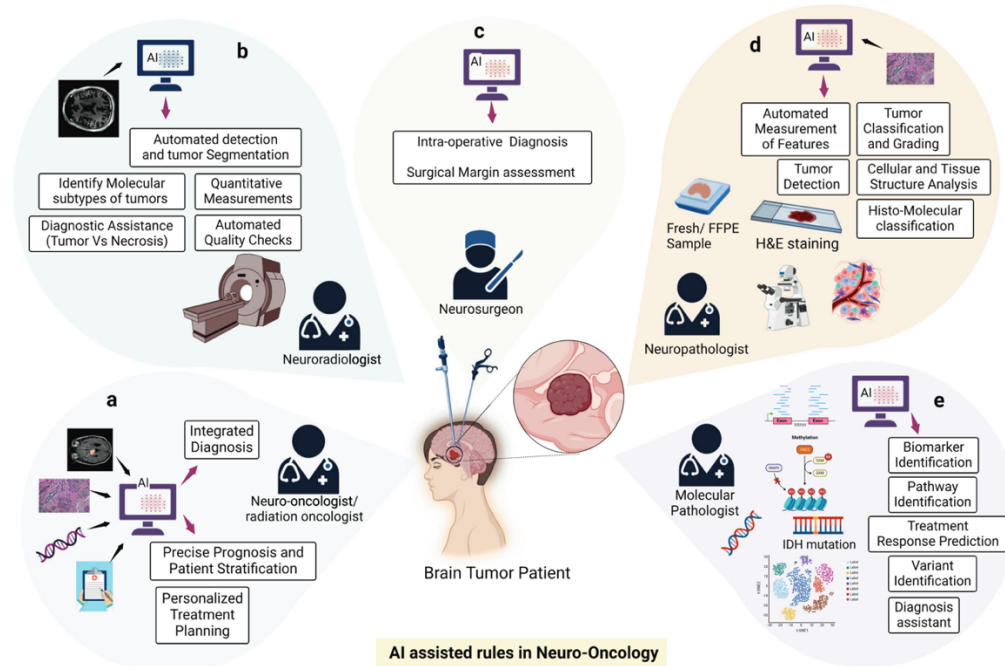


Figure 9: riassunto delle applicazioni dell'AI in neuro-oncologia.⁴¹

Automated Radiology Report Generation (ARRG)

L' ARRG (*Automated Radiology Report Generation*) è un'applicazione dell'intelligenza artificiale finalizzata a generare in maniera automatica referti clinici a partire da immagini radiologiche. Il razionale dietro al suo sviluppo risponde alla crescente pressione verso i servizi di diagnostica per immagini legati all'aumento dei volumi di lavoro e alla carenza di professionisti. Questo è reso possibile dalla combinazione di tecniche di elaborazione del linguaggio naturale e di visione artificiale. Nella loro architettura base, i modelli ARRG si sviluppano su reti *encoder-decoder*: il primo estrae le caratteristiche salienti dall'immagine, il secondo genera il referto. In passato erano utilizzati soprattutto approcci CNN-LSTM, le architetture più recenti si basano su grandi modelli linguistici (LLM) ma gli approcci matematici possibili sono numerosi e diversificati.⁴³ La valutazione di quanto il testo generato si avvicini a quello umano è affidata a metriche NLP (*natural language processing*)

come BLEU, ROUGE o Meteor.⁴⁴ Nell'ambito dell'ARRG la neuroradiologia rappresenta una frontiera naturale e ancora relativamente inesplorata.

Un esempio concreto lo si ritrova nel sistema ADS (*Acute stroke Detection and Segmentation*) sviluppato da Liu et al. nel 2023. ADS è addestrato su 1878 MRIs di pazienti con diagnosi clinica di ictus ischemico acuto: il sistema è risultato in grado di generare referti completamente automatizzati. Oltre a questo, è stato in grado di fornire simultaneamente anche il volume dell'infarto, la segmentazione del core ischemico e un vettore di caratteristiche quantitative delle regioni anatomiche coinvolte. Le prestazioni del sistema si sono mostrate comparabili a quelle di clinici esperti e superiori a quelle di un medico non esperto in neuroradiologia.⁴⁵

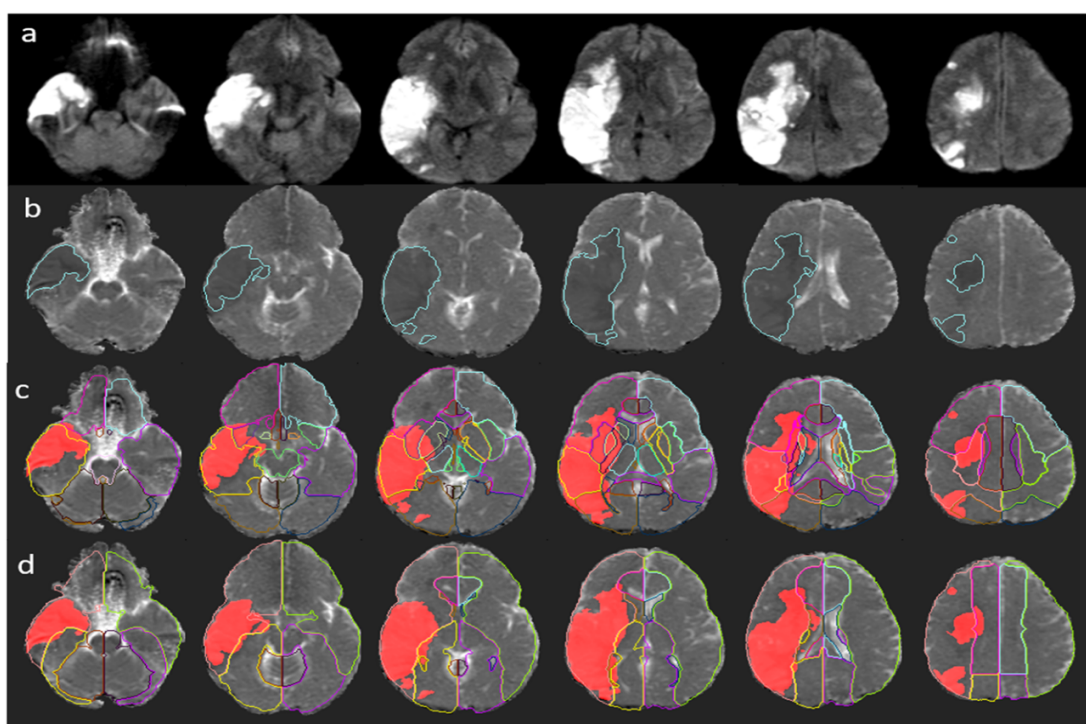


Figure 10: correlazione tra: a) ictus ischemico in DWI, b) segmentazione automatica del core ischemico, c-d) quantificazione dell'area interessata in rapporto a suddivisione anatomica classica atlas-based e territori di vascolarizzazione arteriosa rispettivamente.⁴⁵

Aspetti etici

L'ARRG e l'AI portano a numerosi vantaggi dal punto di vista lavorativo (come il risparmio del tempo che può essere reinvestito in altre attività), tuttavia, dall'altro lato della medaglia, forniscono una serie di sfide etiche importanti inerenti

soprattutto *privacy*, la responsabilità professionale e la paura della sostituzione dei professionisti.

Sostituzione medica

Infatti, nonostante i rapidi progressi dell'intelligenza artificiale in ambito diagnostico, la figura del radiologo non è destinata a essere sostituita, bensì potenziata. Secondo uno studio qualitativo condotto da Coker et al. (2026) su tredici esperti del settore AI in sanità, l'AI fungerà da strumento di supporto piuttosto che da sostituto del clinico, migliorandone le capacità senza eliminarne il ruolo. Gli esperti hanno sottolineato come i sistemi di AI siano in grado di raggiungere in alcuni contesti prestazioni comparabili a quelle degli specialisti; tuttavia, l'interpretazione clinica, la contestualizzazione del dato nel quadro del singolo paziente e, soprattutto, la comunicazione della diagnosi restano prerogative irrinunciabili del medico. Inoltre, il radiologo mantiene un ruolo insostituibile nei contesti in cui l'AI non è applicabile come emergenze tempo-dipendenti, guasti tecnologici o limitazioni infrastrutturali. Ne consegue che a fronte del progredire della tecnologia, rimane indispensabile il mantenimento di solide competenze cliniche e di autonomia di giudizio. L'AI è destinata a ridisegnare il profilo professionale del radiologo, spostandone il focus dalle attività routinarie verso un giudizio esperto, critico, personalizzato e umano, che nessun algoritmo potrà replicare.⁴⁶

Algoretica

La definizione di un codice etico (modernamente definito algoretica) che accompagni la progressività della cosiddetta "condizione tecno-umana" è fondamentale proprio per ovviare ai problemi etici che l'AI solleva. Tra i principi fondativi di questo nuovo paradigma ci sono fattori come la *cognizione sintetica (homo+machina) sapiens* (secondo la quale la visione della realtà mediata dallo strumento deve rimanere un processo attivo e non un acquisizione passiva di un risultato), il *free won't* (finalizzato a determinare un perimetro di ciò che la macchina può e non può fare e di valutarne il grado di autonomia decisionale) e il concetto di *volontà libera*, secondo la quale la macchina deve imparare almeno quattro direzioni d'interazione per una cooperazione adeguata con l'uomo: l'intuizione (le macchine dovrebbero essere in grado di intuire le volontà dell'uomo e

conseguentemente adattarsi ad esse), l'intellegibilità (bisogna fare in modo che l'agire della macchina sia intellegibile per l'uomo e che si possa capire prima cosa la macchina sta effettivamente per compiere), l'adattabilità (la macchina non deve adattarsi solo all'essere umano in quanto tale, ma anche alla personalità con cui interagisce) e la regolazione (davanti alle sue possibilità pressoché illimitate, la macchina deve saper regolare il suo agire in relazione alle caratteristiche precedenti e in relazione all'obiettivo che l'uomo le ha posto; non può essere solo l'assolutezza dell'obiettivo a guidare le sue azioni).⁴⁷

