

Biopsia virtuale delle lesioni polmonari mediante radiomica TC e machine learning

Sviluppo di un modello multiclass per la caratterizzazione non invasiva
di NSCLC, SCLC, metastasi e lesioni benigne

Facoltà di Medicina e Chirurgia

Scuola di specializzazione in Radiodiagnostica

Clary Venzano

Matricola 3767887

Relatore Prof. Carlo Martinoli

Correlatore Dott. Emanuele Barabino

A.A. 2025-2026

Abstract

La caratterizzazione di una lesione polmonare resta uno dei problemi più delicati della diagnostica toracica. La TC è la metodica di riferimento per individuarla e descriverne la morfologia, ma nella maggior parte dei casi la diagnosi definitiva si basa comunque sulla conferma istologica. La biopsia tissutale rimane il gold standard, però è invasiva, non priva di complicanze, e per sua natura descrive solo la porzione di tessuto effettivamente campionata.

Negli ultimi anni la radiomica ha mostrato che dalle immagini si possono estrarre informazioni quantitative non colte dal solo occhio del radiologo. È in questa cornice che si parla di biopsia virtuale: un modo non invasivo di interrogare l'intero volume della lesione, nella speranza di ricavarne indizi sulla sua biologia.

In questo lavoro abbiamo verificato se feature radiomiche estratte da TC potessero classificare le lesioni polmonari in quattro categorie – carcinoma non a piccole cellule (NSCLC), carcinoma a piccole cellule (SCLC), lesioni benigne e metastasi. L'idea di fondo era capire se un modello di machine learning basato su radiomica potesse avvicinarsi, almeno in parte, al concetto di biopsia virtuale pre-istologica.

Il dataset comprendeva 591 lesioni polmonari in 434 pazienti. Abbiamo confrontato 14 feature set e 7 algoritmi di machine learning, per un totale di 98 configurazioni sperimentali, validate con cross-validazione ripetuta 5×5 e separazione dei dati a livello di paziente. In considerazione della marcata disomogeneità tra le classi, come metriche principali sono state balanced accuracy, macro-F1, macro-AUC one-vs-rest e macro-AUPRC, più informative della semplice accuracy in questo contesto.

Il modello con la miglior balanced accuracy è risultato ExtraTrees addestrato sull'intero set di 851 feature (balanced accuracy $0,556 \pm 0,059$; macro-AUC $0,757 \pm 0,044$). Interessante notare che ExtraTrees addestrato solo sulle 669 feature caratterizzate dal Noise Sensitivity Test ha ottenuto risultati praticamente sovrapponibili (balanced accuracy $0,543 \pm 0,034$; macro-AUC $0,755 \pm 0,030$) – segno che si può ridurre la dimensionalità del problema senza perdere granché in termini predittivi.

Guardando alle singole classi, gli SCLC sono la categoria riconosciuta meglio in assoluto: AUC one-vs-rest di 0,907, precision 0,978, recall 0,789, F1 0,874. Molto più

complicata invece la distinzione tra NSCLC, metastasi e lesioni benigne, tra cui il modello confonde frequentemente le predizioni.

Nel complesso, questi risultati dicono che la radiomica TC è in grado di intercettare un segnale quantitativo utile alla caratterizzazione non invasiva delle lesioni, ma che la classificazione multiclass completa è ancora lontana dal poter competere con la diagnosi istologica. Il concetto di biopsia virtuale si dimostra però particolarmente promettente proprio per gli SCLC, il che apre a un possibile sviluppo futuro incentrato su un modello dedicato SCLC vs non-SCLC. Il Noise Sensitivity Test si conferma un buon strumento per selezionare feature più affidabili in un workflow radiomico pensato per la pratica clinica.

Parole chiave: biopsia virtuale, radiomica, TC torace, lesioni polmonari, machine learning, SCLC, NSCLC, metastasi, Noise Sensitivity Test.

Indice

1. Introduzione	6
1.1 Le lesioni polmonari nella pratica clinica	6
1.2 Il ruolo della TC nella caratterizzazione delle lesioni polmonari	7
1.3 Limiti della biopsia tissutale	7
1.4 Il concetto di biopsia virtuale	8
1.5 Radiomica e caratterizzazione quantitativa dell'immagine	9
1.6 Machine learning e modelli predittivi radiomici	9
1.7 Il problema della robustezza radiomica	10
1.8 Razionale dello studio	10
2. Obiettivi dello studio	12
2.1 Obiettivo primario	12
2.2 Obiettivi secondari	12
2.3 Ipotesi sperimentale	13
3. Materiali e metodi	14
3.1 Disegno dello studio	14
3.2 Popolazione in studio	14
3.3 Classificazione diagnostica delle lesioni	15
3.4 Imaging TC e preprocessing	15
3.5 Segmentazione delle lesioni	16
3.6 Estrazione delle feature radiomiche	18
3.7 Noise Sensitivity Test	18
3.8 Feature set analizzati	18
3.9 Modelli di machine learning	20
3.10 Validazione interna	20
3.11 Metriche di valutazione	21
4. Risultati	22
4.1 Descrizione generale del dataset	22
4.2 Feature radiomiche e costruzione della biopsia virtuale	22
4.3 Performance globale dei modelli multiclass	23
4.4 Confronto tra algoritmi	25
4.5 Il modello con maggiore accuracy	26
4.6 Analisi per classe del modello NST_ranked_all_669 + ExtraTrees	27

4.7 Matrice di confusione del modello NST_ranked_all_669 + ExtraTrees.....	28
4.8 Il modello NST_top_300 + ExtraTrees	28
5. Discussione	30
5.1 La radiomica TC come biopsia virtuale	30
5.2 Interpretazione della performance multiclass	30
5.3 Il fenotipo radiomico degli SCLC	31
5.4 Difficoltà nella distinzione tra NSCLC, metastasi e lesioni benigne	32
5.5 Ruolo del Noise Sensitivity Test.....	32
5.6 Implicazioni cliniche.....	33
5.7 Limiti dello studio	33
5.8 Prospettive future	34
6. Conclusioni	36
Bibliografia	37

Introduzione

1.1 Le lesioni polmonari nella pratica clinica

In radiologia toracica si possono incontrare lesioni polmonari molto differenti fra loro: alterazioni benigne, esiti infiammatori, tumori primitivi del polmone, metastasi da neoplasie extrapolmonari. Una diagnosi tempestiva condiziona in modo diretto il percorso diagnostico-terapeutico del paziente.

In caso di una lesione polmonare sospetta, il percorso diagnostico richiede la formulazione di una serie di quesiti fondamentali volti a definirne la natura[1, 2]. In primo luogo, è necessario stabilire se la lesione presenti caratteristiche maggiormente suggestive per una condizione benigna o per una patologia neoplastica. In quest'ultimo caso, occorre differenziare un tumore polmonare primitivo da una localizzazione metastatica e, qualora si tratti di una neoplasia primitiva, procedere alla distinzione tra carcinoma polmonare non a piccole cellule (non-small cell lung cancer, NSCLC) e carcinoma polmonare a piccole cellule (small cell lung cancer, SCLC)[3–5]. Tale classificazione riveste un ruolo clinico determinante, poiché queste entità differiscono significativamente per comportamento biologico, modalità di diffusione, prognosi e strategie terapeutiche.