SCOPO DELLO STUDIO

Lo studio è stato svolto in collaborazione con il laboratorio AlmageLab di UNIMORE, un gruppo di oltre 73 ricercatori, di cui una divisione, AlmageLab-Zip, si dedica ai dati medici, in particolare a immagini (2D e 3D) e a referti testuali. È stato sviluppato un algoritmo di *machine learning atlas-based* finalizzato a valutare il grado di sovrapposizione tra un atlante segmentato nelle singole regioni e la maschera di segmentazione della lesione encefalica: prospettiva sfidante nell'ambiente oncologico viste le criticità legate all'anisotropia neoplastica e alla carenza di annotazioni.^{48,49} Il modello matematico in studio applica un algoritmo deterministico su ciascun volume: risulta diverso dall'utilizzo dei dati di *training* del *machine learning* classico e proprio per questo è complesso ottenerne una validazione. L'utilizzo di un algoritmo non derivante dal *deep learning*, scelta apparentemente non ottimale per le prestazioni, permette di sopperire alla carenza di dati annotati (*ground truth* di segmentazione)⁴⁹ e di ovviare al problema della "black box" insita nei modelli di *deep learning* e *machine learning* più avanzati.³² Gli algoritmi di *deep learning*, infatti, richiedono grandi moli di dati annotati (spesso non presenti) e la segmentazione dell'intero encefalo nelle singole regioni di corteccia, materia bianca sottocorticale e materia bianca profonda è una procedura estremamente lunga e costosa, che può richiedere giorni di lavoro.

L'obiettivo del seguente studio è contribuire alla validazione e al miglioramento dell'algoritmo, processo per cui è fondamentale che il suo stesso lavoro venga effettuato e controllato anche dall'uomo così da ottenere dati confrontabili e verificando l'*agreement* tra la macchina e l'operatore.

Inoltre, al fine della valutazione, si è deciso di tentare un confronto tra i referti testuali di produzione medica e gli output testuali dell'algoritmo così da valutare anche l'implementazione clinica dello stesso nell'ambito della refertazione automatica.

Da un punto di vista clinico, infatti, l'implementazione di un modello di segmentazione automatica garantirebbe risultati costanti, non influenzati dalla soggettività e dalla variabilità inter-osservatore e intra-osservatore.⁵⁰ Allo stesso tempo, l'automazione del processo di segmentazione e di refertazione

consentirebbe una riduzione dei tempi di analisi con possibili ricadute positive sul flusso di lavoro clinico.⁵¹

MATERIALI E METODI

Disegno dello studio

Campione

Questo studio retrospettivo è stato svolto utilizzando un dataset di 96 pazienti con prima diagnosi di glioma diffuso dell'adulto effettuata mediante RMN afferenti al servizio di Neuroradiologia dell'azienda ospedaliero-universitaria di Modena. Il dataset è stato raccolto in un periodo intercorrente tra settembre 2020 e giugno 2025. Per questioni di tempistiche tecniche, il lavoro ivi presentato riguarda la prima metà del campione, ossia 48 pazienti dei 96 totali. Gli esami svolti dai pazienti consistono in RMN con protocollo pre-operatorio con e senza mezzo di contrasto endovenoso effettuati con finalità di approfondimento o diagnosi rispetto all'insorgenza di sintomi neurologici evocativi di neoplasia cerebrale, verso la quale rimangono comunque sintomi caratterizzati da bassissimi valori predittivi positivi in cure primarie.⁵²

Criteri di inclusione ed esclusione

Sono stati inseriti nel dataset i pazienti caratterizzati da:

- Età ≥ 18 anni;
- Assenza di controindicazioni e consenso all'esecuzione dell'esame di RMN;
- Esame di RMN svolto con apparecchiatura di competenza del servizio di Neuroradiologia dell'AOU;
- Diagnosi di primitività tumorale ad atteggiamento espansivo sulla base dei dati clinici, d'imaging e patologici nel contesto del PDTA dell'AOU di Modena;
- Consenso informato raccolto quando possibile, come da comitato etico.

Di conseguenza, sono stati esclusi tutti i pazienti che non soddisfacevano queste caratteristiche.

Protocollo d'imaging

Le immagini sono state ottenute da apparecchiature per l'esecuzione di RMN a 3T e 1,5T localizzate su più ospedali della provincia di Modena e prodotte da fornitori tecnologici diversi. Il protocollo che è stato utilizzato consente un'adeguata valutazione morfologica delle lesioni e prevede l'acquisizione di sequenze 3D ad alta risoluzione atte a produrre:

- Immagini 3D-T1w senza mezzo di contrasto;
- Immagini 3D-T1w con somministrazione di mezzo di contrasto;
- Immagini 3D-T2-FLAIR: metodica di *inversion recovery* finalizzata a sopprimere il liquor cefalo-rachidiano per aumentare la visibilità dell'edema vasogenico o la progressione tumorale *non-enhancing*⁵³;
- Immagini T2w assiali.

L'utilizzo di apparecchiature diverse in luoghi diversi consente di riflettere le reali differenze tecniche e pratiche che si incontrano durante la pratica clinica quotidiana, contribuendo a valorizzare lo studio.

Selezione dei referti

Sono stati ricavati i referti testuali di produzione medica specialistica dei 48 pazienti analizzati. I reali referti complessivamente utilizzati sono 45 per i seguenti motivi:

- Per un totale di 3 pazienti di 48 non è stato possibile ricavare il referto dall'archivio;
- Per 1 paziente di 48 si è deciso di utilizzare i referti di due esami RMN ad esso attribuiti;
- Sono stati esclusi i referti di metodiche che non fossero la RMN.

I referti sono redatti dai diversi medici del servizio di neuroradiologia che utilizzano, pertanto, metodi di refertazione e stili differenti. Inoltre, prima di qualsiasi elaborazione i referti sono stati completamente anonimizzati, ricavando esclusivamente la parte testuale descrittiva della lesione in esame.

Pipeline dell'algoritmo e dell'autogenerazione dei referti

Ai fini dello studio è stato utilizzato un algoritmo di *machine learning* ideato con il gruppo AlmageLab-zip di UNIMORE. La peculiarità di questo algoritmo *atlas-based* è quella di partire da dati volumetrici già precedentemente segmentati (ossia maschere tridimensionali della lesione neoplastica) per trasformarli in informazioni anatomiche e quantitative. Lo stesso algoritmo, a partire dalle informazioni stesse, produce un *json strutturato* (ossia una modalità per organizzare e scambiare informazioni tra programmi e applicazioni) e un testo descrittivo (*txt*). Seguono nel dettaglio le singole fasi della *pipeline* principale.

Segmentazione delle aree tumorali

La segmentazione delle aree cerebrali è un processo particolarmente laborioso e dispendioso in termini di tempo. In un lavoro precedente condotto presso UNIMORE è stato sviluppato un modello di DL per la segmentazione automatica dei gliomi, addestrato sul dataset pubblico BRATS⁵⁴ e successivamente adattato al contesto clinico della provincia di Modena mediante un processo di *fine-tuning*.⁵⁵ Tale modello ha permesso di ottenere segmentazioni automatiche accurate, successivamente corrette singolarmente da neuroradiologi esperti, che sono state utilizzate come punto di partenza per il presente studio.

Per questo lavoro è stato invece sviluppato appositamente un algoritmo di ML il cui obiettivo è valutare il grado di sovrapposizione tra un atlante suddiviso manualmente nelle singole aree cerebrali e le segmentazioni prodotte dal modello di DL precedentemente descritto.

L'algoritmo accede alla segmentazione del tumore e utilizza le diverse componenti fondamentali distinguendo: *core* necrotico, edema associato, *enhancement*, lesioni separate (definibili come multicentriche) e lesioni multifocali.

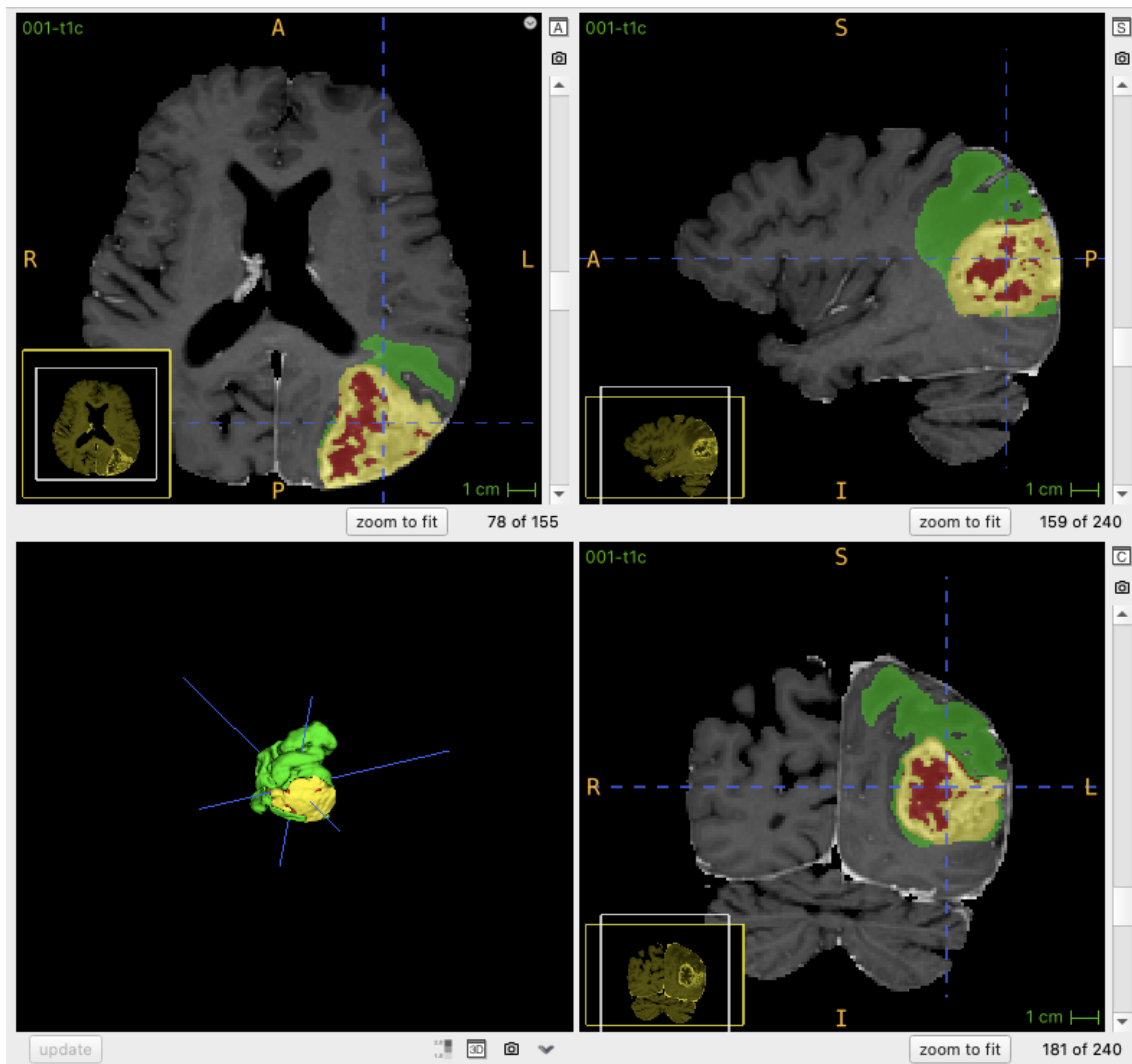


Figure 11: Esempio proveniente dal software ITK-SNAP di maschera di segmentazione registrata su immagine T1w con contrasto: i colori rappresentano le aree di core necrotico (rosso), enhancement (bianco) ed edema (verde).

Secondo la convenzione utilizzata, la differenza tra lesione multicentrica e multifocale consta essenzialmente nella condivisione dell'edema: se due o più lesioni condividono la stessa componente edematosa saranno multifocali, se invece due lesioni presentano edemi in discontinuità, quindi distinti, saranno definite multicentriche.⁵⁶ Nella pratica, davanti a lesioni separate l'algoritmo tenta di trattarle come distinte.

Registrazione dell'atlante anatomico sul paziente

L'atlante anatomico cerebrale SRI24⁵⁷, precedentemente annotato da un medico neuroradiologo, viene deformato nello spazio anatomico del singolo paziente tramite registrazione in T1: a questo fine è stata utilizzata la libreria *open source*

ANTs (*advanced normalization tools*), conosciuta per le registrazioni ad alta qualità.⁵⁸ Questa parte consente di associare ogni voxel della lesione a una specifica area anatomica (come può essere il giro prefrontale o la sostanza bianca profonda del lobo temporale).



Figure 12: atlante anatomico annotato a partire dalle segmentazioni dell'atlante SRI24 rifinite con un'annotazione manuale aggiuntiva per la sostanza bianca.

Identificazione della sede di origine

Per ogni focus interpretato come tumorale il sistema ne calcola il baricentro e associa quest'ultimo all'annotazione anatomica più rappresentata in quel punto sull'atlante. Da questa associazione l'algoritmo ricava: nome anatomico, macroarea, lato e profondità.

Identificazione delle aree coinvolte

L'algoritmo controlla quali aree vengono intersecate dalla lesione e per ogni regione anatomica valuta quantitativamente il numero di voxel coinvolti. Al superamento di

un determinato valore soglia (0,5%), l'area viene inclusa tra le aree interessate. Ogni procedura viene svolta due volte: una per il *tumor core* e l'altra per l'edema.

Valutazione delle aree eloquenti

Oltre all'atlante anatomico generale, sono state utilizzate anche maschere di aree eloquenti, ossia aree di particolare interesse neurologico poiché associate a funzioni specifiche.⁵⁹ Ci si è basati principalmente su quattro aree funzionali maggiori: area motoria primaria, area del linguaggio motorio, area del linguaggio recettivo e area visiva primaria.

Come detto precedentemente, se ciascuna area è interessata almeno per lo 0.5%, allora viene segnalata come coinvolta.

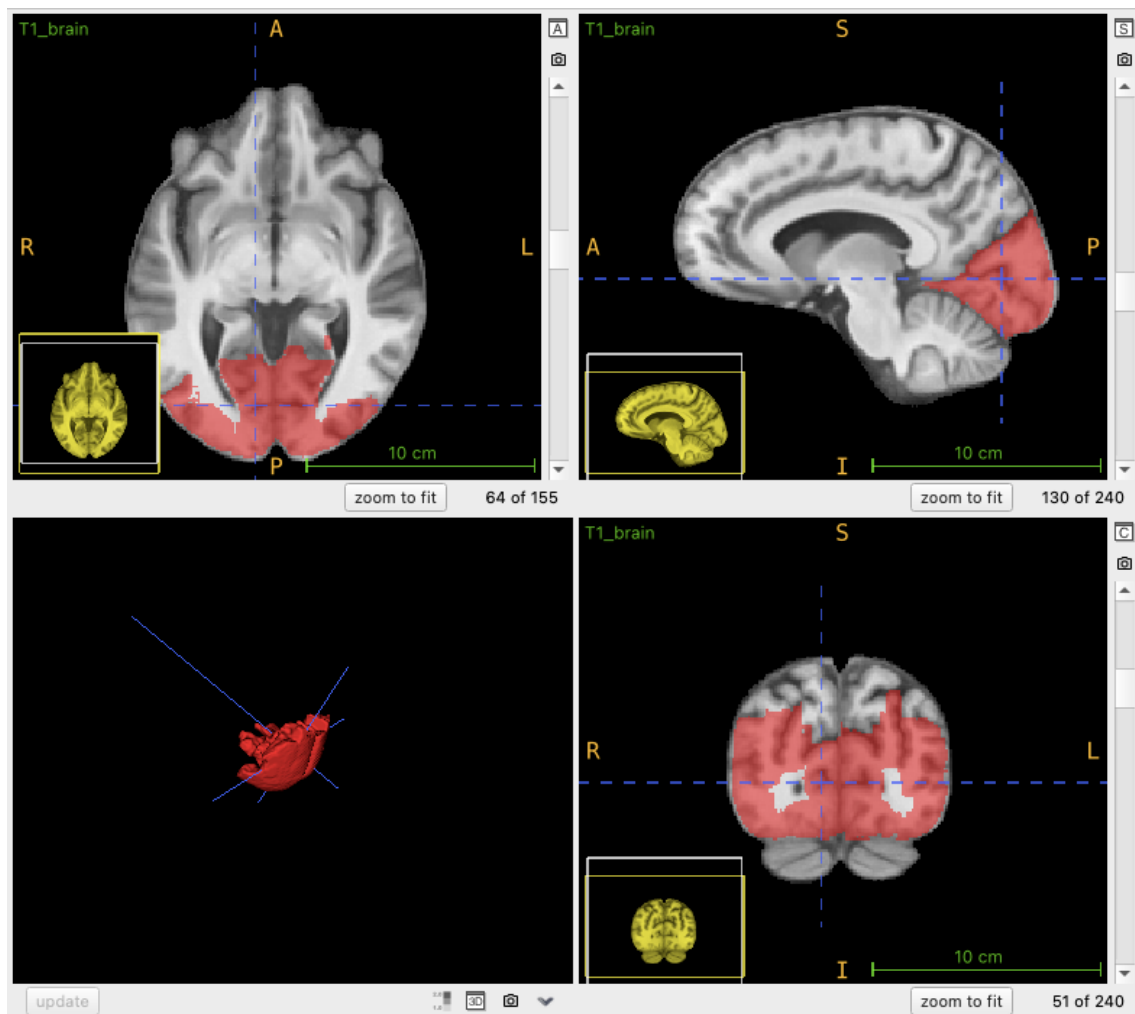


Figure 13: esempio di annotazione per l'area eloquente della visione (colorata in rosso) registrata su immagine T1w.

Valutazione di metriche geometriche aggiuntive

Per ogni lesione sono stati estratti una serie di parametri quantitativi: volume del *core* necrotico, volume della componente di *enhancement*, volume dell'edema, proporzione relativa tra necrosi - *enhancement* - edema, dimensione del *bounding box* della lesione, massimo diametro assiale, punto medio del massimo diametro assiale e spessore medio della componente di *enhancement*.

Il fine di queste valutazioni è di ottenere un dettaglio morfologico elevato anche in vista di un successivo iter di confronto.

```
(...)  
],  
  "size": [  
    23.0,  
    31.0,  
    34.0  
  ],  
  "ed_size": [  
    40.0,  
    68.0,  
    52.0  
  ],  
  "nec_volume": 7827.0,  
  "ed_volume": 22811.0,  
  "et_volume": 0.0,  
  "nec_proportion": 0.2560855908912446,  
  "ed_proportion": 0.746335558173014,  
  "et_proportion": 0.0,  
  "thickness_et": null,  
  "depth": [  
    "cortex",  
    "deep_white_matter",  
    "subcortical_white_matter"  
  ],  
  "largest_axial_diameter_and_midpoint": [  
    30.4138126514911,  
    [  
      132,  
      144,  
      117  
    ]  
  ]  
(...)
```

Figure 14: generated json contenente le valutazioni geometriche aggiuntive di una delle lesioni in esame.