L'NSCLC è di gran lunga il più frequente e comprende a sua volta sottotipi eterogenei – adenocarcinoma, carcinoma squamoso, carcinoma a grandi cellule. L'SCLC è più raro ma decisamente più aggressivo: cresce rapidamente, si dissemina precocemente e richiede spesso un trattamento sistemico tempestivo.

Le metastasi polmonari meritano un discorso a parte: il polmone è tra le sedi più comuni di malattia secondaria, e distinguere un tumore primitivo da una metastasi può risultare difficile, soprattutto quando la lesione è unica, la storia oncologica del paziente è complessa o il tumore primitivo non è noto.

Infine, non va dimenticato che alcune lesioni benigne – processi infiammatori, granulomi, esiti fibrotici o cicatriziali – possono mimare radiologicamente una neoplasia, imponendo ulteriori accertamenti[6].

La diagnosi differenziale, insomma, nasce dall'integrazione di imaging, clinica, anamnesi, follow-up e, spesso, istologia. Ma anche con un imaging moderno e sofisticato

come quello di cui disponiamo oggi, molte lesioni restano indeterminate alla sola lettura morfologica.

1.2 Il ruolo della TC nella caratterizzazione delle lesioni polmonari

La TC resta lo standard per lo studio delle lesioni polmonari: consente di individuare anche noduli molto piccoli e di descriverne dimensioni, margini, densità, eventuali calcificazioni o cavitazioni, componente solida o subsolida, rapporti con le strutture vicine[7, 8].

Nella pratica quotidiana, però, questa valutazione è per lo più visiva e semiquantitativa. Il radiologo legge il reperto alla luce di morfologia, andamento temporale, sede e contesto clinico – un approccio funzionale, ma con limiti evidenti: caratteristiche di lesioni maligne e benigne spesso si sovrappongono, e l'occhio umano difficilmente coglie pattern quantitativi complessi nascosti dentro l'immagine.

Ogni esame TC è un insieme di numeri, voxel per voxel, che descrivono la distribuzione spaziale delle densità dentro la lesione – informazioni che possono riflettere eterogeneità interna, necrosi, componente cellulare, fibrosi, architettura tissutale, vascolarizzazione. Solo una parte minima di questo contenuto informativo finisce però nel referto convenzionale.

La radiomica nasce dall'idea di sfruttare in modo sistematico questo patrimonio di dati, trasformando l'immagine in un insieme di variabili quantitative che si prestano ad analisi statistiche e a modelli di machine learning.

1.3 Limiti della biopsia tissutale

La biopsia tissutale resta il riferimento diagnostico per le lesioni polmonari: permette una diagnosi istologica, distingue tra sottotipi tumorali e, quando il materiale è sufficiente, consente analisi immunoistochimiche e molecolari[9, 10].

La biopsia polmonare può avere limiti concreti. In primis, è una procedura invasiva e, per quanto eseguita da mani esperte, può associarsi a complicanze – pneumotorace, sanguinamento, dolore, emottisi, e talvolta la necessità di osservazione o trattamento. Il rischio varia in base a sede della lesione, dimensioni, distanza dalla pleura, presenza di enfisema, traiettoria dell'ago e condizioni generali del paziente.

In secondo luogo, c'è il problema del campionamento: la biopsia racconta solo la porzione di tessuto prelevata, mentre le lesioni tumorali sono spesso eterogenee al loro interno, con aree a diversa cellularità, necrosi, ipossia, fibrosi o infiltrato infiammatorio. Un singolo prelievo può non essere rappresentativo dell'intera lesione.

Infine, la biopsia può risultare non diagnostica o che il materiale sia insufficiente, con la necessità di ripetere la procedura – e quindi rischi, tempi e costi che si moltiplicano.

Nonostante questi limiti, la biopsia polmonare resta indispensabile nella stragrande maggioranza dei percorsi oncologici. Ma questi limiti aprono uno spazio concreto per strumenti complementari, non invasivi, capaci di dare informazioni aggiuntive prima, durante o dopo il percorso biotico.

1.4 Il concetto di biopsia virtuale

Con **biopsia virtuale** si intende la possibilità di ottenere informazioni sulla natura biologica di una lesione analizzando in modo non invasivo le immagini[11, 12]. A differenza della biopsia tradizionale, che campiona fisicamente una porzione di tessuto, la biopsia virtuale studia l'intero volume lesionale attraverso i dati di imaging.

La biopsia virtuale non deve essere intesa come un sostituto immediato della biopsia istologica, ma come uno strumento complementare. Potrebbe, in linea di principio:

- stimare la probabilità che una lesione appartenga a una certa categoria diagnostica;
- segnalare pattern suggestivi di specifici sottotipi tumorali;
- aiutare a stabilire le priorità del percorso diagnostico;
- orientare la scelta della sede più rappresentativa da biopsiare;
- integrarsi con i dati clinico-radiologici nella stratificazione del rischio;
- seguire nel tempo l'eterogeneità della lesione senza doverla ricampionare.

Il polmone è un terreno particolarmente adatto a questo tipo di approccio, dato che la biopsia può essere tecnicamente impegnativa e che distinguere lesioni primitive, metastatiche e benigne ha conseguenze terapeutiche importanti.

Per dimostrare che questo concetto abbia una qualche credibilità scientifica, occorre provare che le immagini contengano davvero un segnale quantitativo legato alla natura della lesione, e che questo segnale non sia semplicemente rumore, artefatto o variabilità tecnica.

1.5 Radiomica e caratterizzazione quantitativa dell'immagine

La radiomica permette di estrarre dalle immagini un numero elevato di feature quantitative, che descrivono aspetti diversi della regione di interesse: forma, intensità, texture, eterogeneità spaziale[13–15].

Le famiglie principali sono:

- feature di forma – volume, superficie, compattezza, geometria della lesione;
- feature di primo ordine – distribuzione dei valori di intensità dei voxel;
- feature di texture – relazioni spaziali tra livelli di grigio;
- feature derivate da trasformazioni dell'immagine, come wavelet o altri filtri matematici.

L'ipotesi di fondo è che alcune di queste caratteristiche quantitative rispecchino proprietà biologiche della lesione[16]. In questo senso la radiomica è uno strumento di fenotipizzazione quantitativa: non si limita a descrivere ciò che è visibile, ma cerca pattern nascosti, potenzialmente espressione del comportamento biologico del tessuto.

Nell'ottica della biopsia virtuale, le feature radiomiche rappresentano i "marcatori quantitativi" che si ricavano dall'immagine: giocano, sul piano concettuale, un ruolo simile a quello delle variabili istologiche o molecolari, cioè aiutare a inferire la natura della lesione.

1.6 Machine learning e modelli predittivi radiomici

Con centinaia di feature a disposizione, servono strumenti computazionali capaci di riconoscere pattern complessi quali il machine learning, che permette di costruire modelli predittivi in grado di imparare le relazioni tra feature di imaging e categorie diagnostiche.

Nel nostro caso, il modello viene addestrato a riconoscere differenze radiomiche tra NSCLC, SCLC, metastasi e lesioni benigne. Se il modello distingue queste classi con

performance superiori al caso, significa che le immagini TC contengono davvero informazioni quantitative legate alla natura della lesione.

Tuttavia, i modelli di machine learning applicati alla radiomica sono particolarmente esposti al rischio di overfitting[17, 18], soprattutto quando il numero di feature è alto rispetto al numero di pazienti, il cosiddetto problema “large p, small n”[19]; un modello può imparare pattern casuali, specifici del dataset di training, che non si generalizzano affatto a nuovi pazienti.