Attribuzione del tipo di lesione

Il sistema produce tre tipi principali di output di lesione:

- ‘*single*’: lesione con un singolo centro tumorale;
- ‘*multifocal*’: lesione con più centri tumorali e edema condiviso, l’algoritmo provvederà a separare i singoli centri;
- ‘*isolated_edema*’: area infiltrativa-edematosa senza una componente di core tumorale evidente.

Fase generativa di json e testo

Al termine, tutte le informazioni vengono salvate in un file Python strutturato *generated.json*. Proprio da questo file verrà generato un testo con frasi predefinite.

The brain shows **1 lesion**. The lesion originates from the right middle frontal gyrus, at coordinates: (91,120,116), on the frontal lobe, on the right hemisphere, in the cortex.

The tumor core involved the following areas: right precentral gyrus, right superior frontal gyrus, right middle frontal gyrus, right precentral white matter, right middle frontal white matter.

The tumor core involves the motor eloquent area. No deep white matter invasion is observed. The tumor extends across the cortex and subcortical white matter.

The largest axial diameter measures 26 mm and its midpoint is found at coordinates (90,120,115).

The tumor core has size 24 mm x 28 mm x 22 mm.

The necrotic core represents 25.64% of the whole tumor's mass, with volume 2888 mm³ while the enhancing tumor represents 0.66% with volume 74 mm³.

The average thickness of the enhancing margin is 2.03 mm.

The edema surrounding the lesion involves the following areas: right precentral gyrus, right superior frontal gyrus, right middle frontal gyrus, right precentral white matter, right superior frontal white matter, right middle frontal white matter, right supplementary motor area white matter, right frontal deep white matter.

The edema has size 35 mm x 32 mm x 38 mm.

The edema represents 74.67% of the whole tumor's mass, with volume 8409 mm³

Figure 15: esempio di referto generato dall'algoritmo per una delle lesioni in esame.

Annotazioni manuali all'algoritmo

Attraverso l'utilizzo di un programma che consente la scrittura e il test di codici in linguaggio Python, si sono confrontati gli script *generated.json* e i referti *.txt* automaticamente generati dall'algoritmo con le immagini reali provenienti dalle

immagini di RMN anonimizzate e pre-processate. Ai fini del confronto, sono stati utilizzati i software napari⁶⁰, in particolare il suo plug-in specifico per le immagini mediche, e ITK-SNAP, un software *open-source* per la segmentazione e la visione di immagini biomedicali.⁶¹ Ogni script *json* è stato valutato nella sua interezza, in particolare ci si è concentrati su otto campi prioritari:

1. *'type'*: che identifica la lesione come singola, multicentrica, multifocale o edema isolato;
2. *'origin_name'*: che identifica l'area cerebrale dove ritroviamo l'origine geometrica per la lesione neoplastica all'interno di 281 aree annotate;
3. *'origin_macroarea'*: che identifica la regione cerebrale in cui viene a trovarsi l'origine geometrica della lesione neoplastica all'interno di 11 macroaree annotate;
4. *'origin_side'*: che identifica il lato cerebrale in cui viene a trovarsi l'origine geometrica della lesione neoplastica in 3 possibili conformazioni (*left*, *right* e *median*);
5. *'origin_depth'*: che identifica la profondità dell'origine geometrica della lesione neoplastica all'interno di 4 profondità codificate;
6. *'affected_perc'*: che identifica la percentuale di interessamento per ogni singola area cerebrale da parte della lesione tumorale considerando *core* tumorale + *enhancement* contrastografico;
7. *'affected_perc_ed'*: che identifica la percentuale di interessamento per ogni singola area cerebrale da parte della lesione tumorale considerando però anche l'edema associato alla lesione stessa;
8. *'depth'*: che identifica il coinvolgimento in profondità progressivo da parte del tumore.

Ognuno di questi campi è stato annotato in caso si fossero rivelate discordanze con quanto presentato dall'immagine di RMN (alcuni esempi: alcune aree cerebrali categorizzate come interessate sono state rimosse, altre sono state aggiunte, alcuni tumori interpretati come singoli sono stati resi multifocali o viceversa). Ogni area

cerebrale possiede un distintivo codice numerico assegnato univocamente in modo che ci sia corrispondenza tra le aree grigie e le relative aree di bianca sottocorticale.

```
"type": "single",
"centers": [
  {
    "origin": 1001,
    "origin_name": "left precentral white matter",
    "origin_macroarea": "frontal",
    "origin_side": "L",
    "origin_depth": "subcortical_white_matter",
    "barycenter": [
      132.53213236233552,
      144.30893062476045,
      119.02376389421234
    ],
    "affected": [13 elements],
    "affected_names": [13 elements],
    "affected_perc": [
      [
        1019,|
        "left supplementary motor area white matter",
        "18.1551%",
        0.1815510412674077
      ],
      (...)
    ]
  }
]
```

Figure 16: esempio json delle principali aree interessate dal processo di annotazione

Pre-elaborazione dei referti

Affinché le informazioni contenute nei referti fossero utilizzabili dall'algoritmo e fosse possibile il confronto con i *generated.json* annotati precedentemente, si è svolta un'estrazione manuale di *features* rilevanti dai testi medici. Per ognuno di essi, si sono ricavati sei campi:

1. *'num_centri'*: che corrisponde al numero di foci lesionali;
2. *'diametro'*: che corrisponde alla misura in mm delle voci di referto identificate come "diametro massimo" o in un'unica dimensione dello spazio;
3. *'dimensioni'*: che corrispondono alle dimensioni della lesione in millimetri lungo due o tre direzioni dello spazio, se esplicitate;
4. *'aree_coinvolte'*: che corrispondono alle singole aree cerebrali interessate dalla lesione. La nomenclatura adottata si basa su un *pool* di 281 aree;

5. 'macroarea': che corrisponde alla regione cerebrale interessata dal processo patologico. Anche in questo caso la nomenclatura si basa su un *pool* di 11 macroaree;
6. 'lato': che corrisponde al lato interessato dal processo patologico che poteva essere *left*, *right* o *median*.

In aggiunta, nel caso in cui un dato non fosse direttamente deducibile dal testo del referto, o in caso di assenza dello stesso, il campo è stato contrassegnato con una linea di codice *null*.

Ciascun campo è stato compilato manualmente rispettando una determinata sintassi, tale che consentisse il successivo confronto da parte dell'algoritmo.

```
[
  {
    'num_centri': 1,
    'diametro': null,
    'dimensioni': [50,37],
    'aree_coinvolte': ['left temporal deep white matter','left occipital deep white matter'],
    'macroarea': ['parietal', 'occipital'],
    'lato': 'L'
  }
]
```

Figure 17: esempio di json compilato manualmente con le informazioni salienti ottenute dai referti testuali.

Metriche di esecuzione del confronto

Per ultimo, si sono confrontati i dati ottenuti effettuando un confronto tra: una versione strutturata corretta, una versione strutturata generata interamente da algoritmo e le informazioni estratte dai referti testuali.

Ai fini del confronto, sono state utilizzate cinque metriche:⁶²

- *Accuracy*: intesa come la proporzione di tutte le classificazioni corrette, positive o negative che siano. Viene calcolata con la formula:

$$\frac{\text{veri positivi} + \text{veri negativi}}{\text{veri positivi} + \text{veri negativi} + \text{falsi positivi} + \text{falsi negativi}}$$

Un modello ideale non dovrebbe fornire falsi negativi e falsi positivi, pertanto, l'*accuracy* è tanto migliore tanto più si avvicina a 1 (100%).

- *Precision*: intesa come la proporzione di tutte le classificazioni positive del modello che sono effettivamente positive:

$$\frac{\text{veri positivi}}{\text{veri positivi} + \text{falsi positivi}}$$

Una precisione pari a 1 (100%) indica l'assenza di falsi positivi.

- *Recall*: intesa come la proporzione di positivi reali che sono stati correttamente classificati come positivi. Viene definito matematicamente dalla formula:

$$\frac{\text{veri positivi}}{\text{veri positivi} + \text{falsi negativi}}$$

Per lo stesso concetto prima riportato, un modello ideale non dovrebbe possedere falsi negativi, pertanto, il recall è tanto migliore tanto più si avvicina a 1 (100%).

- *F1 score*: una misura sintetica che ha il compito di bilanciare *precision* e *recall* facendone una media:

$$2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Si tratta di una metrica che penalizza i modelli che hanno molta discrepanza tra le due metriche. Ad esempio, un modello con elevata *precision* ma basso *recall* avrà un *f1* basso.

Calcolare insieme le metriche *precision* e *recall* permette anche di valutare il funzionamento su più ampia scala: un modello che classifica il 100% delle volte come “malato” avrebbe un recall elevato, ma una *precision* bassa.

- *Prevalence*: una metrica modificata che rielabora il concetto classico di prevalenza. Viene definita matematicamente dalla formula:

$$\frac{\text{Numero dei campi mediamente presenti}}{\text{Numero di campi possibili}}$$

RISULTATI

Analisi descrittiva della popolazione

Lo studio ha incluso e analizzato 48 dei 98 protocolli di RMN preoperatoria derivanti dai pazienti selezionati dai criteri di inclusione: la statistica descrittiva si basa pertanto sui dati ottenuti da 48 pazienti.

La coorte in esame ha una mediana d'età di 62,75 anni con un range interquartile (IQR) di 52,5-69,7 anni. Il soggetto d'età massima ha 83 anni, quello d'età minima 31. I soggetti di sesso maschile sono 27 (56,3%) mentre quelli di sesso femminile sono 21 (43,7%).