1.7 Il problema della robustezza radiomica

Uno dei limiti più discussi della radiomica riguarda la variabilità delle feature, che possono essere influenzate da un numero considerevole di fattori tecnici: parametri di acquisizione, spessore di strato, kernel di ricostruzione, dose, rumore, preprocessing, resampling, discretizzazione, segmentazione[20–22].

Questo aspetto concerne il concetto di biopsia virtuale: per essere credibile, dovrebbe basarsi su feature che riflettono proprietà della lesione, non caratteristiche accidentali dell'immagine. Se una feature è fortemente influenzata dal rumore o dai parametri tecnici, il suo valore biologico diventa incerto.

Selezionare feature robuste diventa quindi un passaggio cruciale: non basta ottenere una buona performance predittiva, bisogna anche dimostrare che il modello si appoggia a informazioni relativamente stabili, e plausibilmente trasferibili ad altri contesti.

In questo lavoro è stato usato un approccio basato sul Noise Sensitivity Test (NST) proprio per caratterizzare le feature rispetto al rumore[20], selezionare quelle potenzialmente più robuste e capire se ridurre il set radiomico comportasse una perdita di performance.

1.8 Razionale dello studio

Lo studio nasce all'incrocio di tre esigenze.

- Clinica: caratterizzare in modo più accurato e non invasivo le lesioni polmonari, aiutando a distinguere NSCLC, SCLC, metastasi e lesioni benigne.

- Biologica: capire se le lesioni polmonari abbiano fenotipi radiomici differenti, misurabili tramite feature quantitative estratte dalla TC.
- Metodologica: costruire un workflow di biopsia virtuale che non punti solo a massimizzare la performance, ma anche a selezionare feature più robuste e meno sensibili al rumore.

Da queste premesse, in questo studio cercheremo di valutare se radiomica TC, machine learning e selezione delle feature tramite NST insieme possano rappresentare una base solida per lo sviluppo di una biopsia virtuale delle lesioni polmonari.

Obiettivi dello studio

2.1 Obiettivo primario

L'obiettivo primario dello studio è valutare la fattibilità di una **biopsia virtuale radiomica** delle lesioni polmonari, tramite modelli di machine learning applicati a immagini TC.

In particolare, è stata verificata la capacità dei modelli radiomici di classificare le lesioni in quattro categorie:

1. carcinoma polmonare non a piccole cellule;
2. carcinoma polmonare a piccole cellule;
3. lesioni benigne;
4. metastasi polmonari.

2.2 Obiettivi secondari

Gli obiettivi secondari dello studio sono:

- ✓ valutare la performance globale di modelli multiclass basati su feature radiomiche TC;
- ✓ capire quali categorie diagnostiche fossero più facilmente riconoscibili tramite biopsia virtuale radiomica;
- ✓ approfondire in particolare la riconoscibilità radiomica degli SCLC;
- ✓ confrontare diversi algoritmi di machine learning tra loro;
- ✓ valutare l'effetto della selezione delle feature tramite Noise Sensitivity Test;
- ✓ verificare se togliere le feature non caratterizzate dal NST comportasse una perdita significativa di performance;
- ✓ analizzare i principali pattern di errore del classificatore multiclass;
- ✓ individuare possibili direzioni per modelli più specifici e più vicini a un uso clinico reale.

2.3 Ipotesi sperimentale

L'ipotesi principale era che le immagini TC delle lesioni polmonari contenessero un segnale quantitativo legato alla loro natura diagnostica, estraibile con la radiomica e interpretabile con modelli di machine learning.

Una seconda ipotesi era che selezionare le feature tramite Noise Sensitivity Test potesse contribuire a costruire modelli più robusti, riducendo la dimensionalità del problema senza compromettere in modo sostanziale la performance predittiva.

CAPITOLO 3

Materiali e metodi

3.1 Disegno dello studio

Lo studio ha disegno sperimentale retrospettivo, ed è stato condotto su immagini TC di lesioni polmonari. Per ciascuna lesione polmonare sono state estratte feature radiomiche, sulla cui base sono stati poi sviluppati modelli di machine learning per la classificazione multiclass.

Il workflow si è articolato così:

- ✓ identificazione della popolazione in studio;
- ✓ classificazione diagnostica delle lesioni;
- ✓ segmentazione delle lesioni;
- ✓ estrazione delle feature radiomiche;
- ✓ caratterizzazione delle feature tramite Noise Sensitivity Test;
- ✓ costruzione di diversi feature set;
- ✓ addestramento e validazione dei modelli;
- ✓ valutazione della performance globale e per classe.

3.2 Popolazione in studio

Il dataset finale comprendeva 591 lesioni polmonari in 434 pazienti sottoposti a biopsia toracica da Gennaio 2014 a Gennaio 2026 presso la Radiologia Generale e Interventistica dell'IRCCS Ospedale Policlinico San Martino.

In alcuni casi la lesione è stata rappresentata più volte all'interno del dataset, acquisita con modalità diverse ad esempio con TC-PET o con TC mediante mezzo di contrasto. Onde evitare leakage informativo, la suddivisione tra training e test set è stata fatta a livello di paziente, in modo che le lesioni dello stesso soggetto non finissero mai contemporaneamente nei due gruppi.

La distribuzione delle lesioni era la seguente:

Categoria diagnostica	Numero di lesioni	Percentuale
NSCLC	360	60,9%

Categoria diagnostica	Numero di lesioni	Percentuale
SCLC	57	9,6%
Lesioni benigne	83	14,0%
Metastasi	91	15,4%
Totale	591	100%

L'NSCLC rappresentava la categoria prevalente, mentre gli SCLC costituivano la classe più piccola. Questo sbilanciamento ha richiesto particolare attenzione nella scelta e nell'interpretazione delle metriche.

3.3 Classificazione diagnostica delle lesioni

Le lesioni sono state suddivise in quattro categorie:

- **NSCLC** – carcinomi polmonari non a piccole cellule;
- **SCLC** – carcinomi polmonari a piccole cellule;
- **lesioni benigne** – lesioni non maligne;
- **metastasi** – localizzazioni polmonari secondarie da neoplasie extrapolmonari.

La diagnosi di riferimento è stata definita in base al dato istologico disponibile per le lesioni maligne. In caso di biopsia non diagnostica è stato preso come riferimento l'istologico dell'intervento chirurgico, dove presente, o in alternativa il caso è stato escluso. In caso di lesione benigna, la lesione veniva classificata come tale solo in caso di confermata risoluzione o stabilità temporale per almeno 5 anni nel percorso clinico-radiologico.

3.4 Imaging TC e preprocessing

Le immagini TC sono state acquisite nell'ambito del normale percorso diagnostico-assistenziale dei pazienti, utilizzando scanner, protocolli di acquisizione e parametri di ricostruzione eterogenei. Tale variabilità riflette la natura retrospettiva e real-world della coorte analizzata.

Prima dell'estrazione radiomica, le immagini DICOM sono state organizzate e sottoposte a controllo di qualità, verificando inoltre la corretta corrispondenza spaziale tra ciascuna

immagine e la relativa maschera di segmentazione. Non sono state volutamente applicate procedure di preprocessing, normalizzazione delle intensità o resampling spaziale. Le immagini sono state pertanto analizzate nella loro forma nativa, mantenendo invariati voxel spacing, spessore di strato, kernel di ricostruzione e livelli di grigio, al fine di preservare le caratteristiche originarie dei dati e valutare la robustezza dell'approccio radiomico in condizioni rappresentative della pratica clinica reale.

3.5 Segmentazione delle lesioni

Le lesioni sono state segmentate da un gruppo di operatori con esperienza variabile, da 10 anni a nessuna esperienza specifica in ambito di contornamento radiologico, ognuno dei quali ha segmentato lo stesso numero di lesioni, al fine di valutare l'impatto della curva di apprendimento e del background clinico sulla riproducibilità delle feature radiomiche estratte.

Questa scelta metodologica risponde a un'esigenza specifica del disegno sperimentale: mentre gli studi di reliability radiomica pubblicati in letteratura tendono spesso a impiegare operatori esperti e tra loro comparabili per seniority, tale approccio non riflette la reale eterogeneità degli scenari clinici multicentrici, dove la segmentazione manuale o semi-automatica può essere delegata a figure professionali con gradi di expertise molto diversi (radiologi strutturati, specializzandi al primo anno, tecnici di radiologia dedicati al post-processing).

Per ogni lesione è stata definita una regione di interesse (ROI) corrispondente al volume lesionale – passaggio cruciale, dato che tutte le feature vengono calcolate a partire dai voxel inclusi nella maschera.

Nella segmentazione si è cercato di includere l'intero volume della lesione, escludendo strutture non pertinenti come vasi adiacenti, parete toracica, atelettasia o parenchima polmonare sano. L'immagine sottostante rappresenta un esempio di segmentazione di una lesione del lobo inferiore di destra, eseguita con tecnica semiautomatica.



Il significato concettuale di questo passaggio merita di essere ulteriormente approfondito, poiché si colloca al cuore del razionale della biopsia virtuale radiomica: mentre la biopsia tissutale tradizionale campiona per definizione solo una parte, spesso esigua, del volume tumorale — con il rischio intrinseco di errore di campionamento (sampling error) in presenza di eterogeneità spaziale intratumorale — l'analisi radiomica interroga l'intero volume segmentato, restituendo una caratterizzazione quantitativa che integra l'informazione dell'intera massa lesionale in un'unica analisi, non invasiva e ripetibile nel tempo.

Questa differenza non è solo tecnica, ma concettualmente fondativa: la biopsia istologica, per quanto rappresenti tuttora il gold standard diagnostico, fornisce un'istantanea di una porzione di un fenomeno biologico intrinsecamente disomogeneo, sia a livello macroscopico (necrosi, angiogenesi, aree di diversa cellularità) sia microscopico (eterogeneità genomica e clonale). L'estrazione di feature radiomiche dall'intero volume segmentato consente invece di catturare, seppur indirettamente e attraverso un correlato di secondo livello (texture, forma, distribuzione dei livelli di grigio), la componente di eterogeneità strutturale della lesione nella sua interezza, superando il limite di rappresentatività spaziale insito nel prelievo biottico.

3.6 Estrazione delle feature radiomiche

Da ciascuna lesione segmentata è stato estratto un set di feature radiomiche, per un totale di 851 feature, ottenute mediante il pacchetto Pyradiomics[23]. Le feature estratte appartengono a diverse classi radiomiche: le feature di forma (shape), che descrivono le caratteristiche morfologiche e geometriche della lesione indipendentemente dall'intensità dei voxel; le feature di primo ordine (first order), che riassumono la distribuzione statistica dei livelli di grigio all'interno del volume segmentato senza considerare la loro disposizione spaziale; le feature di texture, che catturano invece le relazioni spaziali tra voxel adiacenti attraverso matrici di co-occorrenza e di run-length (GLCM, GLRLM, GLSZM, GLDM, NGTDM), fornendo quindi un'informazione sull'eterogeneità strutturale interna alla lesione.

3.7 Noise Sensitivity Test

Per valutare la robustezza delle feature abbiamo usato il **Noise Sensitivity Test** (NST), che caratterizza ciascuna feature in base alla sua sensibilità al rumore e alla sua capacità di rappresentare un segnale realmente informativo.

Delle 851 feature totali, **669** disponevano di caratterizzazione NST, mentre **182** ne erano prive. Le feature caratterizzate sono state ordinate secondo un ranking di affidabilità e discriminatività.

Il NST non è stato usato solo come procedura tecnica di riduzione dimensionale, ma come criterio metodologico per rendere la biopsia virtuale più affidabile – privilegiando feature meno probabilmente dominate da rumore o instabilità tecnica.

3.8 Feature set analizzati

Abbiamo costruito 14 feature set:

Feature set	Numero di feature	Descrizione
ALL_raw_851	851	Tutte le feature radiomiche disponibili
NST_ranked_all_669	669	Tutte le feature con ranking NST disponibile

Feature set	Numero di feature	Descrizione
NST_top_500	500	Prime 500 feature secondo ranking NST
NST_top_400	400	Prime 400 feature secondo ranking NST
NST_top_300	300	Prime 300 feature secondo ranking NST
NST_top_200	200	Prime 200 feature secondo ranking NST
NST_top_100	100	Prime 100 feature secondo ranking NST
NST_top_50	50	Prime 50 feature secondo ranking NST
NST_top_25	25	Prime 25 feature secondo ranking NST
NST_HD_ge_80_524	524	Feature con alta discriminatività NST $\geq 80\%$
NST_HD_ge_85_470	470	Feature con alta discriminatività NST $\geq 85\%$
NST_HD_ge_90_391	391	Feature con alta discriminatività NST $\geq 90\%$
NST_HD_ge_95_311	311	Feature con alta discriminatività NST $\geq 95\%$
NST_HD_ge_98_176	176	Feature con alta discriminatività NST $\geq 98\%$

Questa strategia ha permesso di capire se una riduzione progressiva delle feature, guidata dal NST, mantenesse (o addirittura migliorasse) la performance del modello.

3.9 Modelli di machine learning

Per la costruzione del modello di machine learning sono stati messi a confronto sette algoritmi differenti, nel tentativo di valutare al contempo famiglie metodologiche concettualmente distanti tra loro. Accanto ai classici alberi decisionali in configurazione ensemble — Random Forest ed ExtraTrees, quest'ultimo caratterizzato da una maggiore componente di casualità nella scelta degli split e quindi tendenzialmente più resistente all'overfitting — sono stati considerati due algoritmi di gradient boosting, LightGBM e Histogram Gradient Boosting, entrambi pensati per gestire in modo efficiente dataset di dimensioni contenute come quello a disposizione, pur mantenendo buone capacità di modellazione delle interazioni non lineari tra le feature.

A questi si è affiancata una Support Vector Machine con kernel RBF, scelta per la sua capacità di definire confini decisionali non lineari anche in presenza di un numero relativamente limitato di osservazioni, condizione tipica degli studi radiomici. Infine, per completare il confronto con un estremo più conservativo dello spettro metodologico, sono stati inclusi due modelli di regressione logistica regolarizzata: uno con penalizzazione L2, più stabile in presenza di feature correlate tra loro (come spesso accade nelle feature radiomiche), e uno con penalizzazione elastic-net, che unendo componente L1 e L2 consente anche una forma di selezione automatica delle variabili più informative.

Questa scelta è stata dettata dalla volontà di rappresentare nel confronto sia modelli in grado di catturare relazioni complesse e non lineari tra le feature, sia modelli più semplici e interpretabili, così da poter valutare se un eventuale guadagno in accuratezza giustificasse, o meno, la perdita di trasparenza tipica dei modelli più sofisticati.