Dal punto di vista patologico, 35 pazienti hanno diagnosi istologica di glioblastoma (72,9%), 4 pazienti hanno diagnosi di astrocitoma (8,3%), un solo paziente ha diagnosi di oligodendroglioma (2,1%). I restanti 8 (16,7%) non possiedono una diagnosi certa dal punto di vista istologico ma solo un'indicazione del grado della lesione (alto o basso).

Un totale di 35 pazienti (72,9%) ha svolto l'esame di RMN con un campo magnetico di 3T, mentre 13 pazienti (27,1%) hanno svolto la RMN a 1,5T.

	Totale (N=48)	GBM (n=35)	Astrocitoma (n=4)	Oligodendroglioma (n=1)	Glioma non specificato (n=8)	1,5T (n=13)	3T (n=35)
Mediana anni d'età [IQR]	62,75 [52,5 - 69,7]	62,8 [53,3 - 70,0]	68,4 [62,0 - 75,9]	36,9	56,8 [47,8 - 67,6]	-	-
Sesso							
Maschi	27 (56,3%)	22 (45,8%)	1 (2,1%)	0	4 (8,3%)	6 (46,2%)	21 (60%)
Femmine	21 (43,7%)	13 (27,1%)	3 (6,3%)	1 (2,1%)	4 (8,3%)	7 (53,8%)	14 (40%)
Campo magnetico							
1,5T	13 (27,1%)	9 (18,8%)	1 (2,1%)	1 (2,1%)	2 (4,2%)	-	-
3T	35 (72,9%)	26 (54,2%)	3 (6,3%)	0	6 (12,5%)	-	-

Table 1

Utilizzando scanner diversi e con campi magnetici differenti il campione riflette una variabilità riscontrabile anche nella realtà clinica quotidiana.

Inoltre, questi dati confermano la tendenza epidemiologica di glioblastoma ad essere la neoplasia più frequente tra i gliomi diffusi dell'adulto, in particolare nel sesso maschile.

Analisi del contenuto dei referti

Dall'analisi dei dati presenti emerge che la totalità dei referti (100%) contiene le informazioni inerenti al lato interessato dalla neoplasia e alla macroarea interessata (dato che scende all'80% se consideriamo le singole aree). Quasi tutti i referti (95,6%) riportano informazioni sul numero delle lesioni e nel 73,3% dei casi è descritta almeno una dimensione spaziale; dato che scende al 51,1% se consideriamo i referti che riportano almeno due dimensioni spaziali.

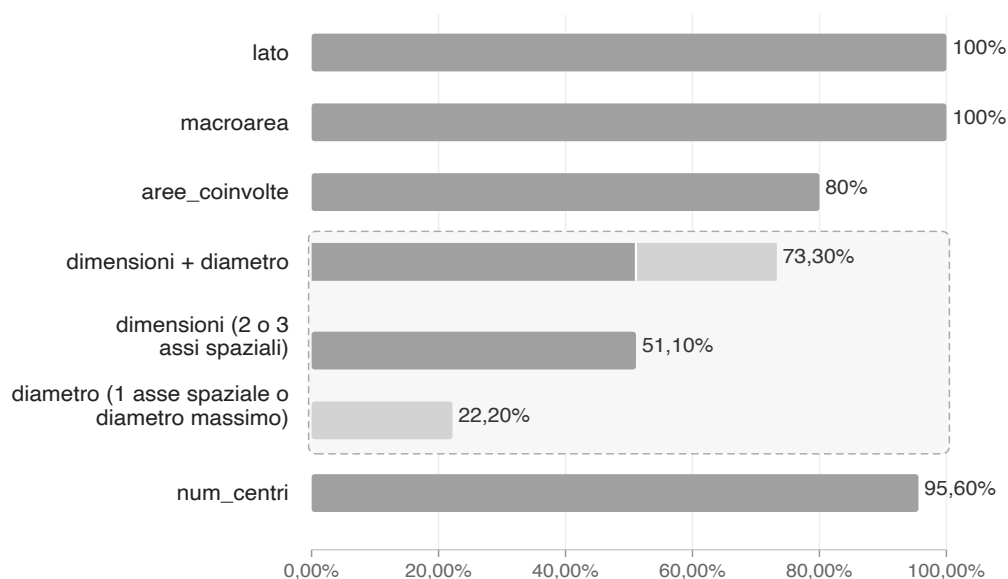


Figure 18

Output annotazioni manuali dell'algoritmo contro algoritmo *raw*

In prima istanza, alla comparazione tra le annotazioni manuali e l'algoritmo *raw* sono emersi i seguenti risultati:

	number of lesions	origin area localization	origin side
accuracy [IC95%]	70,49%	65,18% [50.2–77.9%]	97,92% [88.9–99.9%]
p	-	<0.001 (***)	<0.001 (***)

Table 2: Legenda: $p < 0.001$ (***) , $p < 0.01$ (**), $p < 0.05$ (*), non significativo se $p \geq 0.05$.

I numeri mostrano che l'algoritmo (tenendo conto delle precedenti segmentazioni ottenute con DL) possiede un'*accuracy* elevata nel riconoscere correttamente il numero delle lesioni (70,49%). L'*accuracy* è nettamente più elevata se si tratta di riconoscere il lato d'origine della neoplasia (97,92%) ed è più bassa se si tratta di

localizzare l'area d'origine (65,18%) su un totale di 281 aree annotate, pur mantenendo complessivamente un valore elevato di sovrapposizione rispetto alle annotazioni.

Il calcolo degli intervalli di confidenza è stato eseguito utilizzando il metodo di Wilson per le proporzioni ed è stato calcolato un valore p per testare che il risultato della metrica in relazione al lato e area d'origine fosse significativamente diverso dal valore attribuibile al caso: considerando 3 possibili localizzazioni per il side (*right*, *left* e *median*) e 281 possibili localizzazioni per l'area d'origine, i risultati emergono come fortemente significativi ($p < 0.001$). Per il numero delle lesioni, l'effetto del caso è difficilmente quantificabile in quanto variabile numerica discreta, pertanto il valore p non è stato calcolato.

	<i>precision</i>	<i>recall</i>	<i>f1</i>	<i>prevalence</i>
tumor core affected	62,75%	84,91%	72,16%	4,42%
tumor core eloquent	91,89%	73,91%	81,93%	0,20%
edema affected	63,40%	95,10%	76,08%	12,71%

Table 3

Le altre metriche mostrano un elevato valore di *precision* per quanto riguarda la classificazione di aree eloquenti interessate (91,10%), indicativo di un basso numero di classificazioni falsamente positive, ossia di aree che vengono ritenute interessate dal processo espansivo anche se non lo sono. Un valore elevato di *recall*, che indica invece numeri più bassi di classificazioni falsamente negative, ossia di aree che non vengono ritenute interessate dal processo espansivo anche se lo sono, lo troviamo negli scores delle aree di edema. Complessivamente, si può notare come per tutti gli scores ci siano valori elevati (>70%) di *recall* e *f1*.

Output annotazioni manuali dell'algoritmo contro referti reali

Numero dei centri, lato e macroarea

Le analisi successive sono state svolte confrontando l'output dell'algoritmo dopo le annotazioni con le informazioni contenute nei referti clinici.

	<i>Numero dei centri</i>	<i>Lato interessato</i>	<i>Macroarea interessata</i>
Exact match [IC95%]	76,70%	93% [81.3% – 97.9%]	100% [92.1% – 100.0%]
Off by 1 (exact match ± 1)	14% (90,7%)		

Table 4

Ciò che emerge è che il *matching* dell'algoritmo annotato nei confronti delle informazioni testuali ottiene massimi risultati per la *detection* della macroarea (100%), calando leggermente se si parla del lato d'origine (93%). I risultati peggiori li ritroviamo nella quantificazione del numero di centri (76,7%) con un valore di scostamento di ± 1 centro (*off by 1*) del 14%: questo implica che in questa percentuale aggiuntiva di casi la discordanza tra algoritmo e referto è al massimo di una lesione in più o in meno. Gli intervalli di confidenza al 95% sono stati calcolati con il metodo di Wilson per le proporzioni.

Diametro e dimensioni

	diametro	dimensioni
Mean relative error	51,60%	19,90%

Table 5

In ultimo, se si considerano le dimensioni dell'algoritmo confrontate con le dimensioni ricavate dal referto (derivanti dalla misurazione dei diametri lungo due o tre assi spaziali), l'errore relativo si attesta circa al 20% (19,90%): questo significa che le misurazioni si discostano del 20% tra referto e output algoritmico; ad esempio, una misurazione di 10 mm dell'algoritmo corrisponderà a una misurazione di 8 o 12 mm sul referto.

Successivamente, si nota che il diametro assiale massimo dell'algoritmo confrontato con il diametro massimo del referto (nei referti con l'indicazione di un singolo asse spaziale, quell'asse è stato interpretato come diametro massimo)

mostra un errore relativo del 51,60%: lo scostamento è per più della metà della misurazione.

DISCUSSIONE

I risultati ottenuti nel presente studio offrono una prima valutazione sistematica delle performance dell'algoritmo deterministico di *machine learning* ideato, confrontandone gli *output* sia con le annotazioni dell'operatore sia con i referti testuali prodotti da neuroradiologi esperti.

L'analisi complessiva dei dati suggerisce che il sistema raggiunge livelli di accuratezza clinicamente rilevanti in diversi domini, pur evidenziando aree di miglioramento che meritano approfondimento. I dati vanno interpretati tenendo conto che si valutano risultati preliminari, in quanto riferiti a metà del campione.

Performance nella classificazione del numero di lesioni

Il primo dato degno di nota riguarda l'*accuracy* nella classificazione del numero di lesioni, che si attesta al 70,49% nel confronto algoritmo-annotazioni manuali.

Al contrario l'*accuracy* si attesta al 76,7% nel confronto con i referti clinici: questo risultato è incoraggiante e suggerisce che il sistema possiede una buona capacità di rilevare il numero effettivo delle lesioni. Il valore di *Off by 1* del 14% porta il dato al 90,7%: questo indica che in questa percentuale di casi il riconoscimento si discosta di massimo una lesione rispetto al referto.

La distinzione numerica è un aspetto molto rilevante: le lesioni multifocali e multicentriche (definibili multiple) si associano generalmente a una prognosi peggiore con un aumento del rischio di morte di circa il 70%.⁶³

Il fatto che l'*accuracy* dell'algoritmo annotato migliori nel confronto con i referti testuali rispetto a quello con l'algoritmo *raw* è un dato interpretabile alla luce della natura stessa dei referti: in questo campione i neuroradiologi tendono a descrivere prevalentemente la lesione dominante, semplificando la complessità morfologica in un'unica entità clinicamente rilevante, il che si avvicina maggiormente all'output testuale dell'algoritmo annotato rispetto alla granularità analitica dell'algoritmo *raw*.

Questi fenomeni suggeriscono che la metrica dell'*accuracy*, in questo contesto, misura non solo la correttezza tecnica dell'algoritmo, ma anche il grado di

sovrapposizione con la prassi clinica reale, che è in ultima analisi l'obiettivo applicativo principale del sistema.

Localizzazione anatomica e lateralità

Per quanto riguarda la localizzazione dell'area d'origine misurata dall'algoritmo come il centro geometrico del *tumor core*, l'*accuracy* si attesta al 65,18% nel confronto con le annotazioni manuali. Un vero e proprio confronto con l'area di origine del referto non è possibile vista l'eterogeneità descrittiva che li caratterizza per le singole aree cerebrali: ciò che è possibile valutare è il grado di sovrapposizione a livello di macroarea. Si è ritenuta la classificazione corretta se la macroarea d'origine indicata dall'algoritmo era contenuta tra le macroaree di interessamento tumorale francamente menzionate nel referto: questo dato raggiunge il 100% di sovrapposizione nel confronto con i referti clinici. Allo stesso tempo, la localizzazione del lato mostra valori elevati, con un *matching* del 97,92% nel confronto tra annotazioni e algoritmo *raw* e del 93% nel confronto tra i referti e le annotazioni.

La lateralità, con un'*accuracy* del 97,92%, rappresenta un dominio in cui la significatività clinica del risultato è netta. Questo risultato è importante perché gli errori di lateralità possono avere conseguenze cliniche devastanti. L'algoritmo, basandosi per la localizzazione sul calcolo del baricentro della lesione, adotta un approccio matematicamente definito ma non sempre allineato con la scelta clinica, soprattutto nelle lesioni che attraversano o si trovano in prossimità della linea mediana o che presentano una componente edematosa estesa all'emisfero controlaterale: questo aspetto si è frequentemente riscontrato nel processo di annotazione manuale.

Aree eloquenti, tumor core e edema

Particolarmente rilevante dal punto di vista clinico è la *performance* dell'algoritmo nella classificazione delle aree eloquenti, che raggiunge una *precision* del 91,89%. Questo significa che quando l'algoritmo segnala il coinvolgimento di un'area eloquente (motoria, del linguaggio espressivo, del linguaggio recettivo o visiva) tale segnalazione è corretta in oltre nove casi su dieci. In un contesto neurochirurgico,

dove la pianificazione preoperatoria dipende criticamente dall'identificazione del rischio di deficit post-operatori permanenti, un valore di *precision* così elevato rappresenta un risultato di rilievo applicativo.

Infatti, un falso positivo in questo dominio, ossia la segnalazione erronea del coinvolgimento di un'area eloquente, porterebbe a una maggiore cautela chirurgica, che potrebbe tradursi in una resezione meno estesa ma che preservi la funzione neurologica.²¹

Al contrario, il *recall* per le aree eloquenti si attesta al 73,91%, indicando che circa un quarto dei coinvolgimenti reali non viene rilevato dall'algoritmo. Questo dato merita attenzione, poiché un falso negativo in questo contesto, ossia la mancata segnalazione del coinvolgimento di un'area eloquente, potrebbe indurre il chirurgo a pianificare una resezione con un EOR maggiore con conseguente rischio di deficit neurologici permanenti.²¹

Il miglioramento del *recall* per le aree eloquenti rappresenta pertanto una priorità nello sviluppo futuro dell'algoritmo soprattutto nei suoi possibili sviluppi inerenti la neurochirurgia oncologica: questo algoritmo potrebbe proporsi come alleato per selezionare i pazienti candidabili a *task-based fMRI*, un esame di risonanza magnetica funzionale (basato sul segnale *BOLD*) finalizzato a rilevare in maniera non invasiva l'interessamento patologico di aree eloquenti⁶⁴, o all'*awake surgery*, una tecnica chirurgica che, quando percorribile, è raccomandata nei pazienti con gliomi che interessano le stesse aree precedentemente menzionate rappresentando ad oggi il *gold standard* per la valutazione della funzionalità delle aree cerebrali.⁶⁵

Per quanto riguarda il *tumor core*, la *precision* è del 62,75% e il *recall* dell'84,91%, con un *F1 score* del 72,16%. La prevalenza di un *recall* elevato rispetto alla *precision* indica che l'algoritmo tende a includere un numero maggiore di aree come coinvolte dal nucleo tumorale, generando alcuni falsi positivi ma limitando i falsi negativi. In un contesto neuro-oncologico, questa tendenza conservativa può essere considerata preferibile rispetto al comportamento opposto, poiché la sottostima dell'estensione tumorale ha implicazioni gravi per la pianificazione terapeutica rispetto a una leggera sovrastima: infatti, uno studio ha mostrato che il parametro

RTV (*residual tumor volume*) è un fattore prognostico più forte rispetto all'EOR (*extent of resection*).⁶⁶

L'edema mostra la stessa tendenza in forma ancora più pronunciata, con *precision* al 63,40% e *recall* al 95,10%: l'algoritmo è estremamente sensibile nel rilevare le aree edematose, ma a costo di una specificità ridotta. Questo comportamento è comprensibile dal punto di vista biologico, poiché l'edema peritumorale nei gliomi di alto grado tende ad estendersi lungo le fibre della sostanza bianca in modo difficilmente prevedibile (*finger-like pattern*)⁶⁷, rendendo la delimitazione precisa dei suoi margini una sfida anche per l'occhio esperto del neuroradiologo dove si può sommare ad altre alterazioni come le alterazioni della sostanza bianca su base microangiopatica che l'algoritmo non è ancora in grado di discriminare accuratamente.

In definitiva, emerge come l'algoritmo sia caratterizzato da elevata sensibilità: un aspetto che consente di ridurre al minimo i falsi negativi ma che tende frequentemente a classificare aree come positive quando in realtà non lo sono. Un possibile metodo per correggere questa tendenza sarebbe quello di impostare dei valori soglia più alti per classificare l'area come *affected*, riducendo così la sensibilità del sistema.

In ogni caso, questi dati devono essere interpretati alla luce della complessità insita nel compito stesso. L'atlante anatomico SRI24 annotato che è utilizzato dall'algoritmo suddivide il cervello in 281 aree distinte: una granularità anatomica di questo livello è di gran lunga superiore a quella normalmente adottata nella refertazione clinica routinaria, dove il neuroradiologo descrive tipicamente la sede della lesione in termini di lobi cerebrali, giri principali o strutture profonde facilmente riconoscibili e rilevanti dal punto di vista clinico. Ne consegue che il confronto diretto tra le aree identificate dall'algoritmo e quelle descritte nei referti risulta inevitabilmente penalizzato da questa asimmetria terminologica e di granularità.

In altre parole, una discordanza tra l'algoritmo e il referto non equivale necessariamente a un errore clinicamente significativo: l'algoritmo potrebbe identificare correttamente una lesione nel giro frontale medio sinistro, mentre il

referto descrive semplicemente una “*lesione localizzata al lobo frontale sinistro*”, senza che vi sia una reale divergenza diagnostica e clinica.

Valutazione di diametri e dimensioni

L’analisi del contenuto ha rivelato che, mentre lato e macroarea sono riportati nella totalità dei referti, il diametro massimo (oppure la valutazione lungo un singolo asse spaziale) è presente nel 22,2% dei casi e le dimensioni calcolate lungo due o tre assi spaziali nel 51,1%. Da questi dati è possibile derivare due principali punti di interesse:

- L’algoritmo analizza la lesione tridimensionalmente; questo aspetto gli consente di valutare con più precisione il diametro assiale massimo della lesione nello spazio. Nella quotidianità clinica difficilmente vengono a valutarsi angoli diversi rispetto a quelli canonici delle sezioni coronali, sagittali e assiali: di fatto la valutazione clinica è spesso una valutazione bidimensionale della lesione che tiene meno conto della tridimensionalità;
- L’errore relativo è più elevato quando il clinico esamina solo un’asse dimensionale della neoplasia (51,6%); il dato migliora se consideriamo almeno due assi dimensionali (19,9%): l’errore si riduce all’aumentare della definizione dimensionale che il clinico referta. Questo dato potrebbe essere un ulteriore elemento a favore dell’implementazione di sistemi che consentano una standardizzazione del referto e che permettano di ottenere una valutazione dimensionale più rigorosa.