3.10 Validazione interna

La validazione è stata condotta con cross-validazione ripetuta 5×5, per un totale di 25 fold per ciascuna configurazione sperimentale.

La suddivisione è avvenuta a livello di paziente, non di singola lesione — aspetto fondamentale visto che diversi pazienti avevano più lesioni. Una suddivisione per lesione avrebbe rischiato di introdurre leakage, con lesioni dello stesso paziente presenti sia nel training che nel test set.

In ciascun fold, il training set includeva in media circa 473 lesioni (circa 347 pazienti), mentre il test set circa 118 lesioni (circa 87 pazienti).

3.11 Metriche di valutazione

Per la valutazione delle performance dei modelli sono state impiegate diverse metriche: accuracy, balanced accuracy, macro-F1, weighted-F1, macro-AUC one-vs-rest, weighted-AUC one-vs-rest e macro-AUPRC.

Data la disomogeneità nella distribuzione delle classi all'interno del dataset, si è scelto di non considerare l'accuracy semplice come metrica di riferimento principale: in un contesto sbilanciato come quello in esame, un modello che tendesse a privilegiare sistematicamente la classe maggioritaria (NSCLC) potrebbe infatti raggiungere valori di accuracy apparentemente elevati pur non riuscendo, di fatto, a riconoscere adeguatamente le classi minoritarie, che dal punto di vista clinico rappresentano spesso proprio le categorie di maggiore interesse diagnostico.

Per questo motivo l'interpretazione dei risultati si è basata principalmente su balanced accuracy, macro-F1 e macro-AUC one-vs-rest, metriche più robuste rispetto allo sbilanciamento delle classi e quindi più adatte a restituire un quadro realistico della capacità discriminativa dei modelli testati. Le restanti metriche (weighted-F1, weighted-AUC e macro-AUPRC) sono state comunque riportate a completamento del quadro valutativo, utili in particolare per verificare la coerenza dei risultati anche pesando le classi in base alla loro numerosità reale.

CAPITOLO 4

Risultati

4.1 Descrizione generale del dataset

Il dataset finale comprendeva **591 lesioni polmonari** in **434 pazienti**. La classe più rappresentata era l'NSCLC, con 360 lesioni (60,9% del totale); seguivano 57 SCLC, 83 lesioni benigne e 91 metastasi.

Classe	Numero di lesioni	Percentuale
NSCLC	360	60,9%
SCLC	57	9,6%
Lesioni benigne	83	14,0%
Metastasi	91	15,4%
Totale	591	100%

4.2 Feature radiomiche e costruzione della biopsia virtuale

Il set radiomico completo comprendeva **851 feature**, che rappresentavano la descrizione quantitativa dell'intero volume lesionale – la base informativa, in sostanza, della nostra biopsia virtuale.

Di queste, 669 erano caratterizzate tramite NST e 182 no. Il set **NST_ranked_all_669** ha rappresentato quindi il principale feature set "robusto", ottenuto escludendo le feature non valutabili con il NST.

Confrontare i diversi feature set ci ha permesso di verificare se la biopsia virtuale radiomica mantenesse una capacità discriminativa adeguata anche dopo una selezione delle feature basata sulla robustezza, piuttosto che sulla sola performance.

4.3 Performance globale dei modelli multiclass

Abbiamo testato **98 configurazioni sperimentali**, combinando 14 feature set e 7 algoritmi.

Il modello migliore in termini di balanced accuracy è stato **ExtraTrees sull'intero set ALL_raw_851**:

- accuracy media: **$0,470 \pm 0,066$** ;
- balanced accuracy media: **$0,556 \pm 0,059$** ;
- macro-F1: **$0,510 \pm 0,053$** ;
- macro-AUC one-vs-rest: **$0,757 \pm 0,044$** ;
- macro-AUPRC: **$0,566 \pm 0,051$** .

Subito dietro, **ExtraTrees sul set NST_ranked_all_669**:

- accuracy media: **$0,468 \pm 0,046$** ;
- balanced accuracy media: **$0,543 \pm 0,034$** ;
- macro-F1: **$0,504 \pm 0,033$** ;
- macro-AUC one-vs-rest: **$0,755 \pm 0,030$** ;
- macro-AUPRC: **$0,563 \pm 0,032$** .

Le configurazioni migliori, ordinate per balanced accuracy, sono riassunte nella tabella sottostante.

Feature set	Modello	N feature	Accuracy	Balanced accuracy	Macro-F1	Macro-AUC OvR	Macro-AUPRC
ALL_raw_851	ExtraTrees	851	0,470	0,556	0,510	0,757	0,566
NST_ranked_all_669	ExtraTrees	669	0,468	0,543	0,504	0,755	0,563
NST_ranked_all_669	Random Forest	669	0,461	0,542	0,501	0,753	0,560
NST_top_400	ExtraTrees	400	0,470	0,539	0,500	0,744	0,559
NST_HD_ge_85_470	ExtraTrees	470	0,479	0,539	0,500	0,743	0,555
NST_top_300	ExtraTrees	300	0,494	0,537	0,506	0,743	0,556
NST_HD_ge_80_524	ExtraTrees	524	0,476	0,537	0,499	0,753	0,564

Questi risultati confermano che la biopsia virtuale radiomica riesce a estrarre un segnale informativo reale dalle immagini TC, ma anche che la classificazione multiclass completa resta, nel complesso, moderata.

Una macro-AUC intorno a 0,75 indica una capacità discriminativa superiore al caso; una balanced accuracy intorno a 0,55, però, ci dice che distinguere simultaneamente quattro categorie diagnostiche è estremamente complesso.

4.4 Confronto tra algoritmi

Per quanto riguarda la performance media dei diversi algoritmi, ExtraTrees si conferma il più competitivo in termini di balanced accuracy e macro-AUC.

Modello	Accuracy media	Balanced accuracy media	Macro-F1	Macro-AUC OvR	Macro-AUPRC
ExtraTrees	0,478	0,525	0,491	0,736	0,547
Random Forest	0,473	0,512	0,483	0,730	0,541
SVM RBF	0,443	0,505	0,470	0,711	0,514
Logistic L2	0,463	0,504	0,468	0,710	0,515
Logistic elastic-net	0,466	0,504	0,467	0,710	0,514
LightGBM	0,528	0,502	0,494	0,726	0,542
Histogram Gradient Boosting	0,506	0,500	0,486	0,724	0,538

LightGBM ottiene in media un'accuracy più alta, ma con una balanced accuracy inferiore rispetto a ExtraTrees – comportamento che indica una certa tendenza a privilegiare la classe maggioritaria NSCLC.

In considerazione del fatto che l'obiettivo della biopsia virtuale è distinguere più categorie diagnostiche, non semplicemente massimizzare la correttezza globale, ExtraTrees risulta essere il modello principale più appropriato.

4.5 Il modello con maggiore accuracy

Il modello con l'accuracy media più alta è stato **LightGBM sul feature set NST_HD_ge_90_391**:

- accuracy media: **0,549 ± 0,057**;
- balanced accuracy: **0,512 ± 0,046**;
- macro-F1: **0,505 ± 0,039**;
- macro-AUC one-vs-rest: **0,737 ± 0,031**;
- macro-AUPRC: **0,556 ± 0,031**.

Nonostante l'accuracy migliore, la balanced accuracy è più bassa rispetto al miglior ExtraTrees. La maggiore accuracy sembra quindi riflettere soprattutto una classificazione più efficace della classe maggioritaria, più che una reale superiorità nella biopsia virtuale multiclass.