Implicazioni per la refertazione automatica

Il confronto tra l’output strutturato dell’algoritmo e i referti testuali dei neuroradiologi ha messo in luce alcune lacune legate alla *pipeline* dell’algoritmo e alla disomogeneità intrinseca dei referti stessi. Questa disomogeneità riflette la mancanza di un *template* strutturato e standardizzato nella refertazione neuroradiologica di routine, un problema riconosciuto in letteratura.⁶⁸ La refertazione in linguaggio libero, sebbene ricca di sfumature cliniche, è intrinsecamente soggettiva e dipende dallo stile personale del refertante, dall’esperienza maturata, dal carico di lavoro e dal contesto clinico in cui viene

prodotta.⁶⁹ A questo va aggiunto il dettaglio che si è fatto riferimento solo agli esami di RMN, evitando di includere referti di altre metodiche come la tomografia computerizzata (TC), che però, vengono a volte portati come riferimento dai referti successivi. Paradossalmente, queste variabilità rappresentano uno degli argomenti più forti a favore dell'implementazione di sistemi di ARRG basati sull'AI: un algoritmo produce sempre gli stessi campi informativi per ogni lesione analizzata, garantendo una standardizzazione che favorisce la confrontabilità longitudinale degli esami, la comunicazione multidisciplinare e la raccolta di dati per studi di ricerca.⁷⁰ Tra i tentativi di standardizzazione si trova VASARI (*Visually Accessible Rembrandt Images*): un *feature set* semantico costruito per standardizzare la descrizione visiva su RM dei gliomi e rendere più riproducibili referti e annotazioni. Inoltre, si è visto che alcune *features* di VASARI sono performanti anche nella predizione dell'*overall survival* e dello status mutazionale di IDH.⁷¹

È importante sottolineare che l'obiettivo di un sistema di ARRG in ambito neuroradiologico non è quello di sostituire il referto medico, ma di costituire una bozza strutturata che il neuroradiologo possa verificare, integrare e validare: a conferma di ciò, uno studio ha mostrato che il 91.0% dei report generati da un modello *vision-language* era clinicamente accettabile, ma solo il 37.8% era qualitativamente superiore o comparabile all'originale;⁷² con conclusione implicita che è sempre necessaria una supervisione da parte di un professionista medico.

In questo senso, la *pipeline* sviluppata in collaborazione con il laboratorio Almage-Lab-Zip di UNIMORE rappresenta un possibile approccio metodologicamente solido: partendo da segmentazioni già disponibili ottenute mediante un algoritmo di *deep learning*, estrae in modo sistematico le informazioni anatomiche e metriche più rilevanti e le traduce in un testo descrittivo standardizzato. Il neuroradiologo mantiene il ruolo centrale di supervisore e validatore, in linea con le indicazioni della letteratura più recente che esclude una sostituzione del clinico da parte dell'intelligenza artificiale, prevedendo invece una ridefinizione del suo profilo professionale verso compiti di maggiore complessità cognitiva e relazionale.⁴⁶

Limiti dello studio e prospettive future

Il presente studio presenta alcuni limiti che devono essere riconosciuti. In primo luogo, la dimensione campionaria, seppur adeguata ai fini di un'analisi preliminare e destinata ad aumentare al completamento delle annotazioni, limita la potenza statistica delle conclusioni e non consente un'adeguata analisi stratificata per diagnosi istologica (glioblastoma, astrocitoma e oligodendroglioma). La stratificazione istologica potrebbe rivelare differenze di performance legate alle diverse caratteristiche morfologiche e di estensione tumorale di ciascun sottotipo. In secondo luogo, come già discusso, l'asimmetria di granularità tra la nomenclatura anatomica dell'atlante e quella adottata nei referti clinici rappresenta un limite metodologico che penalizza il confronto diretto.

Un ulteriore limite riguarda la natura monomodale del corpus di referti utilizzato per il confronto testuale. L'esclusiva inclusione di referti di risonanza magnetica, sebbene metodologicamente coerente con il protocollo d'imaging adottato e con le caratteristiche tecniche dell'algoritmo, non riflette pienamente la complessità del percorso diagnostico reale del paziente neuro-oncologico. Nella pratica clinica, la valutazione iniziale di una sospetta lesione cerebrale avviene frequentemente mediante tomografia computerizzata (TC), spesso eseguita in regime d'urgenza (vista la disponibilità maggiore della metodica)⁷³: essa è portatrice di informazioni complementari rispetto alla risonanza magnetica, in particolare riguardo alla presenza di calcificazioni, emorragie acute e alterazioni ossee.⁷⁴ L'estensione futura dell'algoritmo a referti provenienti da metodiche diverse consentirebbe di valutare la robustezza dell'algoritmo in un contesto diagnostico più ecologico e rappresentativo, avvicinandosi maggiormente alla realtà del flusso di lavoro clinico quotidiano e ampliando il potenziale traslazionale dello strumento.

Infine, un ulteriore sviluppo di rilievo riguarderebbe l'integrazione della *pipeline* con modelli di linguaggio di grandi dimensioni (LLM), che potrebbero consentire la generazione di testi descrittivi più fluidi e clinicamente contestualizzati rispetto alle frasi predefinite attualmente utilizzate, avvicinando ulteriormente l'output automatico alla qualità narrativa del referto medico esperto, come suggerito dalle architetture più recenti nell'ambito dell'ARRG.^{43,75}

CONCLUSIONI

I risultati preliminari dimostrano che il sistema raggiunge livelli di accuratezza clinicamente rilevanti nella lateralità e nella localizzazione della lesione neoplastica, nell'identificazione delle aree eloquenti interessate e nella stima dell'estensione di *tumor core* e edema. L'algoritmo si propone pertanto come uno strumento concreto di supporto alla refertazione strutturata e standardizzata, con possibili ricadute positive sul flusso di lavoro clinico e sulla comunicazione multidisciplinare, potenziando l'attività del neuroradiologo, senza sostituirlo. È, infatti, importante considerare l'impatto che i sistemi di AI possono avere sulla vita delle persone e sulle professioni: l'AI non è da considerarsi come uno strumento "moralmente neutro", bensì come uno strumento di cui deve essere valutato sia lo scopo di utilizzo che la progettazione a priori nel rispetto del paziente e del medico stesso; per citare l'enciclica "*Magnifica Humanitas*" di Papa Leone XIV: *"In realtà, ogni artefatto tecnico porta con sé scelte e priorità: ciò che misura, ciò che ignora, ciò che ottimizza e il modo in cui classifica persone e situazioni. Se un sistema viene concepito o impiegato in modo da trattare alcune vite come meno degne, o da escluderle senza possibilità di appello, esso non è un semplice strumento "da usare bene": introduce già un criterio che contraddice la dignità inalienabile della persona"*.⁷⁶

BIBLIOGRAFIA

1. Patel, A. P. *et al.* Global, regional, and national burden of brain and other CNS cancer, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* **18**, 376–393 (2019).
2. Emilia Romagna. Registro tumori Emilia Romagna.
3. Ostrom, Q. T. *et al.* The epidemiology of glioma in adults: a ‘state of the science’ review. *Neuro-Oncology* **16**, 896–913 (2014).
4. Linos, E., Raine, T., Alonso, A. & Michaud, D. Atopy and Risk of Brain Tumors: A Meta-analysis. *JNCI Journal of the National Cancer Institute* **99**, 1544–1550 (2007).
5. Presenting signs and symptoms in brain tumors. in *Handbook of Clinical Neurology* vol. 134 19–26 (Elsevier, 2016).
6. Chen, H. *et al.* Mutant IDH1 and seizures in patients with glioma. *Neurology* **88**, 1805–1813 (2017).
7. Perry, J. R. Thromboembolic disease in patients with high-grade glioma. *Neuro-Oncology* **14**, iv73–iv80 (2012).
8. Caranci, F., Briganti, G., Elefante, A. & Russo, C. *Neuroradiologia*. (Idelson-Gnocchi).
9. Louis, D. N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* **23**, 1231–1251 (2021).
10. Price, M. *et al.* CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2018–2022. *Neuro-Oncology* **27**, iv1–iv66 (2025).

11. McNamara, C. *et al.* 2021 WHO classification of tumours of the central nervous system: a review for the neuroradiologist. *Neuroradiology* **64**, 1919–1950 (2022).
12. Gritsch, S., Batchelor, T. T. & Gonzalez Castro, L. N. Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. *Cancer* **128**, 47–58 (2022).
13. Osborn, A. G., Louis, D. N., Poussaint, T. Y., Linscott, L. L. & Salzman, K. L. The 2021 World Health Organization Classification of Tumors of the Central Nervous System: What Neuroradiologists Need to Know. *AJNR Am J Neuroradiol* **43**, 928–937 (2022).
14. Yamashita, K. *et al.* MR Imaging–Based Analysis of Glioblastoma Multiforme: Estimation of *IDH1* Mutation Status. *AJNR Am J Neuroradiol* **37**, 58–65 (2016).
15. Branzoli, F. *et al.* Highly specific determination of IDH status using edited in vivo magnetic resonance spectroscopy. *Neuro-Oncology* **20**, 907–916 (2018).
16. Hegi, M. E. *et al.* MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *n engl j med* (2005).
17. Patel, S. H. *et al.* T2–FLAIR Mismatch, an Imaging Biomarker for IDH and 1p/19q Status in Lower-grade Gliomas: A TCGA/TCIA Project. *Clinical Cancer Research* **23**, 6078–6085 (2017).
18. Wen, P. Y. *et al.* Glioblastoma in adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro-Oncology* **27**, 2751–2788 (2025).