La matrice di confusione aggregata per questo modello:

Classe reale	Predetto NSCLC	Predetto SCLC	Predetto benigno	Predetto metastasi
NSCLC	254	4	48	54
SCLC	8	44	4	1
Benigno	37	0	29	17
Metastasi	46	1	19	25

Il modello riconosce meglio gli NSCLC rispetto a ExtraTrees, ma non migliora la distinzione delle classi minoritarie, in particolare metastasi e lesioni benigne.

4.6 Analisi per classe del modello NST_ranked_all_669 + ExtraTrees

Abbiamo approfondito l'analisi per classe sul modello ExtraTrees + NST_ranked_all_669, scelto perché combina buona performance complessiva e una selezione delle feature basata sul NST – il candidato più rappresentativo del concetto di biopsia virtuale robusta.

Le performance aggregate out-of-fold:

Classe	Precision	Recall	F1-score	AUC one-vs-rest
NSCLC	0,847	0,414	0,556	0,713
SCLC	0,978	0,789	0,874	0,907
Lesioni benigne	0,240	0,590	0,341	0,715
Metastasi	0,236	0,429	0,305	0,686

Il dato più evidente riguarda gli SCLC: nonostante rappresentassero la categoria meno rappresentata nel dataset, il modello li identifica con performance elevate. Una precision di 0,978 implica che quasi tutte le lesioni predette come SCLC lo erano davvero; una recall di 0,789 dice che circa il 79% degli SCLC reali viene riconosciuto; l'AUC one-vs-rest di 0,907 conferma una buona capacità di separarli dalle altre categorie.

Questo fa pensare che gli SCLC abbiano un fenotipo radiomico distintivo, rilevabile alla TC e interpretabile con il machine learning – probabilmente il risultato più promettente di tutto lo studio, in ottica di biopsia virtuale.

L'NSCLC mostra invece una precision alta ma un recall più bassa: quando il modello predice NSCLC, la predizione è spesso corretta, ma molti NSCLC reali vengono etichettati come benigni o come metastasi.

Lesioni benigne e metastasi mostrano entrambe performance più deboli, con valori bassi di precision – segno di una sovrapposizione radiomica consistente tra queste categorie e le altre classi.

4.7 Matrice di confusione del modello NST_ranked_all_669 + ExtraTrees

Classe reale	Predetto NSCLC	Predetto SCLC	Predetto benigno	Predetto metastasi
NSCLC	149	1	112	98
SCLC	2	45	6	4
Benigno	10	0	49	24
Metastasi	15	0	37	39

La matrice conferma che gli SCLC sono la classe più chiaramente identificabile: dei 57 SCLC reali, 45 vengono classificati correttamente, solo 2 confusi con NSCLC, 6 con benigni e 4 con metastasi.

Al contrario, distinguere NSCLC, metastasi e lesioni benigne resta problematico: dei 360 NSCLC, solo 149 vengono classificati correttamente, mentre 112 finiscono tra i benigni e 98 tra le metastasi.

Le metastasi vengono classificate correttamente in 39 casi su 91, ma 37 finiscono etichettate come benigne e 15 come NSCLC. Anche le lesioni benigne, a loro volta, vengono spesso confuse con le metastasi.

Questo pattern di errore suggerisce che la biopsia virtuale radiomica riesca a intercettare un segnale forte per gli SCLC, ma fatichi molto di più con categorie biologicamente e morfologicamente eterogenee come NSCLC, metastasi e lesioni benigne.

4.8 Il modello NST_top_300 + ExtraTrees

Un altro risultato degno di nota riguarda ExtraTrees addestrato sulle prime 300 feature secondo ranking NST. Nell'aggregazione out-of-fold ha ottenuto:

- accuracy: **0,518**;
- balanced accuracy: **0,561**;
- macro-F1: **0,528**.

Con questa matrice di confusione:

Classe reale	Predetto NSCLC	Predetto SCLC	Predetto benigno	Predetto metastasi
NSCLC	178	7	78	97
SCLC	5	46	3	3
Benigno	19	0	38	26
Metastasi	26	1	20	44

Rispetto al modello con 669 feature, questo mostra un miglior bilanciamento nella classificazione discreta – riconosce meglio NSCLC e metastasi, mantenendo comunque una buona performance sugli SCLC.

La macro-AUC, però, è più bassa rispetto al modello con 669 feature: sembra che una riduzione più spinta possa migliorare alcune metriche di classificazione puntuale, a scapito della qualità complessiva della separazione probabilistica.

Discussione

5.1 La radiomica TC come biopsia virtuale

In questo studio abbiamo esplorato la radiomica TC come strumento di biopsia virtuale per la caratterizzazione non invasiva delle lesioni polmonari. L'idea di fondo è che l'immagine TC non contenga solo informazioni morfologiche visibili al radiologo, ma anche pattern quantitativi più complessi, catturabili tramite feature radiomiche.

La biopsia virtuale non va intesa come una sostituzione della biopsia tissutale, ma come un approccio complementare: la biopsia tradizionale dà una diagnosi istologica diretta, mentre la biopsia virtuale prova a inferire informazioni biologiche a partire dall'intero volume lesionale.

Il vantaggio teorico è duplice: da un lato è non invasiva e sfrutta immagini già disponibili nel percorso diagnostico; dall'altro analizza l'intera lesione, non solo una porzione campionata, il che potrebbe permettere di cogliere aspetti dell'eterogeneità tumorale non sempre rappresentati dal prelievo biotico.

I nostri risultati suggeriscono che un segnale radiomico legato alla categoria diagnostica esiste, ma che è distribuito in modo tutt'altro che uniforme tra le classi. La classificazione multiclass completa resta difficile, mentre l'identificazione degli SCLC appare molto più promettente.

5.2 Interpretazione della performance multiclass

La performance globale del modello multiclass resta moderata: il miglior modello raggiunge una balanced accuracy di circa 0,56 e una macro-AUC di circa 0,76.

Questi numeri suggeriscono che la radiomica TC possiede una capacità discriminativa superiore al caso, ma non ancora sufficiente per una classificazione diagnostica autonoma delle quattro categorie. Il modello intercetta un segnale, ma non riesce ancora a trasformarlo in una biopsia virtuale completa e affidabile per tutte le classi.

Un risultato comunque coerente con la complessità reale del problema. Distinguere NSCLC, SCLC, metastasi e lesioni benigne non dipende solo dalla morfologia della lesione, ma anche dal contesto clinico, dalla storia oncologica, dai dati funzionali, dalla sede, dal numero di lesioni e da informazioni istologiche o molecolari non direttamente accessibili alla TC.

La classificazione multiclass va quindi considerata una prova di fattibilità, non un modello pronto per l'uso clinico.

5.3 Il fenotipo radiomico degli SCLC

Il risultato più rilevante di tutto lo studio riguarda gli SCLC, identificati con performance nettamente superiori rispetto alle altre classi, nonostante fossero la meno rappresentata nel dataset.

Il modello NST_ranked_all_669 + ExtraTrees ottiene per gli SCLC: precision 0,978, recall 0,789, F1-score 0,874, AUC one-vs-rest 0,907.

Numeri che indicano un pattern radiomico fortemente riconoscibile: la precision molto alta dice che quando il modello identifica una lesione come SCLC, la predizione è raramente sbagliata; la recall discreta indica che la maggior parte degli SCLC viene effettivamente intercettata.

Dal punto di vista biologico ha senso: gli SCLC sono tumori ad alta cellularità, crescita rapida, comportamento aggressivo, spesso con una presentazione clinico-radiologica peculiare. È plausibile che queste caratteristiche si traducano in pattern quantitativi di densità, texture o eterogeneità misurabili alla TC.

In ottica di biopsia virtuale, è il risultato più importante che abbiamo trovato. Anche se il modello multiclass globale non è ancora abbastanza forte, la capacità di riconoscere gli SCLC apre a un'applicazione più mirata: un modello binario SCLC vs non-SCLC.

Un modello del genere potrebbe avere un valore concreto nella fase pre-istologica, ad esempio per:

- segnalare lesioni con probabilità elevata di SCLC;
- aiutare a definire le priorità del percorso diagnostico;
- integrare la valutazione radiologica convenzionale;
- orientare il sospetto clinico in pazienti con lesioni polmonari sospette.

Naturalmente, prima di qualsiasi applicazione clinica servirebbe una validazione dedicata, una calibrazione e una conferma esterna.

5.4 Difficoltà nella distinzione tra NSCLC, metastasi e lesioni benigne

La distinzione di NSCLC, metastasi e lesioni benigne si è rivelata più complessa, con un alto tasso di errori reciproci di classificazione frequenti. Le spiegazioni di questo comportamento sono molteplici. In primo luogo, l'NSCLC è una categoria estremamente eterogenea, che include adenocarcinomi, carcinomi squamosi e altre forme meno comuni, ciascuna con caratteristiche biologiche e radiologiche proprie. L'aggregazione in un'unica classe può diluire la riconoscibilità radiomica del gruppo.

Le metastasi sono un'ulteriore fonte di eterogeneità: una metastasi polmonare può derivare da tumori primitivi molto diversi tra loro, ciascuno con caratteristiche biologiche e strutturali proprie. Il pattern radiomico di una metastasi potrebbe non essere uniforme.

Le lesioni benigne, a loro volta, comprendono entità diverse – processi infiammatori, fibrotici, granulomatosi, esiti cicatriziali – alcune delle quali possono mimare lesioni maligne sia alla lettura visiva sia all'analisi quantitativa.

Infine, molte informazioni utili a distinguere queste categorie non sono contenute nella sola ROI lesionale. Storia clinica, presenza di neoplasie note, numero di noduli, distribuzione anatomica, crescita nel tempo, eventuale captazione PET possono fare la differenza.

Una biopsia virtuale davvero efficace dovrà integrare la radiomica con dati clinici e radiologici convenzionali, piuttosto che basarsi solo sulle feature quantitative della lesione.

5.5 Ruolo del Noise Sensitivity Test

Il Noise Sensitivity Test è forse l'elemento metodologicamente più originale di questo lavoro. Lo scopo era capire se selezionare feature più robuste potesse mantenere la performance del modello radiomico.

Il confronto tra ALL_raw_851 e NST_ranked_all_669 mostra che togliere le feature non caratterizzate dal NST non comporta una perdita sostanziale di performance: la macro-AUC passa da 0,757 a 0,755, la balanced accuracy si riduce solo lievemente, da 0,556 a 0,543.

È un risultato importante, perché suggerisce che la performance della biopsia virtuale non dipende necessariamente dall'includere indiscriminatamente tutte le feature disponibili. Un set più controllato, caratterizzato in termini di sensibilità al rumore, può mantenere quasi invariata la capacità predittiva.

In radiomica questo aspetto è estremamente rilevante: un modello predittivo non deve essere valutato solo sulla base della performance interna, ma anche sulla plausibilità e robustezza delle feature su cui si basa. Feature troppo sensibili al rumore possono dare risultati apparentemente buoni nel dataset di sviluppo, ma poco riproducibili altrove.

Il NST può essere valutato, in questo senso, come uno strumento di controllo qualità della biopsia virtuale: avvicina il workflow radiomico a un approccio più traslazionale, in cui la selezione delle feature non dipende solo dalla performance statistica, ma anche dalla robustezza del segnale.

5.6 Implicazioni cliniche

Le possibili implicazioni cliniche dell'approccio proposto vanno valutate con cautela. I risultati non supportano l'uso del modello multiclass come alternativa alla biopsia istologica – ma aprono comunque a prospettive interessanti.

La prima riguarda l'identificazione degli SCLC. Un modello radiomico capace di segnalare un'elevata probabilità di SCLC potrebbe avere valore come supporto decisionale nella fase pre-istologica: data l'aggressività biologica di questi tumori e la necessità di un trattamento rapido, uno strumento di triage radiomico non sarebbe da poco.

La seconda riguarda la stratificazione delle lesioni indeterminate: in pazienti con lesioni polmonari sospette, la biopsia virtuale potrebbe integrare la valutazione radiologica convenzionale con una stima quantitativa del rischio.

La terza riguarda la pianificazione della biopsia reale: analizzando l'intero volume lesionale, la radiomica potrebbe in futuro aiutare a individuare le aree più rappresentative, o più eterogenee, da campionare.

La quarta riguarda la ricerca traslazionale: capire quali feature siano più associate al fenotipo SCLC potrebbe rivelarsi utile anche per comprenderne le basi biologiche.

5.7 Limiti dello studio

Lo studio ha diversi limiti.

Il primo è lo sbilanciamento delle classi: gli NSCLC costituiscono oltre il 60% del dataset, gli SCLC meno del 10%. Questo condiziona l'addestramento dei modelli e impone cautela nell'interpretare l'accuracy.

Il secondo è la natura retrospettiva: le immagini sono state acquisite nella pratica clinica corrente, quindi con una certa eterogeneità nei protocolli di acquisizione.

Il terzo è l'assenza di una validazione esterna: pur avendo condotto la validazione interna con separazione per paziente, non è possibile valutare con certezza quanto il modello si generalizzi ad altri centri, scanner o popolazioni.

Il quarto è l'eterogeneità interna delle classi: NSCLC, metastasi e lesioni benigne includono entità molto eterogenee. Una classificazione più granulare sarebbe probabilmente più informativa, ma richiederebbe una casistica più ampia.

Il quinto riguarda la segmentazione: le feature radiomiche dipendono dalla qualità e dalla coerenza delle maschere lesionali, e differenze nella segmentazione possono influenzare i valori delle feature e quindi la performance del modello.

Il sesto è l'assenza di integrazione con dati clinici: una vera biopsia virtuale utile in clinica dovrebbe probabilmente combinare radiomica, dati clinici, anamnesi oncologica, sede della lesione, numero di noduli, crescita temporale ed eventualmente dati PET.

Il modello, in definitiva, va considerato esplorativo. I risultati sono promettenti, soprattutto per gli SCLC, ma richiedono conferma con analisi dedicate e validazione esterna.

5.8 Prospettive future

Le prospettive future principali riguardano lo sviluppo di modelli più specifici e più vicini a un'applicazione clinica reale.

La prima direzione è costruire un modello binario SCLC vs non-SCLC: i risultati multiclass indicano chiaramente che gli SCLC sono la categoria più riconoscibile dal punto di vista radiomico, e un modello dedicato potrebbe raggiungere performance più alte e più facilmente interpretabili.

La seconda è integrare dati clinici e radiologici convenzionali: una biopsia virtuale davvero utile dovrebbe probabilmente combinare feature radiomiche, caratteristiche morfologiche, dati anamnestici e informazioni cliniche.

La terza è la validazione esterna: solo testando il modello su dati di altri centri sarà possibile capire quanto sia realmente robusto il workflow.