19. Johnson, D. R. *et al.* Genetically Defined Oligodendroglioma Is Characterized by Indistinct Tumor Borders at MRI. *AJNR Am J Neuroradiol* **38**, 678–684 (2017).
20. Park, Y. W. *et al.* The 2021 WHO Classification for Gliomas and Implications on Imaging Diagnosis: Part 1—Key Points of the Fifth Edition and Summary of Imaging Findings on Adult-Type Diffuse Gliomas. *Magnetic Resonance Imaging* **58**, 677–689 (2023).
21. Weller, M. *et al.* EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol* **18**, 170–186 (2021).
22. Goldbrunner, R. *et al.* EANS-EANO guidelines on the extent of resection in gliomas. *Neuro-Oncology* **28**, 38–54 (2026).
23. Mohile, N. A. *et al.* Therapy for Diffuse Astrocytic and Oligodendroglial Tumors in Adults: ASCO-SNO Guideline. *THE BOTTOM LINE* (2026).
24. Vogelbaum, M. A. *et al.* Treatment for Brain Metastases: ASCO-SNO-ASTRO Guideline. *THE BOTTOM LINE* (2026).
25. Roth, P. *et al.* Neurological and vascular complications of primary and secondary brain tumours: EANO-ESMO Clinical Practice Guidelines for prophylaxis, diagnosis, treatment and follow-up. *Annals of Oncology* **32**, 171–182 (2021).
26. Pace, A. *et al.* European Association for Neuro-Oncology (EANO) guidelines for palliative care in adults with glioma. *The Lancet Oncology* **18**, e330–e340 (2017).

27. Ostrom, Q. T. *et al.* National-level overall survival patterns for molecularly-defined diffuse glioma types in the United States. *Neuro-Oncology* **25**, 799–807 (2023).
28. Quintarelli, S. *Intelligenza Artificiale - Cos'è Davvero, Come Funziona, Che Effetti Avrà.* (Bollati Boringhieri).
29. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence – an ESR white paper. *Insights Imaging* **10**, 44 (2019).
30. Hatt, M. *et al.* Joint EANM/SNMMI guideline on radiomics in nuclear medicine: Jointly supported by the EANM Physics Committee and the SNMMI Physics, Instrumentation and Data Sciences Council. *Eur J Nucl Med Mol Imaging* **50**, 352–375 (2023).
31. Minaee, S. *et al.* Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–1 (2021) doi:10.1109/TPAMI.2021.3059968.
32. Janiesch, C., Zschech, P. & Heinrich, K. Machine learning and deep learning. *Electron Markets* **31**, 685–695 (2021).
33. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Preprint at <https://doi.org/10.48550/arXiv.1811.10154> (2019).
34. IONOS. <https://www.ionos.it/digitalguide/online-marketing/marketing-sui-motori-di-ricerca/deep-learning-vs-machine-learning/>.
35. Seo, H. *et al.* Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Medical Physics* **47**, (2020).

36. Melazzini, L. *et al.* AI for image quality and patient safety in CT and MRI. *Eur Radiol Exp* **9**, 28 (2025).
37. Avanzo, M. *et al.* Machine and deep learning methods for radiomics. *Medical Physics* **47**, (2020).
38. Pecchi, A. *et al.* DCE-MRI Radiomic analysis in triple negative ductal invasive breast cancer. Comparison between BRCA and not BRCA mutated patients: Preliminary results. *Magnetic Resonance Imaging* **113**, 110214 (2024).
39. Pecchi, A. *et al.* Preliminary exploration of radiomic mammographic analysis in triple negative breast cancer related to BRCA profile. *Sci Rep* **16**, 8765 (2026).
40. Najjar, R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics* **13**, 2760 (2023).
41. Khalighi, S. *et al.* Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. *npj Precis. Onc.* **8**, 80 (2024).
42. Rajendran, S. *et al.* Automated Segmentation of Brain Tumor MRI Images Using Deep Learning. *IEEE Access* **11**, 64758–64768 (2023).
43. Sloan, P., Clatworthy, P., Simpson, E. & Mirmehdi, M. Automated Radiology Report Generation: A Review of Recent Advances. *IEEE Rev. Biomed. Eng.* **18**, 368–387 (2025).
44. Elliott, D. & Keller, F. Comparing Automatic Evaluation Measures for Image Description. in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* 452–457 (Association for

- Computational Linguistics, Baltimore, Maryland, 2014). doi:10.3115/v1/P14-2074.
45. Liu, C.-F. *et al.* Automatic comprehensive radiological reports for clinical acute stroke MRIs. *Commun Med* **3**, 95 (2023).
 46. Coker, E., Lockspeiser, T. & Hilgenberg, S. L. Artificial Intelligence Healthcare Industry Expert Perspectives: How Artificial Intelligence Will Impact Physician Roles and Medical Education. *Academic Pediatrics* **26**, 103225 (2026).
 47. Benanti, P. *Human in the Loop: Decisione Umane e Intelligenze Artificiali*. (Mondadori Università).
 48. Guan, X. *et al.* 3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework. *BMC Med Imaging* **22**, 6 (2022).
 49. Liu, Z. *et al.* A review of self-supervised, generative, and few-shot deep learning methods for data-limited magnetic resonance imaging segmentation. *NMR in Biomedicine* **37**, e5143 (2024).
 50. Meier, R. *et al.* Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry. *Sci Rep* **6**, 23376 (2016).
 51. Luo, X. *et al.* Automated segmentation of brain metastases with deep learning: A multi-center, randomized crossover, multi-reader evaluation study. *Neuro-Oncology* **26**, 2140–2151 (2024).
 52. Schmidt-Hansen, M., Berendse, S. & Hamilton, W. Symptomatic diagnosis of cancer of the brain and central nervous system in primary care: a systematic review. *Family Practice* **32**, 618–623 (2015).

53. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. *Neuro-Oncology*
<https://doi.org/10.1093/neuonc/nov095> (2015) doi:10.1093/neuonc/nov095.
54. Menze, B. H. *et al.* The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2015).
55. Farioli, F. Deep learning per la segmentazione dei tumori gliali diffusi dell'adulto: fine-tuning e soluzione alle sequenze mancanti. (UNIMORE).
56. Lasocki, A., Gaillard, F., Tacey, M., Drummond, K. & Stuckey, S. Multifocal and multicentric glioblastoma: Improved characterisation with FLAIR imaging and prognostic implications. *Journal of Clinical Neuroscience* **31**, 92–98 (2016).
57. Rohlfing, T., Zahr, N. M., Sullivan, E. V. & Pfefferbaum, A. The SRI24 multichannel atlas of normal adult human brain structure. *Human Brain Mapping* **31**, 798–819 (2010).
58. Avants, B. B., Tustison, N. & Song, G. Advanced normalization tools (ANTS). *The Insight Journal* vol. 2 1–35 (2009).
59. Pasquini, L., Peck, K. K., Jenabi, M. & Holodny, A. Functional MRI in Neuro-Oncology: State of the Art and Future Directions. *Radiology* **308**, e222028 (2023).
60. Chiu, C.-L., Clack, N., & the napari community. napari: a Python Multi-Dimensional Image Viewer Platform for the Research Community. *Microscopy and Microanalysis* **28**, 1576–1577 (2022).
61. Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* **31**, 1116–1128 (2006).

62. <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=it>.
63. Di Carlo, D. T., Cagnazzo, F., Benedetto, N., Morganti, R. & Perrini, P. Multiple high-grade gliomas: epidemiology, management, and outcome. A systematic review and meta-analysis. *Neurosurg Rev* **42**, 263–275 (2019).
64. Koc, N. A. *et al.* Theoretical, Technical, and Analytical Foundations of Task-Based and Resting-State Functional Magnetic Resonance Imaging (fMRI)—A Narrative Review. *Biomedicines* **14**, 333 (2026).
65. Sattari, S. A. *et al.* Awake Versus Asleep Craniotomy for Patients With Eloquent Glioma: A Systematic Review and Meta-Analysis. *Neurosurgery* **94**, 38–52 (2024).
66. Gerritsen, J. K. W. *et al.* Impact of maximal extent of resection on postoperative deficits, patient functioning, and survival within clinically important glioblastoma subgroups. *Neuro-Oncology* **25**, 958–972 (2023).
67. Hu, L. S., Hawkins-Daarud, A., Wang, L., Li, J. & Swanson, K. R. Imaging of intratumoral heterogeneity in high-grade glioma. *Cancer Letters* **477**, 97–106 (2020).
68. Alruwaili, A. R. *et al.* A Critical Examination of Academic Hospital Practices—Paving the Way for Standardized Structured Reports in Neuroimaging. *JCM* **13**, 4334 (2024).
69. Crombé, A. *et al.* What Influences the Way Radiologists Express Themselves in Their Reports? A Quantitative Assessment Using Natural Language Processing. *J Digit Imaging* **35**, 993–1007 (2022).

70. Hasani, A. M. *et al.* Evaluating the performance of Generative Pre-trained Transformer-4 (GPT-4) in standardizing radiology reports. *Eur Radiol* **34**, 3566–3574 (2023).
71. Azizova, A. *et al.* Ten Years of VASARI Glioma Features: Systematic Review and Meta-Analysis of Their Impact and Performance. *AJNR Am J Neuroradiol* **45**, 1053–1062 (2024).
72. Park, Y. W. *et al.* A robust vision language model for molecular status prediction and radiology report generation in adult-type diffuse gliomas. *npj Digit. Med.* **9**, 406 (2026).
73. Lo, B. M. *et al.* Clinical Policy: Critical Issues in the Evaluation of Adult Patients With Suspected Transient Ischemic Attack in the Emergency Department. *Annals of Emergency Medicine* **68**, 354-370.e29 (2016).
74. Gajjar, A. *et al.* Pediatric Central Nervous System Cancers, Version 2.2025, NCCN Clinical Practice Guidelines In Oncology. *Journal of the National Comprehensive Cancer Network* **23**, 113–130 (2025).
75. Busch, F. *et al.* Large language models for structured reporting in radiology: past, present, and future. *Eur Radiol* **35**, 2589–2602 (2024).
76. Leone XIV. *Magnifica Humanitas*. (Libreria editrice Vaticana, 2026).