La quarta è l'analisi dell'interpretabilità: capire quali feature contribuiscano di più alla classificazione degli SCLC potrebbe dare informazioni preziose anche sul piano biologico.

La quinta è usare la radiomica come guida alla biopsia reale: in futuro, la biopsia virtuale potrebbe non sostituire quella tissutale, ma migliorarla, suggerendo quali lesioni o quali aree della lesione siano più rappresentative da campionare.

Conclusioni

In questo lavoro abbiamo valutato una possibile biopsia virtuale basata su radiomica per la caratterizzazione non invasiva delle lesioni polmonari.

I risultati mostrano che le immagini TC contengono un segnale radiomico legato alla categoria diagnostica della lesione. Tuttavia, la classificazione multiclass completa di NSCLC, SCLC, metastasi e lesioni benigne resta, nel complesso, moderata: il miglior modello raggiunge una balanced accuracy di 0,556 e una macro-AUC one-vs-rest di 0,757.

La selezione delle feature tramite Noise Sensitivity Test ha permesso di ridurre la dimensionalità del modello senza perdita sostanziale di performance: il modello basato sulle sole feature caratterizzate dal NST ottiene infatti una macro-AUC di 0,755, praticamente sovrapponibile a quella del modello completo.

Il risultato più rilevante riguarda gli SCLC, identificati con performance elevate (AUC one-vs-rest 0,907, F1-score 0,874) – un dato che suggerisce l'esistenza di un fenotipo radiomico distintivo per il carcinoma polmonare a piccole cellule.

Al contrario, distinguere NSCLC, metastasi e lesioni benigne resta più complesso, verosimilmente per la sovrapposizione morfologica e l'eterogeneità biologica interna di queste categorie.

In conclusione, la radiomica TC ci sembra un approccio promettente per sviluppare una biopsia virtuale delle lesioni polmonari. Allo stato attuale, il modello multiclass non può sostituire la biopsia istologica, ma rappresenta una prova di fattibilità e un buon punto di partenza per sviluppi futuri.

La direzione più promettente resta lo sviluppo di modelli dedicati alla distinzione SCLC vs non-SCLC, integrati con dati clinici e validati esternamente. Il Noise Sensitivity Test si conferma uno strumento utile per rendere il workflow radiomico più robusto e metodologicamente credibile.

Bibliografia

1. Mazzone PJ, Lam L (2022) Evaluating the Patient With a Pulmonary Nodule: A Review. *JAMA* 327:264–273. <https://doi.org/10.1001/jama.2021.24287>
2. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, Petrella F, Spaggiari L, Rosell R (2015) Non-small-cell lung cancer. *Nat Rev Dis Primers* 1:15009. <https://doi.org/10.1038/nrdp.2015.9>
3. Duma N, Santana-Davila R, Molina JR (2019) Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin Proc* 94:1623–1640. <https://doi.org/10.1016/j.mayocp.2019.01.013>
4. Wang Q, Gümüş ZH, Colarossi C, Memeo L, Wang X, Kong CY, Boffetta P (2023) SCLC: Epidemiology, Risk Factors, Genetic Susceptibility, Molecular Pathology, Screening, and Early Detection. *J Thorac Oncol* 18:31–46. <https://doi.org/10.1016/j.jtho.2022.10.002>
5. Kim SY, Park HS, Chiang AC (2025) Small Cell Lung Cancer: A Review. *JAMA* 333:1906–1917. <https://doi.org/10.1001/jama.2025.0560>
6. Franquet T, Müller NL, Giménez A, Guembe P, de La Torre J, Bagué S (2001) Spectrum of pulmonary aspergillosis: histologic, clinical, and radiologic findings. *Radiographics* 21:825–837. <https://doi.org/10.1148/radiographics.21.4.g01jl03825>
7. Heitzman ER (1986) The role of computed tomography in the diagnosis and management of lung cancer. An overview. *Chest* 89:237S–241S
8. Ananthasayanam JR, Saravanan J, Gunasekaran D, Meganathan S, Natarajan P (2024) Comprehensive Evaluation of Pulmonary Masses Using Multi-detector CT: Correlating Morphological Features With CT-Guided Histopathological Findings for Enhanced Diagnostic Accuracy. *Cureus* 16:e69377. <https://doi.org/10.7759/cureus.69377>
9. Saggiante L, Biondetti P, Lanza C, Carriero S, Ascenti V, Piacentino F, Shehab A, Ierardi AM, Venturini M, Carrafiello G (2024) Computed-Tomography-Guided Lung Biopsy: A Practice-Oriented Document on Techniques and Principles and a Review of the Literature. *Diagnostics (Basel)* 14:1089. <https://doi.org/10.3390/diagnostics14111089>
10. Laurent F, Montaudon M, Latrabe V, Bégueret H (2003) Percutaneous biopsy in lung cancer. *Eur J Radiol* 45:60–68. [https://doi.org/10.1016/s0720-048x\(02\)00286-3](https://doi.org/10.1016/s0720-048x(02)00286-3)
11. Gill RR (2023) Virtual Image-based Biopsy of Lung Metastases: The Promise of Radiomics. *Acad Radiol* 30:47–48. <https://doi.org/10.1016/j.acra.2022.10.030>
12. Chassagnon G, De Margerie-Mellon C, Vakalopoulou M, Marini R, Hoang-Thi T-N, Revel M-P, Soyer P (2023) Artificial intelligence in lung cancer: current applications and perspectives. *Jpn J Radiol* 41:235–244. <https://doi.org/10.1007/s11604-022-01359-x>
13. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 278:563–577. <https://doi.org/10.1148/radiol.2015151169>

14. Kim D, Jensen LJ, Elgeti T, Steffen IG, Hamm B, Nagel SN (2021) Radiomics for Everyone: A New Tool Simplifies Creating Parametric Maps for the Visualization and Quantification of Radiomics Features. *Tomography* 7:477–487. <https://doi.org/10.3390/tomography7030041>
15. Pinto Dos Santos D, Dietzel M, Baessler B (2021) A decade of radiomics research: are images really data or just patterns in the noise? *Eur Radiol* 31:1–4. <https://doi.org/10.1007/s00330-020-07108-w>
16. Tomaszewski MR, Gillies RJ (2021) The Biological Meaning of Radiomic Features. *Radiology* 299:E256. <https://doi.org/10.1148/radiol.2021219005>
17. Ngiam KY, Khor IW (2019) Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 20:e262–e273. [https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)
18. Rajput D, Wang W-J, Chen C-C (2023) Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics* 24:48. <https://doi.org/10.1186/s12859-023-05156-9>
19. Barabino E, Rossi G, Fedeli A, Cittadini G, Genova C (2021) Radiomic-based diagnostics in oncology: challenges toward clinical practice. *Oncoscience* 8:72–73. <https://doi.org/10.18632/oncoscience.536>
20. Barabino E, Tosques M, Battaglia AP, Ficarra G, Rossi G, Fedeli A, Genova C, Cittadini G (2025) Hunting for signal in the noise: identifying robust CT radiomic features across variable scanning settings. *Eur Radiol*. <https://doi.org/10.1007/s00330-025-12215-7>
21. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, Sabater S (2018) Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* 288:407–415. <https://doi.org/10.1148/radiol.2018172361>
22. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z (2016) Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep* 6:34921. <https://doi.org/10.1038/srep34921>
23. <https://pyradiomics.readthedocs.io/en/latest/features.html>